

Information Fusion of Metabolite and Brain MRI
parameters on the Cardio-Metabolic Risk Factors
investigating Cognition in the Middle-Aged:
the CARDIA Study

Su-Jin Kang

Thesis submitted to Imperial College London for the
Degree of Doctor of Philosophy

July 2021

Department of Metabolism, Digestion and Reproduction
Faculty of Medicine
Imperial College London
UK

Declaration of originality

I hereby declare that this thesis was conducted and written by myself,
and that where any material could be construed as the work of others,
it is fully cited and referenced, and/or with appropriate acknowledgement given.

Copyright Statement

The copyright of this thesis rests with the author and is made available
under a Creative Commons Attribution Non-Commercial licence.

Researchers are free to copy, distribute or transmit the thesis on the condition
that they attribute it, that they do not use it for commercial purposes and
that they do not alter, transform or build upon it. For any reuse or redistribution,
researchers must make clear to others the license terms of this work.

Acknowledgement

I would like to express my sincere appreciation to my **supervisors** – Dr Timothy Ebbels, Dr Queenie Chan, and Professor Julian Griffin for their support and advice during the study. I would especially like to thank Professor Jeremy K. Nicholson, Professor Elaine Holmes, and Dr Isabel Garcia Perez who provided the opportunity to study for a PhD in the department with a full scholarship. I express my sincere appreciation to my funders **Medical Research Council** and **Merck KGaA**.

I would also like to express my sincere appreciation to Dr Katie Meyer (University of North Carolina), Dr Lenore Launer (National Institutes of Health), Professor Kristine Yaffe (University of California San Francisco) who supported my work on the manuscript proposals, ‘Metabolites, cardio-metabolic risk factors and brain MRI (#A-1903; #A-2017)¹’ and ‘Metabolites, cardio-metabolic risk factors and cognition (#A-1904; #A-2018)²’ which were approved for research by the Coronary Artery Risk Development in Young Adults (CARDIA) Publications and Presentations Committee in May 2019 and September 2020. I am sincerely grateful to the CARDIA committee for permitting the manuscript proposals and the CARDIA data acquisition. I am grateful to Dr Lucia Juarez (Director Statistical Computing in The University of Alabama at Birmingham) who delivered the data and materials. I’d especially like to give thanks to Dr Luke Gray Whiley who supported the high-resolution nuclear magnetic resonance (NMR) spectroscopy, liquid-chromatography mass spectrometry (LC-MS) data acquisition and their interpretations. I am also grateful to Dr Gonalo Correia, Dr Caroline Sands and Dr Beatriz Jimenez in the MRC-NIHR National Phenome Centre, Imperial College London who helped me to handle and interpret the raw NMR data.

I am thankful to Ms Nousheen Tariq, a postgraduate education manager at Faculty of Medicine Centre who supported Medical Research Council Doctoral Training Partnership students, and to Dr Janet De Wilde, a former head of postgraduate professional development at the Graduate School who tried to enhance student experience through retreat courses. I also appreciate Dr Fiona Pereira, a Programme Manager of Stratified Medicine Graduate Training Programme in Systems Medicine and Spectroscopic Profiling. Lastly, I would like to thank the departmental staffs working on the training courses including NMR, LC-MS, GC-MS and metabolic profiling.

I would like to acknowledge the Alzheimer’s Society, and Professor Sarah Lewis, Professor John Gladman and Dr Veronika Van Der Wardt, School of Medicine University of Nottingham who initially provided the motivation of the study and provided encouragement and suggestions regarding dementia in my previous study. I also would like to acknowledge Professor Kosh Ray and Dr Ailsa J McKay, School of Public Health, Imperial College who provided the initial motivation to study cardiovascular

¹ Sujin Kang, Queenie Chan, Timothy Ebbels, Julian Griffin, Elaine Holmes, Lenore Launer, Luke Whiley, Kristine Yaffe, Katie Meyer

² Sujin Kang, Queenie Chan, Timothy Ebbels, Julian Griffin, Elaine Holmes, Lenore Launer, Luke Whiley, Kristine Yaffe

disease. I sincerely thank Professor Jae-Won Lee, Department of Statistics, Korea University, and Dr. Carl Donovan, School of Mathematics and Statistics, University of St Andrews.

This doctoral study is a great opportunity to develop a greater depth of understanding of mechanisms and pathology, which is important in appropriate courses of identification and management of risks or diseases.

Lastly, my deepest thanks and **love to my parents**, family members and friends for their support.

List of Abbreviations

ABPM	Ambulatory Blood Pressure Monitoring
ACE	Angiotensin Converting Enzyme
ACG	Anterior Cingulate Gyrus
AD	Alzheimer's Disease
AF	Atrial Fibrillation
AHT	Antihypertensive Treatment
APOE ϵ 4	Apolipoprotein ϵ 4 allele
APP	Amyloid Precursor Protein
ARBs	Angiotensin II Receptor Blockers
ASE	Average Square Error
ATV	Abnormal Tissue Volume
AUC	Area Under the Curve
AV	Activated Voxels
BBB	Blood-Brain Barrier
BLSA	Baltimore Longitudinal Study of Aging
BMI	Body Mass Index
CABG	Coronary Artery Bypass Graft
CAIDE	Cardiovascular Risk Factors, Aging and Incidence of Dementia
CARDIA	Coronary Artery Risk Development in Young Adults
CBF	Cerebral Blood Flow
CBM	Mean of Cerebral Blood Flow values in an ROI
CCSMHA	Cache County Study of Memory Health and Aging
CCTs	Controlled Clinical Trials
CES-D	Center for Epidemiologic Studies-Depression
CHD	Coronary Heart Disease
CHS	Cardiovascular Health Study
C-index	Concordance Index
CI _s	Confidence Intervals
CrI _s	Credible Intervals
CSF tau	Cerebrospinal Fluid tau protein
CT	Computerized Tomography
CVD	Cardiovascular Disease
DBP	Diastolic Blood Pressure
DMA	Dirichlet-Multinomial Allocation
DSST	Digit Symbol Substitution Test
D-statistic	Cohen's d statistic
DTI	Diffusion Tensor Imaging
EC	Entorhinal Cortex
eGFR	Estimated Glomerular Filtration Rate
EPIC	European Prospective Investigation into Cancer and Nutrition
ERT	Oestrogen Replacement Therapy
FA	Factional Anisotropy
FAM	Mean of FA values in an ROI

FINMONICA	Finnish contribution to the World Health Organization Multinational MONItoring of Trends and Determinants in CARdiovascular Disease study
GEE	Generalized Estimating Equations
GM	Gray Matter
GPCOG	General Practitioner Assessment of Cognition
HBP	High Blood Pressure
HBPM	Home Blood Pressure Monitoring
HDL-C	High-Density Lipoprotein Cholesterol
HOMA-IR	Homeostatic Model Assessment for Insulin Resistance
HP	High-Performance
HR	Hazard Ratio
HTR	Hormone Trial Randomization assignment
ICV	Intracranial Volume
LC-MS	Liquid-Chromatography Mass Spectrometry
LDL-C	Low-Density Lipoprotein Cholesterol
LOESS	Locally Estimated Scatterplot Smoothing
MCI	Mild Cognitive Impairment
MCMC	Markov Chain Monte Carlo
MCS	Mental Component Score
MeSH	Medical Subject Headings
MFC	Medial Frontal Cortex
MI	Myocardial Infarction
MoCA	Montreal Cognitive Assessment
MRIs	Magnetic Resonance Imaging-derived parameters
NMF	Non-negative Mmatrix Ffactorization
NMR	Nuclear Mmagnetic Rresonance
NPHS	National Population Health Survey
OIM	Observed Information Matrix
OPoRT	Obesity Population Risk Tool
OR	Odd Ratio
PCA	Principal Component Analysis
PCS	Physical Component Score
PET	Positron Emission Tomography
PQN	Probabilistic Quotient Normalisation
QCs	Pooled Quality Control samples
QOF	Quality and Outcomes Framework
QRISK	QRESEARCH Cardiovascular Risk Algorithm
RAVLT	Rey Auditory Verbal Learning Test
RCTs	Randomized Controlled Trials
ROIs	Anatomical Regions Of Interest
ROS	Reactive Oxygen Species
RSD	Relative Standard Deviation
RSS	Residual Sum of Squares
SAT	Subcutaneous Adipose Tissue
SBP	Systolic Blood Pressure
SCORE	Systematic COronary Risk Evaluation
SEM	Structural Equation Modelling

SF-12	12-Item Short Form Health Survey
6CIT	Six-Item Cognitive Impairment Test
TBV	Total Brain Volume
TC	Total serum Cholesterol
THIN	The Health Improvement Network
3MS	Modified Mini-Mental State
TRM	Mean of Time to Repeat values in an ROI
TTV	Total Tissue Volume
UPLC-MS	Ultra-Performance Liquid Chromatography-Mass Spectrometry
VaD	Vascular Dementia
VLDL	Very-Low-Density Lipoprotein
VRA	Number of Activated Voxels within the ROI
VRM	Mean percent change of Vascular Reactivity in the ROI
WC	Waist Circumference
WHIMS	Women's Health Initiative Study
WHO	World Health Organization
WHR	Waist to Hip Ratio
WM	White Matter
WML	White Matter Lesions

Abstract

Background: There is a lack of evidence of the role of cardio-metabolic risk factors in brain health, especially those which are comprised of physiological information. The aim of this study is to investigate whether urinary metabolites representing systemic metabolism and brain magnetic resonance imaging (MRI)-derived parameters are prognostic markers regarding cognition with a cardio-metabolic risk factor.

Materials and methods: The study focuses on sub-cohort data from the Coronary Artery Risk Development in Young Adults (CARDIA) Study, with targeted metabolomics data by high-resolution nuclear magnetic resonance spectroscopy and liquid-chromatography mass spectrometry (N=606), and brain MRI data with 280 participants (48–60 years) at Year 30, 2015–16; brain MRI data for 185 participants were also acquired at Year 25, 2010–11. A *path analysis* was conducted to investigate mediators of adjusted cognition with cardio-metabolic factors measured at Year 30. Bayesian multilevel models were employed to examine the degeneration of cognition recorded at Years 25 and 30.

Results: Phenylalanine, aminoadipic acid, and tryptophan showed positive associations with fasting glucose in terms of hyperglycaemia, while indole-3-acetic acid showed a negative association. Measures of brain physiology, i.e. cerebral blood flow (CBF) in temporal lobe white matter left, and CBF in left entorhinal area showed negative associations with waist circumference, respectively, in terms of (abdominal) obesity in the cohort (N=74). For cognitively impaired individuals, significant exposures (race, education, family income, sex, age, center for epidemiologic studies-depression, smoking), metabolites (glycine, methionine, aminoadipic acid, tryptophan and kynurenine) and brain MRI-derived parameters (the number of activated voxels within temporal lobe, right anterior cingulate gyrus, and fractional anisotropy in right medial frontal cortex) were identified as physiological markers.

Conclusions: This study, employing a fusion of biostatistical analytic methods using longitudinal epidemiological data, was able to discover complex multivariate epidemiologic, pathologic and phenotypic relationships across neurodegenerative disease networks that were previously unidentified.

Keywords: Cardio-metabolic risk factor, Cognitive impairment, Dementia, Metabolites, Brain MRI parameters, Preventive medicine, Structural equation modelling, Bayesian multilevel regression, Decision-tree, Dimensionality-reduction

Table of Contents

Acknowledgement.....	3
List of Abbreviations	5
Abstract	8
Chapter I. Background.....	13
1. Cardio-metabolic dysfunctions.....	13
1-1. Hypertension.....	13
1-2. Dyslipidemia	14
1-3. Hyperglycaemia.....	15
1-4. (Abdominal) Obesity	15
1-5. The Framingham Risk Score for a cardiovascular disease assessment	15
2. Cardio-metabolic dysfunctions and its impact on cognition.....	16
Hypertension and its impact on cognition.....	16
Dyslipidaemia and its impact on cognition	17
Hyperglycaemia and its impact on cognition	17
Obesity and its impact on cognition	17
Lifestyle factor and its impact on cognition.....	17
3. Systematic review and meta-analysis: coronary heart disease and myocardial infarction impacts on cognition	18
1-1. Cognitive decline or dementia on coronary heart disease	20
1-2. Cognitive decline or dementia on myocardial infarction	21
4. Objectives of the study.....	23
Chapter II. Methods	26
1. Study design and population	26
1-1. The CARDIA Study.....	26
The CARDIA Study: baseline to Year 30.....	26
The study population and data at Year 25 and Year 30	26
Cardio-metabolic risk factors	27
Other exposures and confounders.....	27
Outcome measures: cognition scores.....	28
Prognostic markers: urinary metabolites at Year 30.....	28
Prognostic markers: brain MRI parameters	28
1-2. Demographic information at Year 30	29
1-3. Cardio-metabolic dysfunctions at Year 30.....	33
Hypertension	35
Antihypertensive treatments records in the CARDIA Study at Year 30.....	35
Dyslipidemia	36
Hyperglycaemia.....	36
(Abdominal) Obesity	38

The Framingham Risk Score.....	38
1-4. Cognitive impairment at Year 25 and Year 30.....	39
Construction of a composite cognitive score	40
1-5. Brain MRI parameters at Year 25 and Year 30	42
Total volume	43
Abnormal tissue volume	44
Fractional anisotropy	45
Cerebral blood flow	46
Brain structure involved in the successive stages of memory processing.....	47
Brain structure as early markers of Alzheimer's disease.....	51
1-6. Multi-analytical platforms used in metabolic phenotyping at Year 30.....	57
Method of NMR spectroscopy	57
Method of UPLC–MS.....	58
2. Metabolomics.....	58
2-1. Normalisation of urinary metabolites at Year 30	58
Process of the probabilistic quotient normalisation.....	59
2-2. Normalisation verification.....	60
Identical urinary metabolites assayed by both nuclear magnetic resonance (NMR) and liquid-chromatography mass spectrometry (LC–MS)	64
2-3. The NMR-based urinary metabolites	66
2-4. The LC–MS-based Amino Acids.....	68
2-5. The LC–MS-based tryptophan pathway metabolites.....	72
2-6. Metabolic syndrome with the metabolites	74
Metabolic syndrome with the NMR-based urinary metabolites.....	74
Metabolic syndrome with the LC–MS-based amino acids	74
Metabolic syndrome with the LC–MS-based tryptophan pathway metabolites.....	75
3. Statistical methods for analysing prognostic markers for brain health	76
Chapter III. Predictions of cognition at Year 30.....	78
1. Predictions of cognition based on urinary metabolites	78
1-1. Mediations on cardio-metabolic risk factors	78
1-2. Total path diagram of the NMR-based urinary metabolites.....	82
1-3. Total path diagram of the LC–MS-based amino acids.....	86
1-4. Total path diagram of the LC–MS-based tryptophan pathway metabolites	91
1-5. Total path diagram of the metabolites – Summary and discussion	93
2. Predictions of cognition with the brain magnetic resonance imaging (MRI)-derived parameters.....	95
2-1. Brain MRI parameters on cardio-metabolic risk factors.....	97
Brain volume loss	97
Gray matter volume loss/ gray matter lesions	98
White matter volume/ white matter lesions.....	98
2-2. Total path diagram of the brain MRI parameters	100

2-3. Total path diagrams of the metabolites and brain MRI parameters – Summary and discussion	105
3. Predictions of cognition from metabolites and brain MRI parameters	106
3-1. Total path diagram of the metabolites and brain MRI parameters.....	107
Chapter IV. Degeneration of cognition at Year 25 and Year 30	114
1. Degeneration of cognition with a cardio-metabolic risk factor	115
2. Degeneration of the brain MRI parameters with a cardio-metabolic factor	118
3. Degeneration of cognition with the brain MRI parameters adjusting for a cardio-metabolic factor	124
Chapter V. Data fusion for cognition – Integrated map of risk-specific markers	131
1. Network analysis measured at Year 30.....	131
1-1. Network analysis of the metabolites.....	133
1-2. Network analysis of the brain MRI parameters	135
1-3. Network analysis of the metabolites and brain MRI parameters	136
2. Network analysis of the brain MRI parameters at Year 25 and Year 30	137
3. Classification and profiling using the metabolites	139
3-1. Tree-based statistical models for the metabolites	139
3-2. The principal components and profiling using the metabolites	142
4. Classification and profiling using the brain MRI parameters	145
4-1. Tree-based statistical models for the brain MRI parameters	145
4-2. The principal components and profiling using the brain MRI parameters.....	147
Discussion on profiling using the brain MRI parameters	150
Summary of integrated map	152
Chapter VI. Design of a risk assessment tool for brain health	153
1. Conventional risk prediction tools and their limitations	153
1-1. Conventional risk tools for cardiovascular disease and the limitations	153
1-2. Conventional risk tools for obesity and their limitations	154
1-3. Conventional risk tools for dementia and their limitations.....	154
2. Development of a novel risk assessment tool of brain health	156
2-1. Identifying people who are liable to develop dementia in later life.....	156
2-2. Metabolic phenotyping and its application to stratified medicine	157
Omics data fusion and decision level fusion	157
2-3. Development of a novel risk assessment tool of cardio-metabolic risk-specific brain health	157
Chapter VII. Discussion	160
1. Main findings.....	160
The total path diagram of metabolites	160
The total path diagram of the brain MRI parameters.....	161
The total path diagram of the metabolites and the brain MRI parameters	161

The degeneration of cognition with the brain MRI parameters adjusting for a cardio-metabolic factor	161
To provide a cardio-metabolic risk-specific integrated map	162
1-1. Discussion and suggestions for future research	163
Metabolic pathways for the gut-brain axis or changes in metabolites associated with brain inflammation	163
Metabolic group separation and network-based modelling	164
1-2. Strengths and limitations	165
Strengths	165
Limitations	165
2. Conclusions	166
References	167
Appendix	179

Chapter I. Background

Cardio-metabolic syndrome is a combination of metabolic dysfunctions identified by a variety of metabolic disorders or risk factors. Metabolic syndrome is defined as the presence of at least three of the following cardio-metabolic risk factors according to the adult treatment panel (ATP) III guidelines and the CARDIA Study: waist circumference (WC) (> 88 cm for women or > 102 cm for men), triglycerides (≥ 150 mg/dL), high-density lipoprotein cholesterol (HDL-C) (< 50 mg/dL for women or < 40 mg/dL for men), either systolic blood pressure (SBP) (≥ 130 mmHg) or diastolic blood pressure (DBP) (≥ 85 mmHg), and fasting glucose (≥ 110 mg/dL) [4]. Participants who reported using medications for hypertension control were classified as having met the criterion for elevated blood pressure in the CARDIA Study [5]. The age-adjusted (interquartile range, 33–59) weighted prevalence of metabolic syndrome among US adults was 34.3 ± 0.8 (SE) % using the National Health and Nutrition Examination Survey (NHANES) 2007–2014 data [6]. A total of 20.9% of men without clinical prostatic hyperplasia had metabolic syndrome using the UK Clinical Practice Research Datalink (CPRD) in 2011 [7]. The prevalence of cardio-metabolic syndrome has risen and affected use of health care resources, morbidity and mortality, resulting in an enormous economic burden [8].

In this chapter, I introduce each of the main cardio-metabolic dysfunctions – hypertension, dyslipidaemia, hyperglycaemia, abdominal obesity, and the heart disease risk score.

1. Cardio-metabolic dysfunctions

1-1. Hypertension

Hypertension (i.e. high blood pressure, HBP) is defined as 140/90 mmHg (SBP/DBP) or higher (including subsequent ambulatory blood pressure monitoring (ABPM) daytime average or home blood pressure monitoring (HBPM) average blood pressure 135/85 mmHg or higher for those under 80 years, without diabetes or chronic kidney disease) [9]. Hypertension is known to be a major risk factor for heart attacks or failure including myocardial infarction (MI), atrial fibrillation (AF, irregular and often abnormally fast heart rate) by about two-fold, ischaemic and haemorrhagic stroke, chronic kidney disease, cognitive decline and premature death [9, 10]. Hypertension is also a risk factor for long term organ damage, as a result of injury to the arterial supply. HBP is known to thicken the blood vessels, and leads to the development of plaque, which eventually bursts into the blood vessel and causes direct occlusion or more often a blood clot within the blood vessel. This can either block the vessel suddenly or break off, and float and block the blood vessel downstream, which explains the resultant heart attack and stroke [11, 12]. HBP also can damage the small blood vessels by provoking a reaction whereby cells in the wall of the blood vessel become overgrown and slowly occlude them, which explains the resultant kidney damage [13]. Hypertension also leads to damage to the brain, a probable cause of

complications such as cerebral hemorrhage, and cerebral infarction which has often been observed to be a risk factor for vascular dementia (VaD) and Alzheimer's disease (AD) [14, 15]. These organs have one or a small number of arteries that perfuse them (rather than a network) and are therefore particularly vulnerable to occlusion of the arteries. An important feature of hypertension is that most cases, HBP presents no symptoms, only seen when the blood pressure is high enough to cause actual organ damage. Hypertension is common in the elderly population, and it is one of the most frequent comorbidities (alongside diabetes) in people with dementia [16].

1-2. Dyslipidemia

Dyslipidemia, which carries a risk of cardiovascular disease (CVD) based on elevated triglycerides (TG) and low high-density lipoprotein (HDL), is defined as an abnormal amount of lipids in the blood. It has been reported that the most important secondary cause of dyslipidemia in developed countries is a sedentary lifestyle with excessive dietary intake of saturated fat, cholesterol, and trans fats [17]. There is a decreased concentration of HDL-C as well as qualitative changes in low-density lipoprotein (LDL), notably the presence of small and dense LDL particles (Figure 1).

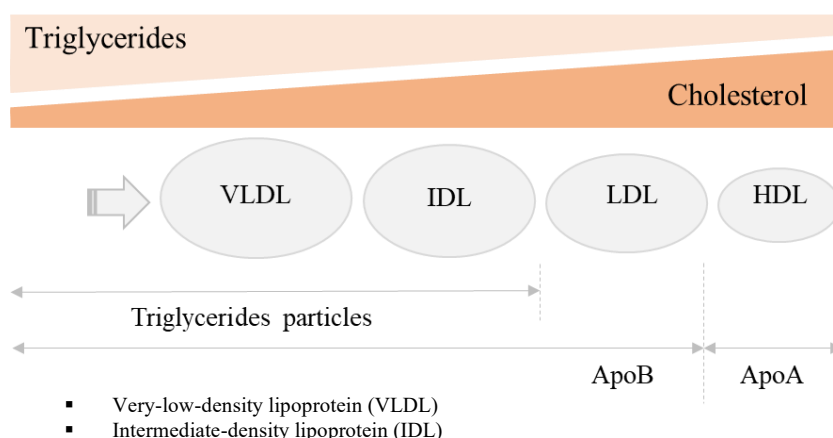


Figure 1 Proportion of triglycerides and cholesterol in lipoprotein particles [18]

It has been shown that obesity with dyslipidemia promotes the onset of CVD and this is strongly related to oxidative stress [19]. Low levels of circulating HDL, enhanced clearance of HDL particles, and elevated plasma very-low-density lipoprotein (VLDL) levels promote reactive oxygen species (ROS) generation in the endothelium [20]. It has been reported that ROS-mediated changes in lipid expression causes further oxidation-derived products, including oxidized LDL which can play a further critical role in CVD [21].

1-3. Hyperglycaemia

Hyperglycaemia (i.e. high amount of glucose in blood plasma) is a condition in which an excessive amount of glucose circulates in blood plasma, and its severe condition is a potential risk factor for CVD, particularly in people with type 2 diabetes, even when taking comorbidity into account such as renal and liver diseases, terminal cancer and cognitive decline [22]. Hyperglycaemia is confirmed when glucose reaches 80-130 milligrams per decilitre (mg/dL) before a meal, or less than 180 mg/dL around 1-2 hours after the beginning of the meal (American Diabetes Association) [23].

Hyperinsulinemia (i.e. high insulin levels in the blood) as a consequence of insulin resistance (IR) (i.e. the cells are resistant to the insulin and are unable to use it as effectively) is thought to increase blood pressure by enhancing sodium reabsorption through its effects on distal renal tubules [24]. Glucose and IR are considered to be comorbidities of obesity.

1-4. (Abdominal) Obesity

As one of the vascular health risk factors, obesity is evidently a major public health problem, leading to a variety of endocrine disturbances and influences on inflamm-aging [25]. The World Health Organization (WHO) reported that the worldwide prevalence of obesity nearly tripled between 1975 and 2016, and 11% of men and 15% of women of the world's adult (aged 18+ years) population were obese (body mass index, BMI ≥ 30 kg/m²) in 2016 [26]. The proportion of those categorized as obese is 26.9% in men and 26.8% in women in England in 2015; this increased from 13.2% in men and 16.4% in women in 1993 [27]. The overall prevalence of obesity among US adults was 39.8% in 2015–16; 42.8% in middle-aged (40–59) adults [28]. Obesity is a condition of abnormal or excess fat accumulation in adipose tissue and a disorder of energy regulations, to the extent that health may be impaired [29, 30]. It has been reported that a number of health related disturbances are strongly associated with obesity; it increases risks associated with the development of other cardio-metabolic risk factors such as coronary heart disease (CHD), type 2 diabetes, IR, hypertension, and hypertriglyceridemia (greatly increased blood levels of triglycerides), and in addition the risk of certain cancers and non-alcoholic fatty liver disease [30]. Visceral or abdominal obesity is thought to be central to metabolic syndrome which is associated with a pro-inflammatory state. Visceral obesity leads to the alteration of the normal physiological balance of adipokines, IR, endothelial dysfunction and a pro-atherogenic state [31]. Accurate identification and the early stage of stratification of these diseases are one of the greatest unmet medical needs.

1-5. The Framingham Risk Score for a cardiovascular disease assessment

The Framingham Risk Score was based on a US white middle-income community cohort (12 years of follow-up, 1971–74). It has been suggested that the tool may overestimate risk in countries with a low

CVD mortality rate (and vice versa). Exposures incorporated with the risk functions in the study were age, BP, smoking history, serum lipoproteins (total serum cholesterol (TC), LDL-C, HDL-C levels) and diabetes [32]. The Framingham Risk Score is considered as the comprehensive and effective measures to assess CVD risk (i.e. one of the cardio-metabolic dysfunctions) in the CARDIA Study population [33].

2. Cardio-metabolic dysfunctions and its impact on cognition

Hypertension and its impact on cognition

Long term higher levels of blood pressure are associated with eventual damage to blood vessels and the organs served by them including the brain. Hypertension is one of the most frequent *comorbidities* (alongside diabetes) in people with *dementia* [16]. It has been reported that hypertension is one of the risk factors of age-related white matter lesions (WML), and age-related WML are associated with increasing cognitive impairment and dementia [34]. Hypertension leads to damage to the brain, a probable cause of complications which often has been observed to be a risk factor for VaD and AD [14, 15].

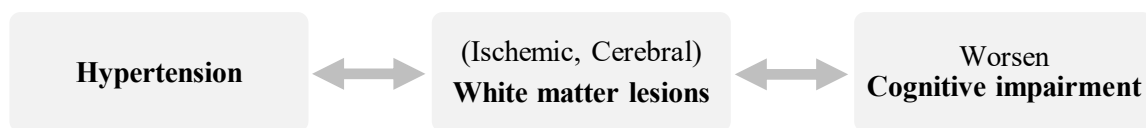


Figure 2 Possible mechanism of hypertension and cognitive impairment [35]

It has been reported that hypertension is associated with increasing cognitive impairment and VaD. Yano *et al.* (2004), examining the CARDIA Study at Year 25 (N=2,326), reported that long-term blood pressure variability (BPV) over 25 years from the baseline was associated with worse psychomotor speed and verbal memory tests in midlife, independent of cumulative exposure to BP during follow-up. The higher standard deviation of BP was associated with lower scores of digit symbol substitution test (DSST) (symbols) and rey auditory verbal learning test (RAVLT) (words), adjusting for age, sex, race, education, and site of collection [36]. The Whitehall II cohort study (N=8,639) indicated that SBP (≥ 130 mmHg) in midlife (aged 50) was significantly associated with an increased risk of dementia (N=385), adjusted HR (95% CIs): 1.38 (1.11, 1.70) [37].

Dyslipidaemia and its impact on cognition

A literature review and meta-analysis of the onset of AD pathology for people with high TC in mid-life reported the pooled relative risk (95% CIs): 2.14 (1.33, 3.44) [38-40]. Individual studies also showed high TC in midlife increased the risk of mild cognitive impairment (MCI) and cognitive decline in late life. The study underlined that there were insufficient data to examine other cholesterol sub-fractions, sex differences, or Apolipoprotein E (APOE) interactions [41].

Hyperglycaemia and its impact on cognition

A previous study reported that acute hyperglycaemia (blood glucose >15 mmol/l) is associated with mild cognitive impairment (i.e. significant slowing of all cognitive performance tests) and an increased number of mental subtraction errors for both type 1 (N=105, mean aged 37 years) and type 2 (N=36, mean aged 50 years) diabetic subjects [42]. Another study similarly reported that during acute hyperglycaemia, cognitive function was impaired and mood state deteriorated in type 2 diabetes subjects (N=22, median aged 62 years, BMI 29.8 kg/m² and Hemoglobin A1c 7.5%) [43].

Obesity and its impact on cognition

It has been reported that there is a strong association between obesity (and its comorbidities) and impaired cognitive function including reducing neural integrity (e.g. gray and white matter atrophy) in mid-life. Increasing age coupled with the negative metabolic consequences of obesity (e.g. type 2 diabetes mellitus) are likely to significantly contribute to cognitive decline and incidence of dementia; a 27 year longitudinal US population based study estimated that being obese at ages 40–45 years increases the risk of later dementia by 74% relative to those of a normal weight [44, 45]. Future studies need to reveal the linkage between adiposity and excess risk of dementia and AD. Adipocyte secreted proteins and inflammatory cytokines explain the association between obesity and increased risk of dementia. Late-life elevated BMI confers a lower risk of having dementia. The hormone leptin explained the mechanism for the reverse association [46].

Lifestyle factor and its impact on cognition

The Cognitive Function and Ageing Study Wales (CFAS-Wales) study (N=2,315, aged 65+ years) found that people with a healthier lifestyle (i.e. a healthy diet, more physical, social and mentally stimulating activities, moderate alcohol consumption, and refraining from smoking) had better mental fitness which was partly accounted for by their level of cognitive reserve. Despite the issue of the data being collected at one point in time, the finding implies earlier life experiences build cognitive reserve, which helps to maintain mental fitness in later life [47].

3. Systematic review and meta-analysis: coronary heart disease and myocardial infarction impacts on cognition

The systematic review and meta-analysis presented in this Section were conducted to understand how major forms of CVD have been identified to be associated with cognition in prospective epidemiological studies, and to investigate whether a CVD risk confers a higher risk of having cognitive impairment or dementia. It will be helpful to discover additional risk factors that might affect the associations.

Atherosclerosis is a disease in which plaque builds up inside arteries (i.e. blood vessels that carry oxygen-rich blood to the heart and other parts of the body). Plaque is made up of fat, cholesterol, calcium, and other substances. This plaque causes the arteries to harden and narrow over time, restricting the flow of oxygen-rich blood to organs and other parts of the body, and increasing the risk of blood clots that could potentially block the flow of blood to the heart or brain [48].

CHD is a disease in which these waxy plaques build up especially inside the coronary arteries. The coronary arteries supply oxygen-rich blood (directly) to the heart muscle [49]. CVD is a general term for conditions affecting the heart or blood vessels and types of CVD include CHD, cerebrovascular disease, stroke, peripheral arterial disease and aortic disease [50]. Coronary artery calcification (CAC) is a state of calcium build-up in the heart's arteries and a marker of atherosclerotic burden. Studies have reported that CHD and CAC are associated with an increased prevalence of WML [51, 52] and gray matter (GM) changes [53].

Figure 3 presents 10-year non-acute CVD Kaplan-Meier estimates using the UK population-based clinical practice research datalink (CPRD) (2001–16, N=325,736) [54]. The middle-aged population shows similar levels of risk as the elderly population, which emphasises the importance of risk management.

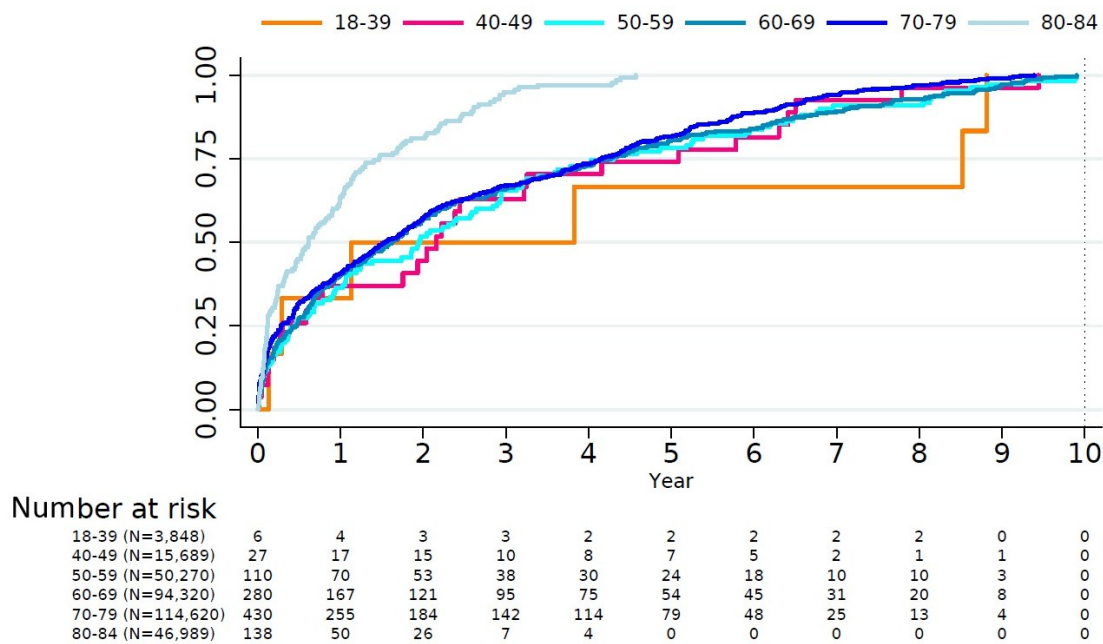


Figure 3 Estimation of 10-year risk of non-acute CVD using CPRD (2001-16) [54]

Research has shown that untreated CVD leads to or influences brain structures and cognitions including cognitive impairment and dementia [55]. A clear understanding of CVD-induced cognitive impairment is needed to research in detail, to understand the findings of previous studies have found and furthermore to research the underlying mechanisms. In the following systematic review and meta-analysis, cognitive impairment outcomes by subdivided CVD risks were examined in relation to the previous prospective studies. Investigations focused on whether there is evidence of CVDs and if so, whether they impact on cognitive impairment as in the previous studies.

In this section, the methodology of a systematic review for how CHD, MI, and AF impact on cognition or dementia is described. Electronic searches (made on 20th Feb 2019) were carried out using free terms, i.e. not medical subject headings (MeSH), in the databases. The papers were identified using the following search keywords in the databases:

((hypertension) OR (cerebral small vessel) OR (CSVD) OR (coronary microvascular) OR (CMD) OR (small artery disease) OR (coronary artery calcification) OR CAC OR (angina pectoris) OR AP OR (coronary heart disease) OR (myocardial infarction) OR (cardiovascular disease)) AND ((cognit) OR (dement*)) AND ((prospective) OR (cohort)) AND ((odd*) OR (hazard*) OR (effect))*

CVD types were subdivided by keywords, i.e. *(coronary heart disease) OR (CHD)*, *(myocardial infarction)*, *(atrial fibrillation)*, in any fields including the full texts.

In total, 5,575 papers remained after excluding duplications (N=1,339); medial database: PUBMED or Medline (N=3,278), Cochran Library (N=318), non-medical databases: CINAHL (N=765) and EMBASE (N=2,261), and social science databases: PsychINFO (N=292). Types of study included were prospective (ideal designs to test associations), retrospective, observational, case control, quasi-

experimental, mix method (quantitative and qualitative), randomized controlled trials (RCTs), controlled clinical trials (CCTs) and some open-label studies. Cross-sectional studies (which are not prospective), systematic review papers, and abstract based reports or journals (in an annual meeting/ congress/ conference) were excluded. Qualitative studies, studies with less than or equal to 100 participants, studies without risk outcomes of odd ratio (OR), hazard ratio (HR), effect size, coefficients and prevalence were excluded. The study population was adults aged 18 and over with no regional restriction. Cognitive related outcomes such as cognition, depression, Alzheimer's, Parkinson's were extracted, excluding death or mortality rates.

1-1. Cognitive decline or dementia on coronary heart disease

Keywords of *(coronary heart disease) OR (CHD)* in abstracts produced 1,373 papers. This was reduced to 882 papers after applying the inclusion/exclusion criteria (described above) and deleting duplications; title exclusion – *(presentation) or (session) or (congress) or (meeting) or (conference) or (abstract) or (poster)* (N=1,313); title exclusion – *(systematic review) or (meta-analysis) or (meta analysis) or (meta-analyses) or (review)* (N=1,262); inclusion keywords in any field (incl. the manuscript) – *(odd ratio) or (odds ratio) or (hazard ratio) or (hazards ratio) or (effect size)* (N=948); inclusion keywords in any field (incl. the manuscript) – *(prospective) or (retrospective) or (observational) or (case control) or (case-control) or (quasi-experimental) or (quasi experimental) or (quantitative) or (RCTs) or (CCTs)* (N=906); exclusion for duplicate manuscript (N=882).

Fifteen studies were included in the quantitative synthesis, and nine studies reported OR or HR on cognitive or dementia outcomes, which are presented in Figure 4. The six studies reporting coefficients or effect size on cognitive, or depression outcomes are not reported in this chapter. The median age of the study populations ranged between 58 and 93 years in the studies examining the impact of CHD. The Honolulu-Asia Aging study adjusted for all variables significantly associated with vascular dementia in a multivariate logistic regression model (i.e. age, education, hypertension, diabetes, western diet preference, use of vitamin E, and 1-hour postprandial glucose at examination) [56]. The Zutphen Study adjusted for age, education, and the baseline Mini-Mental State Examination (MMSE) score [57]. The adjusted odds or hazards ratios were presented (using the 'metan' package in Stata/IC 16.0, StataCorp LLC). The overall OR/HR (from both random and fixed effect models) suggested a **small effect size** (0.15–0.21%) of the impact (Figure 4) [56-64]; fixed-effect summary of OR is 1.32 (1.20, 1.44) and HR is 1.46 (1.25, 1.68), respectively [65].

CHD in Prospective cohort study

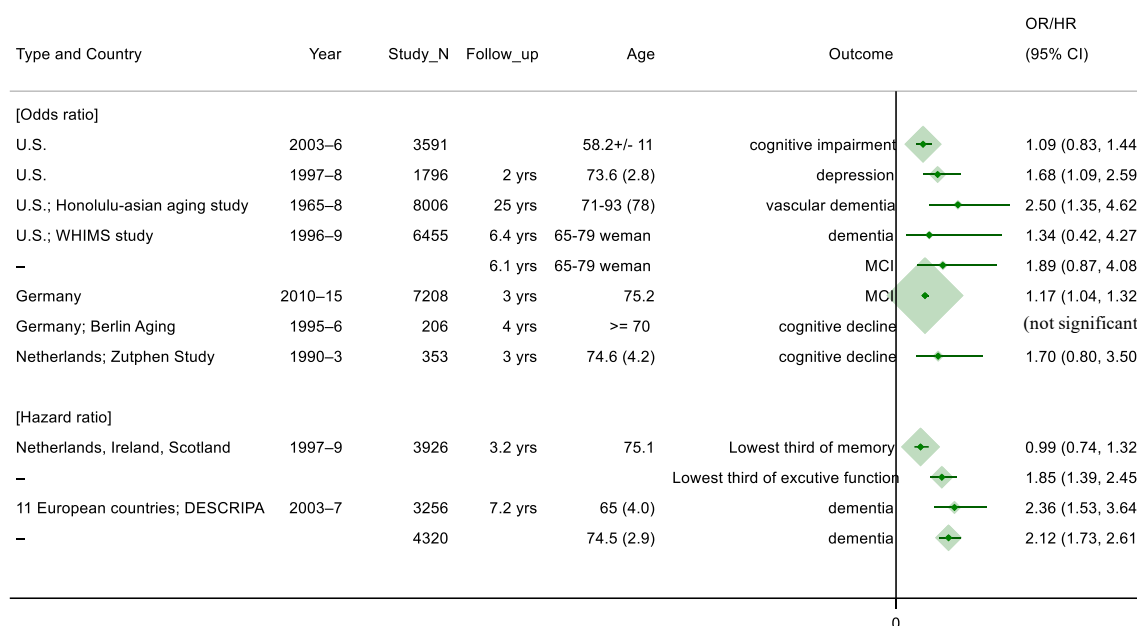


Figure 4 CHD and its impact on cognitive impairment or dementia in prospective cohort studies, OR or HR (95% CIs)

1-2. Cognitive decline or dementia on myocardial infarction

Keywords of (*myocardial infarction*) in abstracts produced 1,887 papers in a search. After applying the inclusion/exclusion criteria and deleting duplications, the remaining papers are 1,080. Twelve studies reported OR or HR on cognitive or dementia outcomes are presented in Figure 5.

The median age of the study populations ranged between 55 and 93 years in the studies examining the impact of MI. The studies adjusted other exposures (i.e. potential confounding factors); Women's Health Initiative Study (WHIMS) fully adjusted for age, education, race, hormone trial randomization assignment (HTR) arm, baseline modified mini-mental state (3MS) examination, alcohol consumption, smoking, physical activity, diabetes, sleep hours, hypertension, BMI, depression, waist-hip ratio, hypercholesterolemia, and aspirin use) [64]. The Cache County Study of Memory Health and Aging (CCSMHA) adjusted for age, sex, education, hypertension, high cholesterol, diabetes, obesity, stroke, coronary artery bypass graft surgery (CABG), and APOE genotype [66]. The Cardiovascular Health Study (CHS) adjusted for age at baseline, education, race, income, APOE genotype, and modified MMSE score at the time of the brain MRI [67]. The Rotterdam study adjusted for age, sex, SBP, DBP, BMI, AF, diabetes, current smoking, intima media thickness, total cholesterol, HDL, and APOE genotype [51]. The North Karelia Project and the Finnish contribution to the WHO Multinational MONItoring of Trends and Determinants in CArdiovascular Disease study (FINMONICA) (N=2,293, aged 65–79; 21 years follow-up) which adjusted for age, sex, education, smoking, alcohol consumption,

and APOE genotype had an OR (95% CIs): 2.5 (1.2, 5.4) for MI on AD or VaD (N=52) [55]. The nationwide population-based cohort study in Denmark, controlled for matching factors by study design and adjusted for extensive founders including age, sex, comorbidities, and socioeconomic status. However, the study acknowledged lack of information on smoking, physical activity and the APOE genotype, which are known to be associated with both MI and dementia [68-70].

The overall adjusted OR/HR suggested a **small effect size (0–0.2%)** of the impact (Figure 5) [51, 55, 64, 66, 67, 70-76]; fixed-effect summary of HR is 0.98 (0.96, 1.0), ORs are 1.38 (0.85, 1.9) and 1.43 (1.07, 1.79), respectively.

MI in Prosepective cohort study

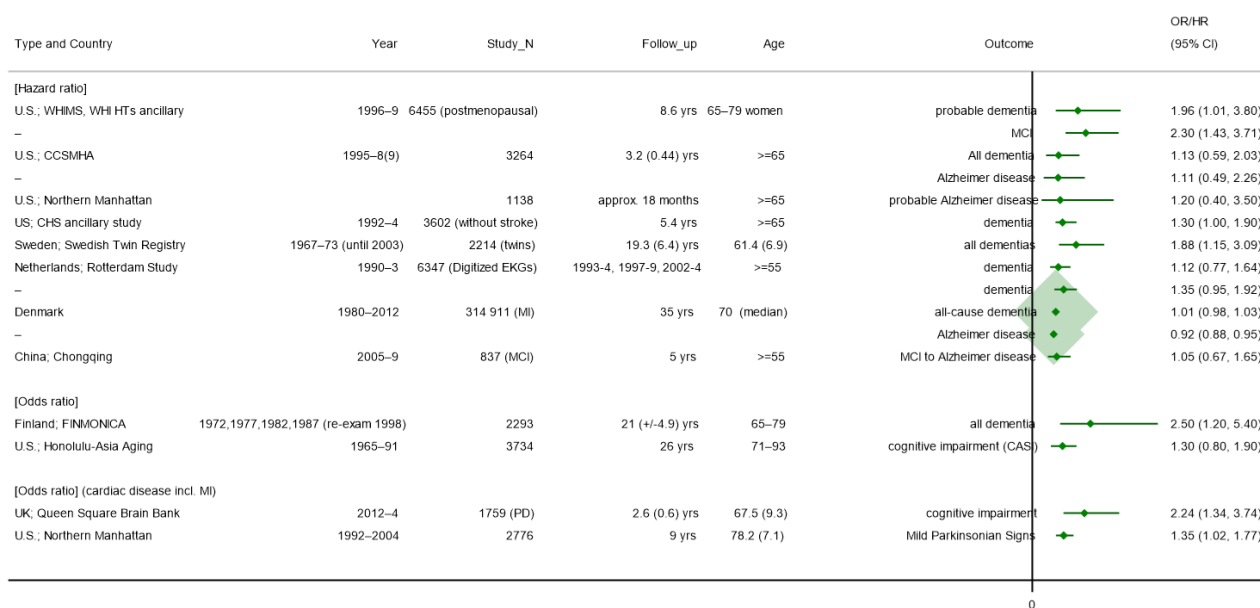


Figure 5 MI and its impact on dementia or MCI in prospective cohort studies, HR or OR (95% CIs)

Women's Health Initiative Study (WHIMS) and the nationwide population-based cohort study in Denmark adjusted for lifestyle factors such as smoking and physical activity [64]. The Honolulu-study reported mid-life smoking, alcohol consumption, serum cholesterol, serum uric acid, use of vitamin C, and presence of APOE ε4 were not significantly associated with VaD [56]. Although studies showed inconsistent results on the association of smoking with all-cause dementia and VaD, increased risk of dementia from smokers, and the increased risk of AD from smokers which is more pronounced in APOE ε4 noncarriers were reported in the meta-analysis of prospective cohort studies [77]. In addition, the Cardiovascular risk factors, Aging and Incidence of Dementia (CAIDE) study reported that leisure-time physical activity (at least twice a week) at midlife was associated with a reduced risk of dementia and AD, after adjustments for other risk factors (e.g. age, sex, education, follow-up time, locomotor disorders, APOE genotype, vascular disorders, smoking, and alcohol consumption) [78].

In summary, CHD and MI were significantly associated with an increased risk of cognitive impairment or dementia in prospective cohort studies. In addition, the overall adjusted odds/hazard ratio from 21 prospective cohort studies suggested a small effect size (0.21–0.28%) for the impact of AF on dementia or cognitive impairment (the outcome of AF is not reported in this Section).

I have conducted an evidence-based systematic review on CVD risks and cognitive impairment or dementia. Small effect sizes of CVD-induced cognitive impairment or dementia were observed in elderly individuals. The association between vascular disease and cognitive impairment may differ by confounders that affect CVD, and there was a variety of adjusted confounders among the studies. Nonetheless, the information on potential interaction between CVDs and cognitive impairment or dementia has been limited by the observational designs. The pathogenesis of AD is characterized by an uncontrolled inflammatory activation of microglial cells [79, 80]. The brain is highly susceptible to ROS and oxidative stress also seems to be involved in the pathogenesis of both AD and VaD [81, 82]. A study reported urinary excretion of oxidative DNA damage repair products (8-oxodG and 8-oxoGua) were higher in patients with AD or VaD (N=18; aged 53–84 years) than in the control group (N=33; matched eating habits, age and weight) [83]. Further detailed understanding of confounders and mediators of the association is important in the prevention and management of cognitive decline, including therapeutic treatments. Biological pathways of underlying atherosclerosis and common risk factors of those diseases will need to be further investigated to ascertain the link.

4. Objectives of the study

Dementia is considered to be a national socioeconomic problem, currently generating annual costs of more than £34 billion a year in the UK, and \$277 billion in the US health and social care (not including unpaid caregiving) [84, 85]. The association between vascular health and late-life lower cognitive function has been investigated. Nevertheless, there is a lack of evidence in terms of the role of underlying physiological mechanisms in the association of cardio-metabolic risk factors and brain health in middle-aged adults, to refine preventive approaches. This study focuses on the prognosis of cognitive reserve (defined as the resilience or plasticity of cognitive networks in the face of disruption) examining the physiological phenotype network underlying it (using the metabolites and brain data), which is important in terms of preventive medicine (Figure 6) [86].

This study hypothesizes that urinary metabolites are mediators regarding cognition on (given) a cardio-metabolic risk. The additional important hypothesis of the study is that a cardio-metabolic risk factor has a detrimental effect on the brain as measured by magnetic resonance imaging (MRI)-derived parameters (MRI parameters), which are markers of cognitive functions. This study also hypothesizes that the identified physiological mediators from the study populations can be influenced by individuals' exposure characteristics (e.g. clinical, lifestyle, socio-demographic and other health factors).

The primary aim of the study is to investigate whether the targeted urinary metabolites and the brain MRI parameters are prognostic markers (i.e. mediators) regarding cognition on a cardio-metabolic risk factor using the CARDIA cohort measured at Year 30. Significant exposures including clinical, lifestyle, and socio-demographic factors, and confounding factors were examined, and the information was used to explore the possibility of developing a risk assessment tool for brain health (i.e. cognition or dementia).

This study benefits from important clinical aspects, physiological markers, and demographic information associated with cognitive reserve. A routine electronic population-based primary care dataset provides a way of investigating some of these complicated factors. However, concern has been raised regarding whether the diseases are recorded well enough [87]. This study provided an opportunity to examine these complicated specific questions with the cross-sectional cohort studies which were longitudinally followed up in the US. CARDIA, with its longitudinal study design, will allow confounding factors and other exposures that might be associated with the relationship, to be taken into account. This study will focus on a cross-sectional analysis with the data measured at Year 30 with participants aged 48 to 61 in 2015–16. The Year 25 data (with the same participants aged 43 to 56 in 2010–11) is also available except for the urinary metabolites data.

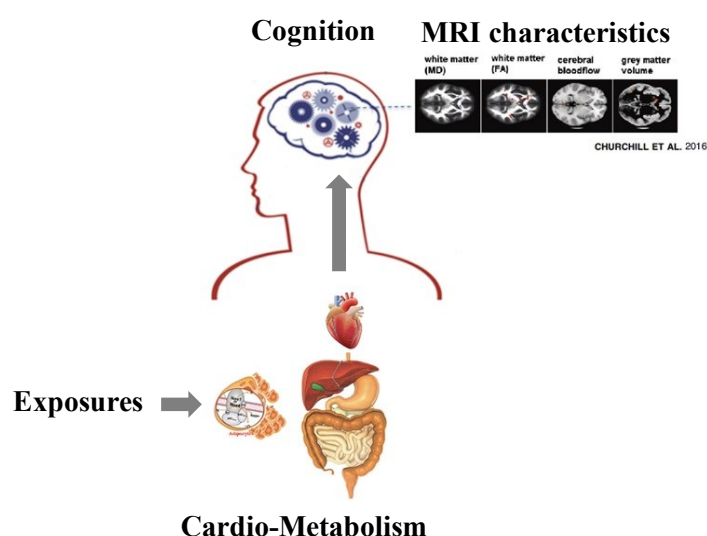


Figure 6 The associations of cardio-metabolism and brain health; investigation of the role of underlying physiological mechanisms

Specific objectives of the study are summarised in Figure 7 and detailed as follows:

Objective 1: To investigate the cognitive outcome (Y) on cardio-metabolic risk factors (X) measured at Year 30. Significant other exposures and confounding factors which are causing spurious significances of cardio-metabolic risk factors regarding cognition will be examined.

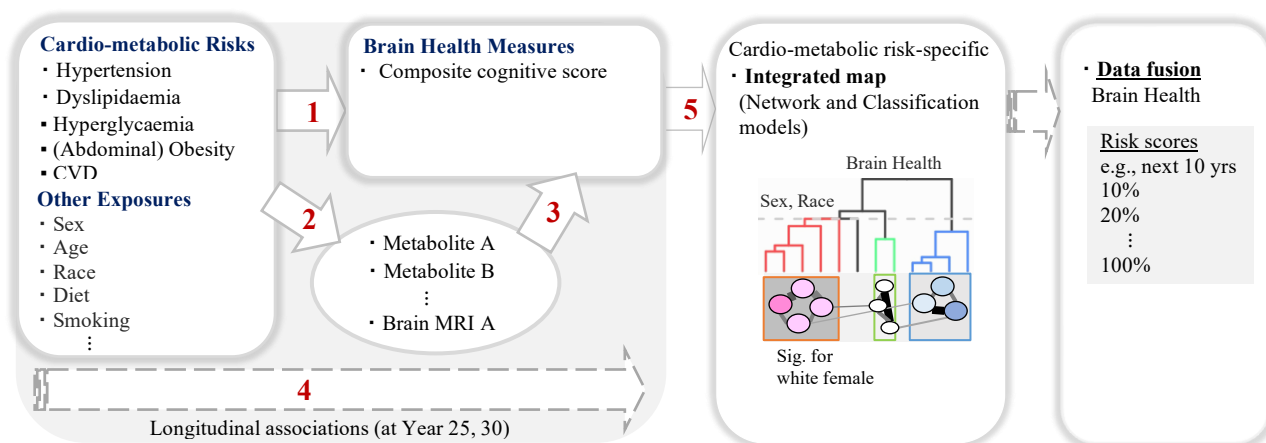
Objective 2: To investigate if urinary metabolites and/or brain MRI parameters (M) measured at Year 30 are significantly associated with cardio-metabolic risk factors (X).

Objective 3: To investigate if each cardio-metabolic risk factor (X) and urinary metabolites or the brain MRI parameters (M), which are significant in objective 2, are significant regarding cognitive outcome (Y), adjusting for other exposures and confounding factors.

Objective 4: To investigate the degeneration of the cognitive outcome (Y) on the brain MRI parameter (X) per cardio-metabolic risk factor using longitudinal (e.g. Year 25 and Year 30) data.

Objective 5: To provide a cardio-metabolic risk-specific integrated map (i.e. profile of the markers) with regard to cognition with the significant other exposures, confounding factors, and mediators which were identified from Objectives 1 to 4.

Lastly, the development of a novel risk assessment tool of brain health (i.e. cognition or dementia) can be examined. Information from the network-based approaches can be integrated. Decision level data fusion constructs separate prognostic models for each data set (e.g. men and women) and integrates results.



Note: Objective 5. In the network cluster enrichment map, each box represents a cluster of metabolites. Edges are links between nodes and indicate the relationship. Node colours indicate a risk factor (e.g. gender) effect. The box background colours represent the enrichment p-value of the cluster.

Figure 7 Schematic of the study objectives and synthesis methodology

Chapter II. Methods

1. Study design and population

1-1. The CARDIA Study

The CARDIA Study: baseline to Year 30

CARDIA is a prospective cohort study of the development and risk factors of cardio-metabolic disease in black and white young adults since 1985–86. The study population consisted of initial 5,115 participants who were recruited from four US metropolitan areas (Birmingham, AL; Chicago, IL; Minneapolis, MN; Oakland, CA); enrolment was balanced by age (18–24 years, or 25–30 years), sex (male/female), race (African-/European-American) and education ($\leq$ high school/ $>$ high school). Follow-up examinations were conducted in 1987–88 (Year 2), 1990–91 (Year 5), 1992–93 (Year 7), 1995–96 (Year 10), 2000–01 (Year 15), 2005–06 (Year 20), 2010–11 (Year 25) and 2015–16 (Year 30) (with retention rates 91%, 86%, 81%, 79%, 74%, 72% and 72%) (Figure 8) [88].

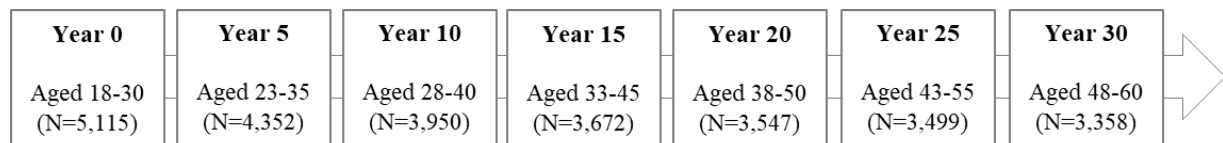


Figure 8 Study design and number of participants for the CARDIA Study [88]

The study population and data at Year 25 and Year 30

In total, 600 participants with high-resolution nuclear magnetic resonance (NMR) spectroscopy data were available in the present study, with 596 completed cognitive tests such as DSST. In addition, 599 participants had data from liquid-chromatography mass spectrometry (LC–MS) of urine and 281 had complete brain MRI data at Year 30 (2015–16) were available. In addition, the same 579 participants within participants at Year 30, and the brain MRI data for 185–193 participants were available at Year 25 (2010–11); N=185 for measurement of cerebral blood flow (CBF) and other measurements are N=193 (Appendix Table 7). Urinary metabolites were not available at Year 25 (Figure 9).

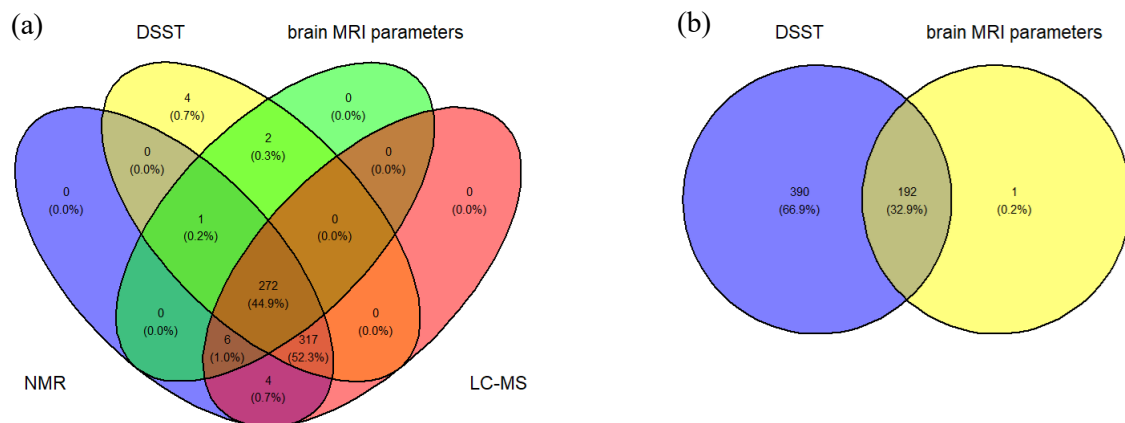


Figure 9 A Venn diagram of participants in (a) Year 30 (lavender blue: NMR, light yellow: DSST, yellow-green: brain MRI parameters, light red: LC-MS) and (b) Year 25 (lavender blue: DSST, light yellow: brain MRI parameters)

To analyse the multiple regression of Objective 1, expecting a small effect size ($f^2 = 0.06$) from Year 25 CARDIA³, a sample size of 270 is required for 9 variables (e.g. centre, education, physical activity, smoking, diet, medication usage, and age, sex, race as a default adjustment) to achieve 80% power at alpha 0.05. To detect the mediated effect for Objective 2, empirical sample sizes 562 are required for small effect sizes (i.e. τ' , α and β are 0.14) to achieve 80% power in Baron and Kenny's (1986) test; for $\tau' = 0.14$, α and $\beta = 0.26$ –0.59, empirical sample sizes 445–414 are required (Fritz *et al.*, 2007) [89, 90].

Cardio-metabolic risk factors

Cardio-metabolic risk factors of the study include measures of blood pressure (SBP/DBP), lipids (HDL-C, LDL-C, and triglycerides), high blood glucose or sugar (glucose, homeostatic model assessment for insulin resistance (HOMA-IR), estimated glomerular filtration rate (eGFR), and insulin), adiposity (BMI and WC), and CVD (10-Year Framingham risk).

Other exposures and confounders

Lifestyle exposures included were physical activity, (summary) diet score, smoking and alcohol consumption. Socio-demographic exposures are sex, age, race, centre, education up to Year 30 (2015–16), and family income. Health information such as depression score, medical/family history of dementia, sleep quality, use of medication (for antihypertensive treatment (AHT), HBP, cholesterol, and depression), and serum and urinary creatinine (mg/dL) were included. In addition, menopausal

³ f^2 (F test of the multiple R-squared) from Year 25 CARDIA were -0.06, -0.06, -0.07 for each BMI < 25 kg/m², hypertension, diabetes, adjusted for age, sex and race. Sample sizes of the outcome (e.g. each total brain volume (TBV), abnormal white matter (AWM), white matter fractional anisotropy (WM-FA)) were 680 (Launer *et al.*, 2005).

status, which is an important factor for body fat distribution and incidence of CVD and hypertension in women, were examined.

Outcome measures: cognition scores

The three cognitive function tests reported in the CARDIA Study at Year 25 and Year 30 are RAVLT, DSST and the Stroop test. In addition, the Montreal Cognitive Assessment (MoCA) is a brief cognitive screening tool with high sensitivity and specificity for detecting MCI was measured at Year 30 [91] (documented at <https://www.cardia.dopm.uab.edu/exam-materials2/data-collection-forms/year-30>).

Higher scores on the RAVLT, DSST and MoCA reflect a better cognitive test, on the other hand, higher scores on STROOP reflect poorer cognitive test performance. The z-scores from RAVLT, DSST, reversed Stroop, and MoCA were generated for the analyses. A composite cognitive score is comprised of z-scores. The detailed information is described in Chapter II, Sub-section 1-4.

Prognostic markers: urinary metabolites at Year 30

Targeted urinary metabolic data was acquired at Year 30 by NMR (N=600) and LC-MS (N=599). Up to 300 participants from Chicago, IL and up to 100 participants from each of the other three field centres (Birmingham, AL; Minneapolis, MN; Oakland, CA) were recruited from urinary analysis at Year 30. The multiplatform metabolite profiling method is ideal to achieve comprehensive coverage of the human metabolic phenome. A targeted metabolic profiling selects analytes based on the hypothesis; it optimizes and verifies real-world samples to support or refute the hypothesis [92]. Detailed information with respect to urinary data including data processing is described in Chapter II, Section 2.

Prognostic markers: brain MRI parameters

Multiple MRI sequences can be used to measure a variety of variables associated with brain structure, physiology and function, and molecular composition of tissues [93]. Brain MRI measurements were performed in three field centres (Birmingham, AL; Minneapolis, MN; Oakland, CA). The overlap with Year 25 is a function of those who attended both exams and who among those, had both MRI data at Year 25 and Year 30. One of the key themes in the present study is investigating brain MRI parameters in addition to cognition measurements, as indicators of brain health in middle-aged community-based adult. As previously described, the present study hypothesised that higher cardio-metabolic risks have detrimental effects on ‘neurocognitive’ pathology which can be derived from the MRI measurements, i.e. higher cardio-metabolic risks are associated with an MRI profile reflective of a relatively higher load of macro and microstructural alterations. For instance, investigation of microstructural alterations of the brain white matter (WM) are common in brain pathology [94]. In this study, I investigated multiple brain regions and fluid including total brain, GM, WM, cerebrospinal fluid, and the brain

structures for the successive stages of memory processing (i.e. hippocampus, medial frontal cortex, and anterior cingulate gyrus) and early markers of AD (i.e. the (medial) temporal lobe volume and entorhinal area) in terms of volumes (cc³), abnormal tissue (load) volumes (cc³), microstructural integrity measured using fractional anisotropy (FA), CBF, the number of activated voxels (three-dimensional pixels of the brain divided by functional MRI machines, about five cubic millimetres in size each) and percent change of vascular reactivity (the responsiveness of a blood vessel to a specific stimulus). The brain MRI parameters and analyses are presented in Chapter II, Sub-section 1-5.

1-2. Demographic information at Year 30

Standardized questionnaires were used to obtain the demographic information for the CARDIA participants at each examination (documented at <https://www.cardia.dopm.uab.edu/exam-materials2/data-collection-forms/year-30>).

The quantified physical activity (reported as exercise units-year) was assessed based on the CARDIA Physical Activity History Questionnaire (Pereira *et al.*, 1997), an interviewer-based self-report of duration and intensity of participation in 13 categories of exercise over the previous 12 months [95, 96]. There were significant differences in the median alcohol consumption per day (mean ml) and physical activity at Year 30 by race and sex ($p < 0.001$, Kruskal-Wallis equality-of-populations rank test), respectively; physical activity was significantly higher in men (Appendix Table 1).

The diet quality score ranged from -16 to 21 points (higher scores indicate healthier choices). A diet quality score consists of cereal fibre consumption (grams per day) and servings of fruits and/or vegetables, nuts, sugar sweetened beverages, and red and/or processed meats (servings per day) [97]. Diet quality score differed significantly for females by race ($p < 0.001$, two-sample Wilcoxon rank-sum test) (Table 1).

The center for epidemiologic studies-depression (CES-D) score which is a 20-item scale, with a score range of 0–60 (higher scores reflect increased severity of depressive symptoms), was used in the CARDIA Study. A score of 16 or greater identifies individuals at risk for clinical depression [98]; black ethnicity had significantly higher risk for clinical depression (57/276) than white ethnicity (44/330) ($p=0.021$, Fisher's exact test).

The 12-item short form health survey (SF-12) consists of mental and physical functioning and overall health-related-quality of life. The self-reported responses to the 12 questions were summarized through standardized scoring into two summary scores the mental component score (MCS) and the physical component score (PCS) from Year 15 to Year 30 [99]. Higher scores representing better self-reported health, and the patterns of the past Y15, Y20, Y25 and Y30 are presented in Appendix Figure 1 and Appendix Table 2 presents the SF-12 at Years 15 to 30 by race and sex. The CARDIA participants reported that their physical functioning slowly decreased over time whereas mental functioning slowly increased over time up to Year 30 (medians).

Table 1 Lifestyle exposures at Year 30 (males and females by race)

Total		Mean (SD)	Black Male, Mean (SD)	White Male, Mean (SD)	Black Female, Mean (SD)	White Female, Mean (SD)
Alcohol consumption per day (mean ml)††	600	11.4 (17.7)	114 14.6 (21.4)	155 16.3 (19.6)	157 5.4 (10.8)	174 10.5 (16.7)
Physical activity exercise units †††	601	341.6 (287.2)	114 318.1 (280)	155 470.2 (311.2)	157 215.7 (235)	175 356 (262.2)
Diet quality score (v2) ††	598	2.6 (5.8)	113 1 (5.4)	153 2.5 (5.5)	157 1.6 (5.7)	175 4.7 (5.8)
CES-D score*	598	8.2 (7.5)	114 9.3 (7.8)	155 7.1 (5.9)	155 9.4 (8.8)	174 7.3 (6.9)
SF-12 Physical Components ††	603	50.2 (8.6)	116 49.2 (8.7)	155 52 (6.8)	158 47.7 (9.6)	174 51.4 (8.3)
SF-12 Mental Components	603	52 (8.8)	116 52.4 (8.5)	155 52.6 (7.5)	158 51.7 (10.2)	174 51.6 (8.7)
		Total, N(%)	Black Male, N(%)	White Male, N(%)	Black Female, N(%)	White Female, N(%)
My sleep was restless	Rarely/none	238 (39.4)	49 (42.2)	66 (42.6)	60 (38)	63 (36)
	Some	268 (44.4)	51 (44)	71 (45.8)	61 (38.6)	85 (48.6)
	Much	65 (10.8)	11 (9.5)	14 (9)	21 (13.3)	19 (10.9)
	Most/all	33 (5.5)	5 (4.3)	4 (2.6)	16 (10.1)	8 (4.6)
	Total	604 (100)	116 (100)	155 (100)	158 (100)	175 (100)
Participate in the PRF and Glucose Tolerance in Women Study	Not restrict carbs intake due to medical conditions ¹⁾	175 (78.8)			77 (82.8)	98 (76)
	Oral glucose tolerance test not done	47 (21.2)			16 (17.2)	31 (24)
	Total	222 (100)			93 (100)	129 (100)
On weight reducing diet because of disease ²⁾	No	117 (79.6)	0 (0)	33 (80.5)	30 (76.9)	54 (85.7)
	Yes	30 (20.4)	4 (100)	8 (19.5)	9 (23.1)	9 (14.3)
	Total	147 (100)	4 (100)	41 (100)	39 (100)	63 (100)
Family income	[1] < \$5,000	22 (3.6)	6 (5.1)	3 (1.9)	10 (6.3)	3 (1.7)
	[2] \$5,000 - \$11,999	26 (4.3)	9 (7.6)	3 (1.9)	10 (6.3)	4 (2.3)
	[3] \$12,000 - \$15,999	33 (5.5)	9 (7.6)	6 (3.9)	16 (10.1)	2 (1.1)
	[4] \$16,000 - \$24,999	29 (4.8)	11 (9.3)	3 (1.9)	9 (5.7)	6 (3.4)
	[5] \$25,000 - \$34,999	44 (7.3)	14 (11.9)	5 (3.2)	18 (11.4)	7 (4)
	[6] \$35,000 - \$49,999	44 (7.3)	11 (9.3)	5 (3.2)	16 (10.1)	12 (6.9)
	[7] \$50,000 - \$74,999	86 (14.2)	15 (12.7)	20 (12.9)	28 (17.7)	23 (13.1)
	[8] \$75,000 - \$99,999	89 (14.7)	18 (15.3)	22 (14.2)	22 (13.9)	27 (15.4)
	[9] ≥ \$100,000	227 (37.5)	23 (19.5)	88 (56.8)	27 (17.1)	89 (50.9)
	Don't know	1 (0.2)	1 (0.9)	0 (0)	0 (0)	0 (0)
	No response	5 (0.8)	1 (0.9)	0 (0)	2 (1.3)	2 (1.1)
Total		606 (100)	118 (100)	155 (100)	158 (100)	175 (100)

1) Not including low-carbohydrate diet

2) For participants have been on a weight reducing diet during the past five years (N=147)

* Center for epidemiologic studies-depression (CES-D) for males by race ($p=0.025$)†† $p < 0.001$ for females by race††† $p < 0.001$ for males and females by race, respectively

The educational attainment (maximum years up to Year 30) was significantly lower in black participants (mean 15 years; SD 2.4) than white participants (mean 17 years; SD 2.5) ($p < 0.001$, Kruskal-Wallis equality-of-populations rank test). In addition, smoking status was significantly different by race, e.g. current smoking in black (58/273) and white (19/327) ($p < 0.001$, Fisher's exact test) (Appendix Table 3).

Table 2 Socio-demographic exposures at Year 30 (males and females by race)

		Total, Mean (SD)	Black Male, Mean (SD)	White Male, Mean (SD)	Black Female, Mean (SD)	White Female, Mean (SD)
Age †		606 55.2 (3.5)	54.1 (3.6)	55.8 (3.1)	54.7 (3.8)	55.9 (3.3)
		Total, N(%)	Black Male, N(%)	White Male, N(%)	Black Female, N(%)	White Female, N(%)
Centre	Birmingham	97 (16.0)	25 (21.2)	14 (9)	34 (21.5)	24 (13.7)
	Chicago	300 (49.5)	62 (52.5)	80 (51.6)	74 (46.8)	84 (48)
	Minneapolis	108 (17.8)	16 (13.6)	31 (20)	25 (15.8)	36 (20.6)
	Oakland	101 (16.7)	15 (12.7)	30 (19.4)	25 (15.8)	31 (17.7)
	Total	606 (100)	118 (100)	155 (100)	158 (100)	175 (100)
Educational attainment maximum years up to Y30 †††	<12	10 (1.7)	5 (4.2)	0 (0)	4 (2.5)	1 (0.6)
	12–<16	241 (39.8)	69 (58.5)	45 (29)	85 (53.8)	42 (24)
	16+	355 (58.7)	44 (37.3)	110 (71)	69 (43.7)	132 (75.4)
	Total	606 (100)	118 (100)	155 (100)	158 (100)	175 (100)
Smoking status †††	Never	387 (64.5)	65 (55.6)	112 (73.2)	104 (66.7)	106 (60.9)
	Former	136 (22.7)	20 (17.1)	33 (21.6)	26 (16.7)	57 (32.8)
	Current	77 (12.8)	32 (27.4)	8 (5.2)	26 (16.7)	11 (6.3)
	Total	600 (100)	117 (100)	153 (100)	156 (100)	174 (100)
		Total, N(%)				
Have a usual source of medical care	No	33 (5.5)	9 (7.6)	10 (6.5)	6 (3.8)	8 (4.6)
	Yes	573 (94.6)	109 (92.4)	145 (93.6)	152 (96.2)	167 (95.4)
	Total	606 (100)	118 (100)	155 (100)	158 (100)	175 (100)

† $p < 0.001$ for males by race††† $p < 0.001$ for males and females by race, respectively

Antihypertensive treatments (AHTs) basically seek to lower blood pressure to below 140/90mmHg to prevent complications. There are different types of AHTs to achieve oligemia (decreasing volume of circulating blood), angiectasia (lowering gastrointestinal bleeding), and negative chronotropic (decreasing heart rate). It is considered that the use of antihypertensives to lower blood homoeotic will last only one day; and thus, it requires continuous drug treatment. Many of these drugs reach a steady state in the blood stream within a few doses. However, their effects on blood pressure take weeks to occur; it is believed that there is some re-setting of the body's internal homeostatic systems [57].

Currently, there is no comorbidity-specific hypertension guideline and GPs in the UK treat hypertension in people with comorbidities in the same way they do for people without them.

Allowing multiple counts per individual in the CARDIA Study (N=603), AHTs include angiotensin converting enzyme (ACE) inhibitors (lisinopril N=351, enalapril N=7), alpha-blockers (tamsulosin N=7), angiotensin II receptor blockers (ARBs) (losartan N=7), calcium channel blockers (amlodipine N=10) and diuretics (hydrochlorothiazide N=21, furosemide N=11, hydrochlorothiazide N=5, spironolactone N=4) and relatively older antihypertensive agents such as beta-blockers (carvedilol N=10, metoprolol N=7, atenolol N=6) [56]. Besides, AHTs included centrally acting alpha-agonist (clonidine N=3), and nitrate (isosorbide mononitrate N=1). In the study, those currently taking medication for HBP differed significantly for males and females by race, respectively ($p < 0.001$, Fisher's exact test). In addition, those currently taking medicine for diabetes differed significantly for males by race ($p < 0.001$, Fisher's exact test) (Table 3).

Table 3 Clinical exposures at Year 30 (males and females by race)

		Total, N(%)	Black Male, N(%)	White Male, N(%)	Black Female, N(%)	White Female, N(%)
Immediate family member diagnosed with dementia? ††	No	461 (77.4)	102 (89.5)	121 (78.1)	127 (81.9)	111 (64.5)
	Yes	135 (22.7)	12 (10.5)	34 (21.9)	28 (18.1)	61 (35.5)
	Total	596 (100)	114 (100)	155 (100)	155 (100)	172 (100)
Family member diagnosed with dementia before 65 years? (for immediate family member diagnosed with dementia)	No	124 (91.9)	9 (75)	32 (94.1)	27 (96.4)	56 (91.8)
	Yes	11 (8.2)	3 (25)	2 (5.9)	1 (3.6)	5 (8.2)
	Total	135 (100)	12 (100)	34 (100)	28 (100)	61 (100)
Gone through menopause, past 12 months	No	60 (18.0)			36 (22.8)	24 (13.7)
	Yes	229 (68.8)			98 (62)	131 (74.9)
	Don't Know	44 (13.2)			24 (15.2)	20 (11.4)
	Total	333 (100)			158 (100)	175 (100)
Kidney urine infection? (for participants with kidney problems)	Yes	13 (24.1)	1 (11.1)	1 (5.3)	0 (0)	11 (57.9)
	Don't Know	2 (3.7)	1 (11.1)	1 (5.3)	0 (0)	0 (0)
	NA	39 (72.2)	7 (77.8)	17 (89.5)	7 (100)	8 (42.1)
	Total	54 (100)	9 (100)	19 (100)	7 (100)	19 (100)
Currently taking medications for HBP †††	No	423 (69.9)	66 (55.9)	123 (79.4)	85 (54.1)	149 (85.1)
	Yes	181 (29.9)	52 (44.1)	31 (20)	72 (45.9)	26 (14.9)
	Don't Know	1 (0.2)	0 (0)	1 (0.7)	0 (0)	0 (0)
	Total	605 (100)	118 (100)	155 (100)	157 (100)	175 (100)
Currently taking medications to lower choles	No	473 (78.2)	84 (71.2)	127 (81.9)	115 (73.3)	147 (84)
	Yes	125 (20.7)	33 (28)	24 (15.5)	41 (26.1)	27 (15.4)
	Don't Know	7 (1.2)	1 (0.9)	4 (2.6)	1 (0.6)	1 (0.6)
	Total	605 (100)	118 (100)	155 (100)	157 (100)	175 (100)
Currently taking medications for heart condition*	No	574 (94.9)	104 (88.1)	152 (98.1)	148 (94.3)	170 (97.1)
	Yes	30 (5.0)	14 (11.9)	3 (1.9)	9 (5.7)	4 (2.3)
	Don't Know	1 (0.2)	0 (0)	0 (0)	0 (0)	1 (0.6)
	Total	605 (100)	118 (100)	155 (100)	157 (100)	175 (100)
Currently taking medications for asthma	No	545 (90.1)	106 (89.8)	149 (96.1)	131 (83.4)	159 (90.9)
	Yes	60 (9.9)	12 (10.2)	6 (3.9)	26 (16.6)	16 (9.1)
	Total	605 (100)	118 (100)	155 (100)	157 (100)	175 (100)
Currently taking medications for depression	No	531 (87.8)	110 (93.2)	142 (91.6)	129 (82.2)	150 (85.7)
	Yes	73 (12.1)	8 (6.8)	12 (7.7)	28 (17.8)	25 (14.3)
	Don't Know	1 (0.2)	0 (0)	1 (0.7)	0 (0)	0 (0)
	Total	605 (100)	118 (100)	155 (100)	157 (100)	175 (100)
Currently taking medications for diabetes †	No	546 (90.3)	100 (84.8)	149 (96.1)	131 (83.4)	166 (94.9)
	Yes	58 (9.6)	18 (15.3)	5 (3.2)	26 (16.6)	9 (5.1)
	Don't Know	1 (0.2)	0 (0)	1 (0.7)	0 (0)	0 (0)
	Total	605 (100)	118 (100)	155 (100)	157 (100)	175 (100)
Currently taking any other medications	No	370 (61.2)	87 (73.7)	104 (67.1)	91 (58)	88 (50.3)
	Yes	232 (38.4)	30 (25.4)	49 (31.6)	66 (42)	87 (49.7)
	Don't Know	3 (0.5)	1 (0.9)	2 (1.3)	0 (0)	0 (0)
	Total	605 (100)	118 (100)	155 (100)	157 (100)	175 (100)
Gone through menopause, past 12 months	No	60 (18.0)			36 (22.8)	24 (13.7)
	Yes	229 (68.8)			98 (62)	131 (74.9)
	Don't Know	44 (13.2)			24 (15.2)	20 (11.4)
	Total	333 (100)			158 (100)	175 (100)
	Total	Mean (SD)	Black Male, Mean (SD)	White Male, Mean (SD)	Black Female, Mean (SD)	White Female, Mean (SD)
Serum Creatinine (mg/dL) †††	601	0.9 (0.2)	1.1 (0.3)	1 (0.1)	0.8 (0.2)	0.7 (0.2)
Urinary Creatinine (mg/dL) †††	606	124.6 (84.6)	183.3 (107.1)	122 (73.2)	129 (70.2)	83.4 (62.8)

* For males by race ($p=0.002$)† $p < 0.001$ for males by race†† $p < 0.001$ for females by race††† $p < 0.001$ for males and females by race, respectively

1-3. Cardio-metabolic dysfunctions at Year 30

Cardio-metabolic risk factors include measures of blood pressure, lipids, high blood glucose or sugar, adiposity and CVD at Year 30 are presented in Table 4.

Table 4 Cardio-metabolic risk factors at Year 30 (males and females by race)

		Total	Mean (SD)	Black Male	Mean (SD)	White Male	Mean (SD)	Black Female	Mean (SD)	White Female	Mean (SD)
Hypertension	SBP ††	606	119.3 (16.1)	118	123.8 (13.5)	155	118.4 (13)	158	125.4 (19.1)	175	111.6 (13.8)
	DBP ††	606	72.8 (11.4)	118	76.2 (10.3)	155	72.5 (9.3)	158	76.9 (11.1)	175	67.1 (11.6)
Dyslipidaemia	HDL-C	601	60.2 (20.5)	116	52.5 (14)	155	49.2 (11.9)	155	67.1 (26.1)	175	68.9 (17.9)
	LDL-C	593	110.6 (32.7)	114	106.3 (35.1)	151	113.6 (29.9)	154	109.9 (35.1)	174	111.5 (31)
	Triglycerides (mg/dL)	599*	105.8 (66.7)	115	106.1 (64.8)	154	127.7 (76.9)	155	98.5 (64.7)	175	92.8 (54.6)
Hyperglycaemia	Fasting glucose (mg/dL)	601	100.5 (25.2)	116	107.7 (39.2)	155	102.7 (22.4)	155	99.5 (22.3)	175	94.7 (15)
	HOMA-IR ††	593	3.1 (2.8)	112	3.4 (2.7)	153	3.5 (3.5)	154	3.4 (2.8)	174	2.2 (1.9)
	eGFR ††	596	91.3 (15.5)	113	92 (19)	153	87.6 (11.1)	155	96.1 (18.1)	175	89.9 (12.6)
	Fasting insulin ††	598	11.9 (8.7)	115	12.8 (9)	155	12.8 (9)	154	13.5 (9.6)	174	9.1 (6.6)
(Abdominal) Obesity	BMI ††	605	29.4 (6.2)	117	30.3 (5.8)	155	29.1 (4.6)	158	32.2 (6.9)	175	26.6 (5.8)
	WC ††	605	94.6 (15.2)	118	100.8 (14.7)	155	99.8 (12.6)	158	95 (14.1)	174	85.2 (14)
CVD	10-Year Framingham risk, male	261	7.7 (4.7)	112	7.8 (5.4)	149	7.7 (4.1)				
	10-Year Framingham risk, female ††	323	4.7 (3.7)					153	5.8 (4.5)	170	3.7 (2.4)

* Omitted two distinctive observations of triglycerides > 600 (2366 and 3234 mg/dL)

†† $p < 0.001$ for females by race

The risk thresholds of the clinical conditions from CARDIA data was defined in Table 5 (documented at <https://www.cardia.dopm.uab.edu/study-information/derived-variables-from-cardia-data>).

SBP category ≥ 140 mmHg and/or DBP ≥ 90 mmHg [19] were used as a definition of stage 1 hypertension and above. Resistant hypertension was defined as on ≥ 4 antihypertensive medications [21]. For lipid metabolism (the synthesis and degradation of lipids in cells), hypertriglyceridemia was defined as triglycerides category > 200 mg/dL. LDL-C category > 160 mg/dL were defined as high LDL-cholesterolemia and HDL-C category < 35 mg/dL for male and < 45 mg/dL for female were defined as low HDL-cholesterolemia.

A fasting plasma glucose result of ≥ 100 mg/dL (or ≥ 5.5 mmol/l) were defined as being at higher risk of developing type 2 diabetes and hyperglycaemia. A fasting plasma glucose result of ≥ 100 (or 110) mg/dL were defined as prediabetes. Obesity was defined as BMI ≥ 30 kg/m². In addition, WC > 102 cm/ 40 inches for men and > 88 cm/ 35 inches for women [28] and waist to hip ratio (> 0.9 for men and > 0.85 for women [84] were used as reference information. Cholesterol-based 10-Year Framingham

Risk points 10 for male and 8 for female (based on age-band 55–59 which is similar to mean (SD) age of the present cohort; 55.2 (3.5)) were used as the multivariable risk of CVD stratification (Table 5) [100].

Table 5 Cardio-metabolic risk factors by risk threshold at Year 30 (males and females by race)

			Total, N(%)	Black Male, N(%)	White Male, N(%)	Black Female, N(%)	White Female, N(%)
Hypertension	SBP ††	< 140	545 (89.9)	106 (89.8)	145 (93.6)	125 (79.1)	169 (96.6)
		140+	61 (10.1)	12 (10.2)	10 (6.5)	33 (20.9)	6 (3.4)
		Total	606 (100)	118 (100)	155 (100)	158 (100)	175 (100)
	DBP †††	< 90	569 (93.9)	105 (89)	152 (98.1)	138 (87.3)	174 (99.4)
		90+	37 (6.1)	13 (11)	3 (1.9)	20 (12.7)	1 (0.6)
		Total	606 (100)	118 (100)	155 (100)	158 (100)	175 (100)
Dyslipidaemia	HDL–C, male	35+ mg/dL	246 (90.8)	106 (91.4)	140 (90.3)		
		< 35 mg/dL	25 (9.2)	10 (8.6)	15 (9.7)		
		Total	271 (100)	116 (100)	155 (100)		
	HDL–C, female	45+ mg/dL	303 (91.8)			141 (91)	162 (92.6)
		< 45 mg/dL	27 (8.2)			14 (9)	13 (7.4)
		Total	330 (100)			155 (100)	175 (100)
	LDL–C	< 160 mg/dL	548 (92.4)	107 (93.9)	143 (94.7)	137 (89)	161 (92.5)
		160+ mg/dL	45 (7.6)	7 (6.1)	8 (5.3)	17 (11)	13 (7.5)
		Total	593 (100)	114 (100)	151 (100)	154 (100)	174 (100)
	Triglycerides*	< 200 mg/dL	552 (92.2)	105 (91.3)	131 (85.1)	148 (95.5)	168 (96)
		200+ mg/dL	47 (7.9)	10 (8.7)	23 (14.9)	7 (4.5)	7 (4)
		Total	599 (100)	115 (100)	154 (100)	155 (100)	175 (100)
Hyperglycaemia	Fasting glucose	< 100 mg/dL	391 (65.1)	60 (51.7)	90 (58.1)	101 (65.2)	140 (80)
		100+ mg/dL	210 (34.9)	56 (48.3)	65 (41.9)	54 (34.8)	35 (20)
		Total	601 (100)	116 (100)	155 (100)	155 (100)	175 (100)
	HOMA-IR ††	< 2.3	298 (50.3)	43 (38.4)	68 (44.4)	63 (40.9)	124 (71.3)
		2.3+	295 (49.8)	69 (61.6)	85 (55.6)	91 (59.1)	50 (28.7)
		Total	593 (100)	112 (100)	153 (100)	154 (100)	Total (100)
	eGFR	< 90	264 (44.3)	51 (45.1)	84 (54.9)	56 (36.1)	73 (41.7)
		90+	332 (55.7)	62 (54.9)	69 (45.1)	99 (63.9)	102 (58.3)
		Total	596 (100)	113 (100)	153 (100)	155 (100)	175 (100)
	Fasting insulin ††	< 20 micro U/ML	518 (86.6)	96 (83.5)	131 (84.5)	127 (82.5)	164 (94.3)
		≥ 20 micro U/ML	80 (13.4)	19 (16.5)	24 (15.5)	27 (17.5)	10 (5.8)
		Total	598 (100)	115 (100)	155 (100)	154 (100)	174 (100)
(Abdominal) Obesity	BMI ††	< 30	356 (58.8)	62 (53)	100 (64.5)	62 (39.2)	132 (75.4)
		30+	249 (41.2)	55 (47)	55 (35.5)	96 (60.8)	43 (24.6)
		Total	605 (100)	117 (100)	155 (100)	158 (100)	175 (100)
	WC, male	< 102 cm	168 (61.5)	75 (63.6)	93 (60)		
		102+ cm	105 (38.5)	43 (36.4)	62 (40)		
		Total	273 (100)	118 (100)	155 (100)		
	WC, female ††	< 88 cm	161 (48.5)			44 (27.9)	117 (67.2)
		88+ cm	171 (51.5)			114 (72.2)	57 (32.8)
		Total	332 (100)			158 (100)	174 (100)
CVD	10-Year Framingham Risk, Male	< 10	194 (74.3)	84 (75)	110 (73.8)		
		10+	67 (25.7)	28 (25)	39 (26.2)		
		Total	261 (100)	112 (100)	149 (100)		
	10-Year Framingham Risk, female ††	< 8	277 (85.8)			118 (77.1)	159 (93.5)
		8+	46 (14.2)			35 (22.9)	11 (6.5)
		Total	323 (100)			153 (100)	170 (100)

†† $p < 0.001$ for females by race

††† $p < 0.001$ for males and females by race, respectively

Hypertension

The information of SBP and DBP including the median, quartiles, interquartile range (IQR), mean and outliers (i.e. values more than 1.5 IQRs from the upper and lower quartiles) are presented in Figure 10. The median SBP and DBP differed significantly by race and sex, respectively ($p < 0.01$, Kruskal-Wallis equality-of-populations rank test). There are distinctive SBP outliers (far more than 1.5 IQRs) in black females, e.g. 169.8, 175.6, 181.3, 193.8 mmHg.

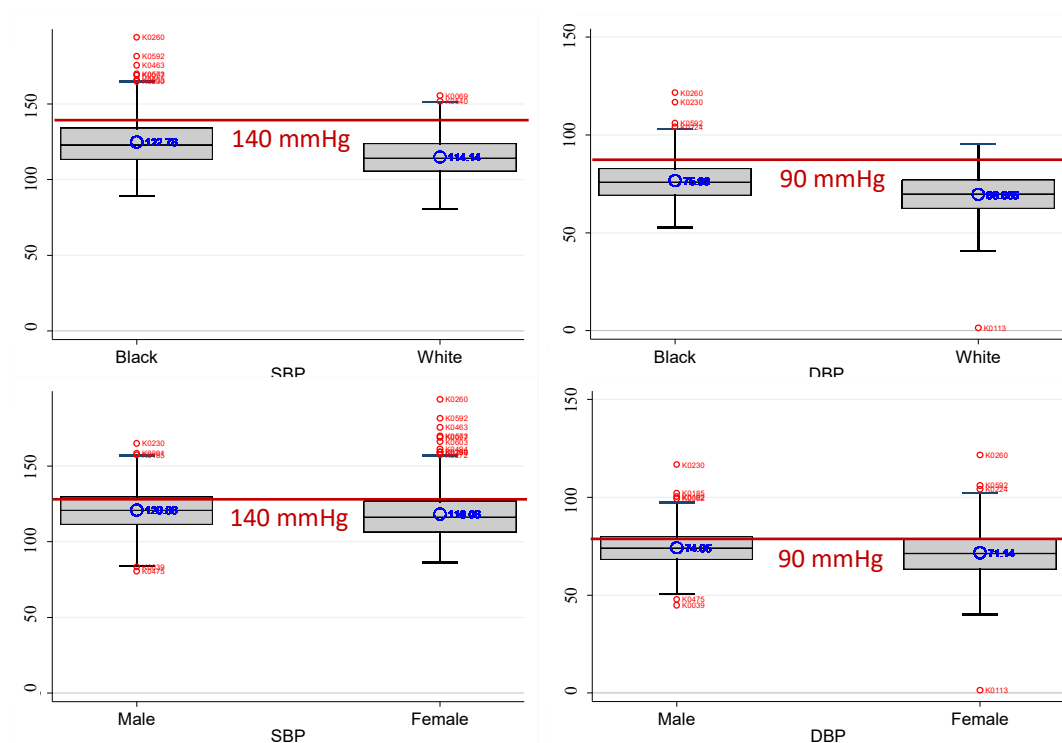


Figure 10 Boxplot of SBP and DBP (by race and sex)

Antihypertensive treatments records in the CARDIA Study at Year 30

The AHTs list at Year 30 consists of ACE inhibitors, alpha blockers, angiotensin II, beta blockers, calcium, centrally acting alpha-agonists, hypertensive agents, diuretic, nitrate and vasodilators. About 16% of the participants had at least one AHT (Table 6).

Table 6 Antihypertensive treatments (AHTs) at Year 30 (by Race and Sex)

		Total, N(%)	Black, N(%)	White, N(%)	Male, N(%)	Female, N(%)
AHTs count	0	464 (77.0)	209 (76.3)	255 (77.5)	216 (80.0)	248 (74.5)
	1	98 (16.3)	47 (17.2)	51 (15.5)	40 (14.8)	58 (17.4)
	2	25 (4.2)	14 (5.1)	11 (3.3)	9 (3.3)	16 (4.8)
	3	11 (1.8)	4 (1.5)	7 (2.1)	3 (1.1)	8 (2.4)
	4	5 (0.8)	0 (0)	5 (1.5)	2 (0.7)	3 (0.9)
	Total	603 (100)	274 (100)	329 (100)	270 (100)	333 (100)

Dyslipidemia

There was a significant difference in the median HDL-C between black and white female participants ($p=0.035$, Kruskal-Wallis equality-of-populations rank test). The median TG concentration differs significantly between males and females in the study ($p < 0.001$, Kruskal-Wallis equality-of-populations rank test). There are distinctive outliers in females (far more than 1.5 IQRs) in TG, i.e. 515 and 554 mg/dL which are greater than very high level of 500 mg/dL (Figure 11).

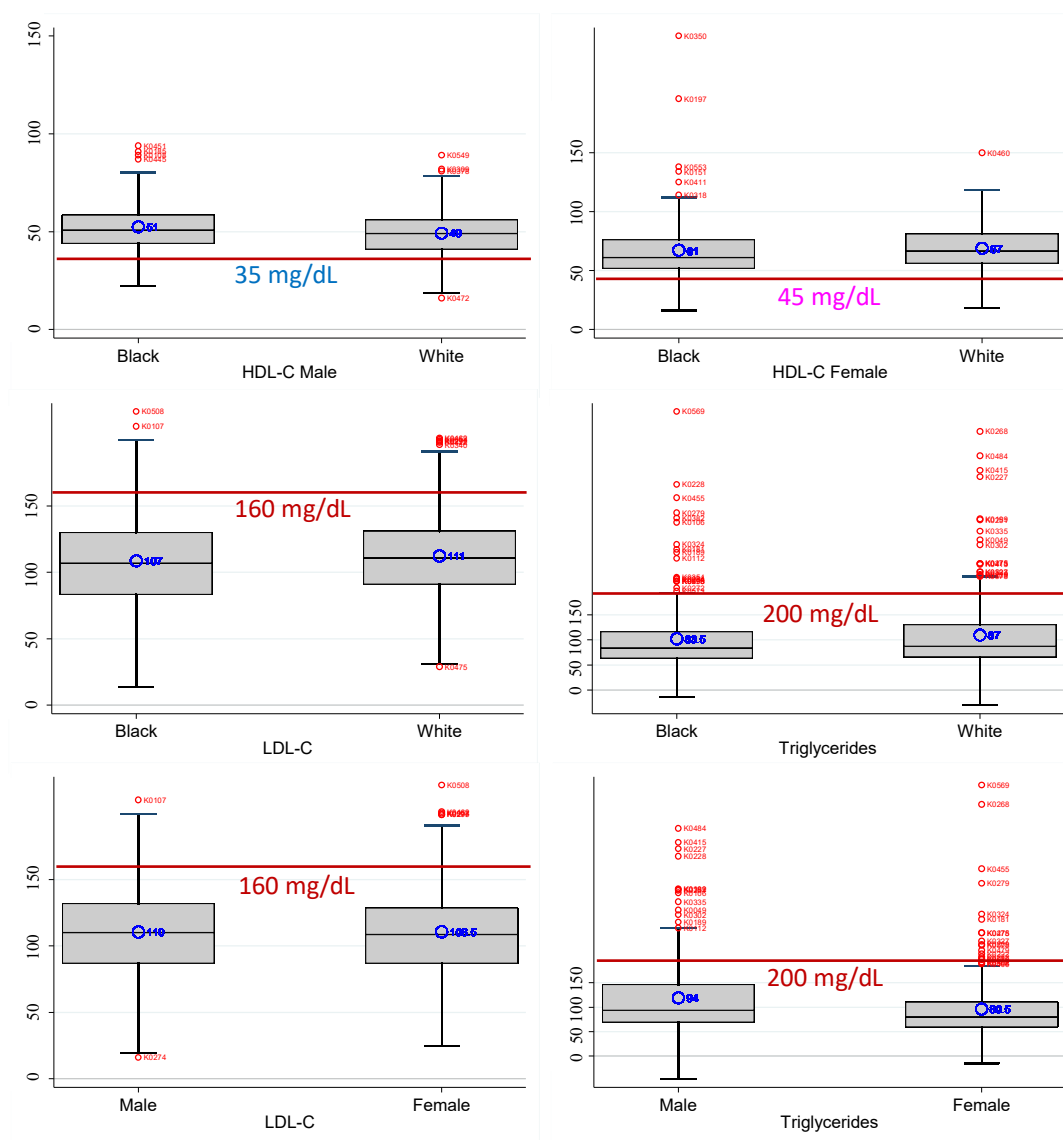


Figure 11 Boxplot of cholesterol and triglycerides in terms of dyslipidemia (by race and sex)

Hyperglycaemia

A fasting blood sugar level less than 100 mg/dL (5.6 mmol/L) is considered to be normal. The median fasting glucose was significantly different between males and females in the study ($p < 0.001$, Kruskal-Wallis equality-of-populations rank test). The median fasting insulin and eGFR differed significantly by race and sex, respectively ($p < 0.01$, Kruskal-Wallis equality-of-populations rank test). The median

HOMA-IR showed significant differences in race and sex ($p < 0.001$, Kruskal-Wallis equality-of-populations rank test) (Figure 12). It is known that women are more likely to develop impaired glucose tolerance (IGT) [101].

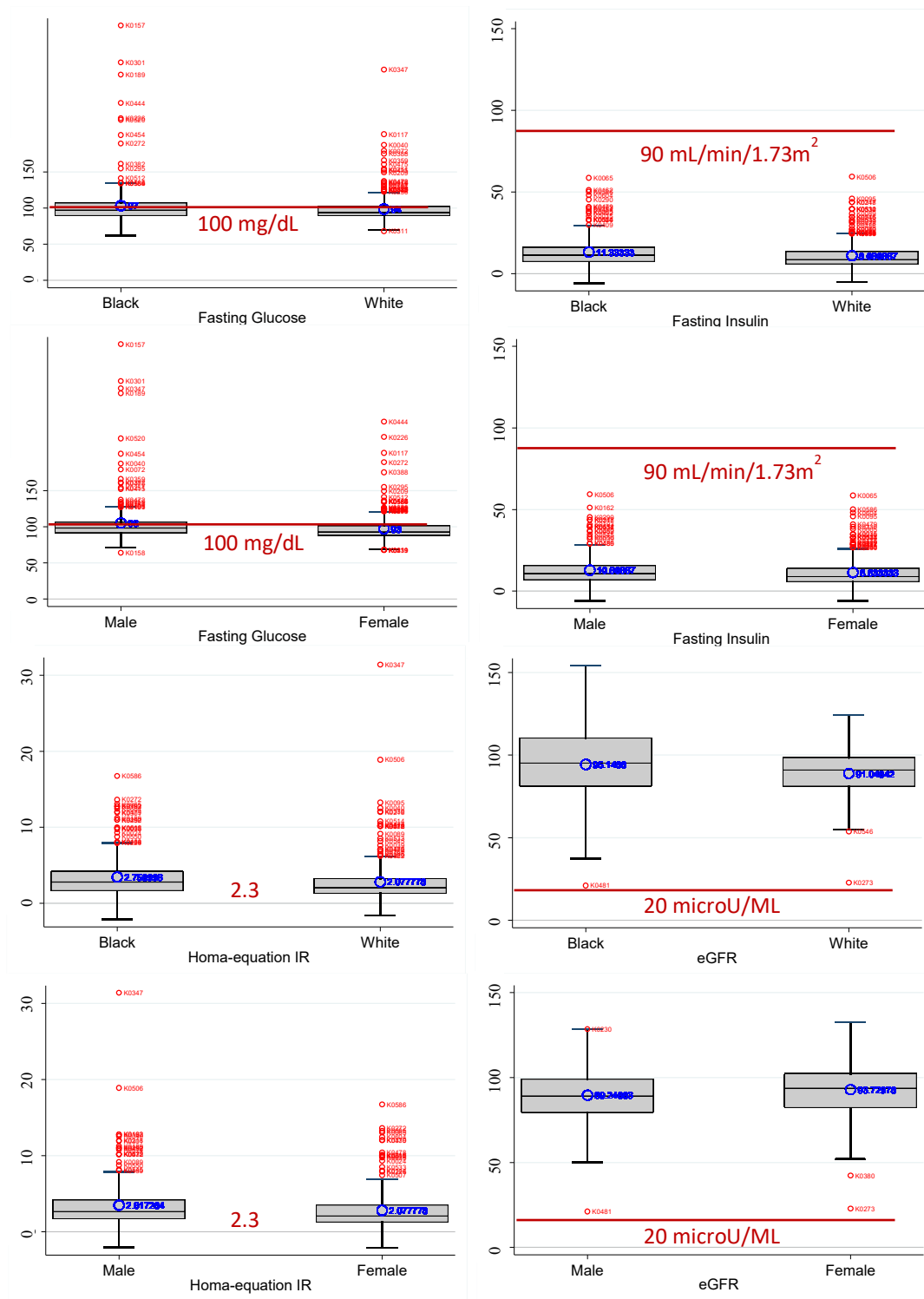


Figure 12 Boxplot of fasting glucose, HOMA-IR, eGFR and fasting insulin in terms of hyperglycaemia (by race and sex)

(Abdominal) Obesity

Obesity is a condition of abnormal or excess fat accumulation in adipose tissue and arises as a disorder of energy regulations, to the extent that health may be impaired [29, 30]. It has been reported that a number of health-related disturbances are strongly associated with obesity. Obesity increases risks associated with the development of other cardio-metabolic risk factors such as CHD, type 2 diabetes, IR, hypertension, and hypertriglyceridemia, in addition to certain cancers and non-alcoholic fatty liver disease [30]. There was a significant difference in the median BMI between black and white participants ($p < 0.001$, Kruskal-Wallis equality-of-populations rank test) (Figure 13).

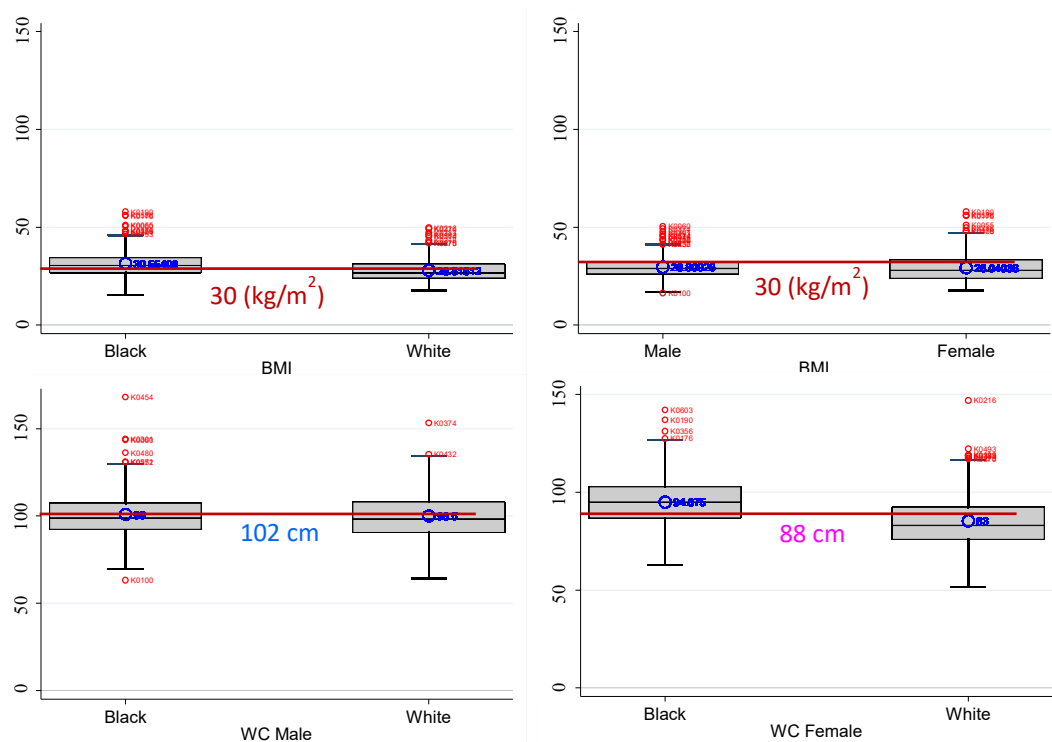


Figure 13 Boxplot of BMI and WC in terms of (abdominal) obesity (by race and sex)

The Framingham Risk Score

The cholesterol-based 10-year Framingham Risk Score was used as a measure of CVD risk in this sub-cohort from the CARDIA Study. A study reported potent relationships between Framingham risk scores and components of metabolic syndrome (including SBP, DBP, WC, and fasting serum glucose) in the Iranian cohort [102]. In the Framingham study, individuals with overt CHD at the baseline examination were excluded in the development of the risk equations. The age-adjusted Cox proportional hazards regression was used. The model was fitted separately for men and women [103].

There was a significant difference in the median 10-Year Framingham Risk between black and white female participants ($p < 0.001$, Kruskal-Wallis equality-of-populations rank test) (Figure 14).

range from zero to 30, with a score of 26 and higher generally considered normal for normal elderly individuals [91].

Construction of a composite cognitive score

A composite cognitive score is comprised of the z-scores from cognitive function tests [111]. The summation score of the standardized four cognitive tests (RAVLT, DSST, reversed Stroop, and MoCA) was generated as the **adjusted composite cognitive score** for the analyses at Year 30. As the MoCA was not measured at Year 25, the summation score of the standardized three cognitive tests (RAVLT, DSST, and reversed Stroop) was generated as a composite cognitive score for the analyses of degeneration at Year 25 and Year 30.

The mean (SD) of the cognitive measurements for the DSST was 69 (16.9) with the range (17,112); the STROOP was 22.3 (11.6) with the range (0, 107); the RAVLT was 9.3 (2.0) with the range (2.8, 14) at Year 30. The mean (SD) of the adjusted composite cognition score at Y30 was 0.02 (3.2) with the range (-14.6, 6.7). A negative score indicates below-normal cognitive function or (mild) cognitive impairment. The Mean (SD) of cognitive measurements at Year 25: the DSST was 71.8 (15.1) with the range (29,111); the STROOP was 21.2 (8.8) with the range (5, 71); The RAVLT was 8.8 (2.6) with the range (1.5, 14.5) (Table 7).

Table 7 Cognitive measurements at Year 25 and Year 30

Year 30	Total Mean (SD)	Black Mean (SD)	White Mean (SD)	Male Mean (SD)	Female Mean (SD)
Cleaned DSST score, Y30	596 69.0 (16.9)	270 62.3 (16.4)	326 74.6 (15.2)	267 64.4 (16.5)	329 72.8(16.3)
Calculated STROOP score, Y30	589 22.3 (11.6)	266 27.2 (11.9)	323 18.3 (9.5)	263 22.7 (12.7)	326 22.0(10.6)
Cleaned RAVLT score, Y30	596 9.3 (2.0)	270 8.3 (1.7)	326 10.0 (1.8)	267 8.8 (1.9)	329 9.7 (1.9)
Cleaned MOCA score, Y30	595 24.4 (3.6)	270 22.5 (3.7)	325 25.9 (2.8)	267 24.2 (3.7)	328 24.5(3.6)
Composite cognitive score, Y30¹⁾	589 0.02 (2.4)	266 -1.3 (2.2)	323 1.1 (2.1)	263 -0.6 (2.5)	326 0.5 (2.3)
Adjusted composite cognitive score, Y30²⁾	588 0.02 (3.2)	266 -1.8 (2.9)	322 1.5 (2.6)	263 -0.6 (3.2)	325 0.5 (3.0)
Year 25	Total Mean (SD)	Black Mean (SD)	White Mean (SD)	Male Mean (SD)	Female Mean (SD)
Cleaned DSST score, Y25	582 71.8 (15.1)	260 65.8 (14.9)	322 76.7 (13.4)	262 68 (14.6)	320 74.9(14.8)
Calculated STROOP score, Y25	581 21.2 (8.8)	259 25.6 (10)	322 17.8 (5.6)	262 21.7 (9.3)	319 20.9(8.3)
Cleaned RAVLT score, Y25	580 8.8 (2.6)	260 7.7 (2.3)	320 9.8 (2.4)	261 8.1 (2.5)	319 9.4 (2.5)
Composite cognitive score, Y25¹⁾	579 0.001 (2.3)	259 -1.3 (2.2)	320 1.1 (1.8)	261 -0.6 (2.3)	318 0.5 (2.2)

1) **Composite cognitive score, Y25 and Y30***: z-scores scores from the three tests (without MoCA) were generated

2) **Adjusted composite cognitive score, Y30***: z-scores scores from all four tests were generated (with the reverse-Stroop score)

* Note: RAVLT is an average of words recalled in immediate, short, and long

The Mean (SD) difference of cognitive measurements between Years 25 and 30 is presented in Table 8. The DSST has worsened 2.5 (9.1) with a right skewed distribution in the change score; the STROOP has worsened 0.99 (8.9) (i.e. lower score is better); the RAVLT has improved 0.44 (1.8) in comparison to the cognitive measurements at Year 25. (The Wilcoxon signed-rank tests, i.e. equality tests on

matched data, $p < 0.001$; Figure 15). A negative composite cognitive score is an indication of having below-normal cognitive function or (mild) cognitive impairment. The mean (SD) of the composite cognitive score (without the standardized MoCA) was 0.02 (3.2) with the range (-12.9, 5.4) at Year 30; the mean (SD) 0.001 (2.29) and the range (-7.9, 6.0) at Year 25. The degeneration of the composite cognitive score (Year 30 - Year 25) of females was mean (SD) -0.05 (1.3).

Table 8 The changed cognitive measurements (i.e. Year 30 - Year 25)

	Total Mean (SD)	Black Mean (SD)	White Mean (SD)	Male Mean (SD)	Female Mean (SD)
DSST Y30 - DSST Y25*	572 -2.5 (9.1)	254 -2.96 (8.0)	318 -2.129 (10)	256 -2.89 (7.7)	316 -2.19 (10.2)
STROOP Y30 - STROOP Y25*	565 0.99 (8.9)	249 1.39 (9.9)	316 0.677 (8.1)	252 1.07 (9.2)	313 0.93 (8.8)
RAVLT Y30 - RAVLT Y25*	570 0.44 (1.8)	254 0.68 (1.9)	316 0.251 (1.7)	255 0.63 (1.8)	315 0.29 (1.8)
CompositeY30 - CompositeY25*	563 0.016 (1.3)	251 0.03 (1.3)	312 0.004 (1.4)	249 0.11 (1.4)	314 -0.05 (1.3)

Note: z-scores scores from all three tests were generated (with the reverse-Stroop score)
MoCA was not reported at Year 25

* The changed cognitive measurements are significant ($p < 0.001$) by the Wilcoxon signed-rank test

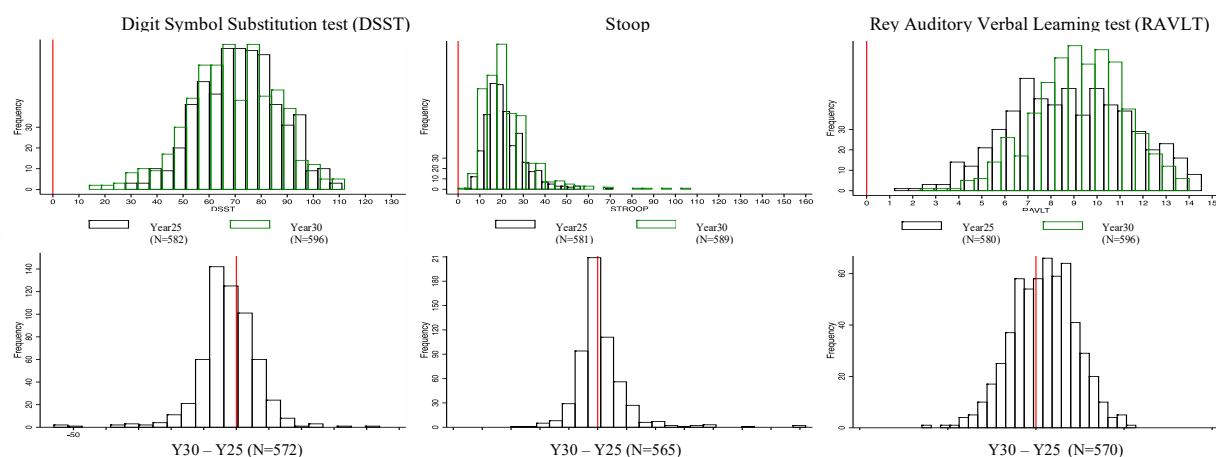


Figure 15 The DSST, Stroop and RAVLT at Year 25 and Year 30, and the changed scores

The adjusted composite cognitive score (i.e. z-scores from the four tests, N=588) was spread across a range of -14.62 to 6.70 with mean (SD) 0.024 (3.18); a positive composite cognitive score indicates normal cognition (Figure 16). The distributions of the DSST, Stroop, RAVLT, MoCA and the adjusted composite cognitive score by race and sex are presented in Appendix Figure 12–16; the median adjusted composite cognitive score at Year 30 differed significantly by race and sex, respectively ($p < 0.001$, Kruskal-Wallis equality-of-populations rank test). While all four median cognitive measurements at Years 25 and 30 differed significantly by race, Stroop and MoCA differed significantly by sex, respectively ($p < 0.001$, Kruskal-Wallis equality-of-populations rank test).

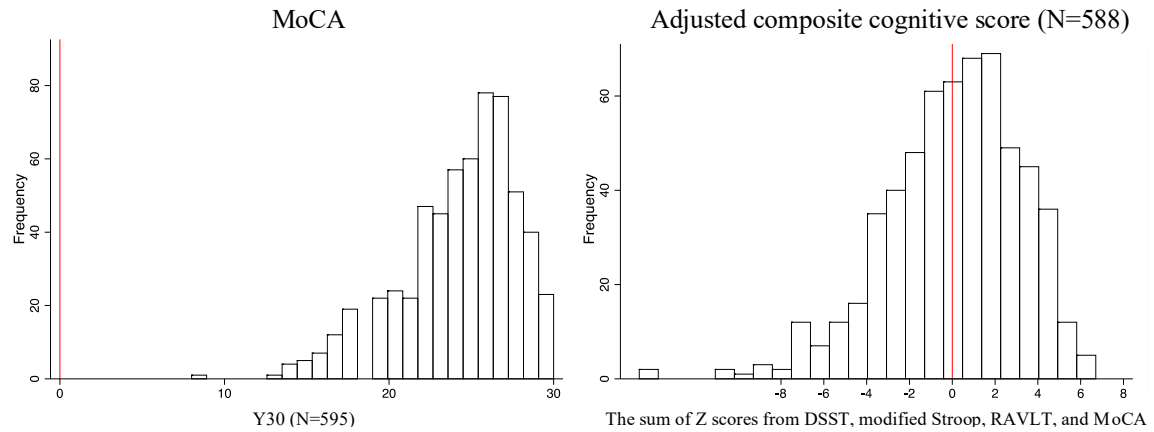


Figure 16 The MoCA and adjusted composite cognitive score at Year 30

1-5. Brain MRI parameters at Year 25 and Year 30

Information relating to the brain MRI parameter is reported for Years 25 and 30 in the CARDIA Study for a sub-set of the volunteers.

Armstrong *et al.* (2017) investigated the association between early cardiac dysfunction and brain abnormalities in middle-aged adults at Year 25 in CARDIA (N=719) [112]. The study used brain health information including cerebral MRI diffusion tensor imaging (DTI), WM fractional anisotropy (FA) and GM CBF. Zhu *et al.* (2015) reported that greater physical fitness was associated with more brain volume and greater WM integrity in middle-aged adults (N=565 at Year 25 in CARDIA) [113]. In the study, the brain MRI parameters were analysed using measures of means and percentiles, as group characteristics. The low total brain volume (TBV), low WM normal tissue volume, low WM FA and low WM CBF were defined as less than equal to 15th percentile of the sample values. High abnormal tissue volume was defined as greater than equal to 85th percentile [113].

In the CARDIA Study, measurements of (1) total volume (cc³), (2) abnormal tissue volume (cc³), (3) CBF (mm³/100gms/min) and (4) FA (normalized) at Year 25 and Year 30 were investigated for clinically important measurements. Clinically important measures include intra-cranial, brain tissue, GM and WM, cerebrospinal fluid and in addition, the brain structures for the successive stages of memory processing from previous research. The distributions of the brain measurements are presented in this section.

Total volume

The histograms of total volumes measured at Years 25 and 30, and their changes (i.e. Year 30 - Year 25) are presented in Figure 17. Twenty-five percentiles of change in brain (-6952.3 cc^3) and fiftieth percentiles of change in GM (-2722.7 cc^3) volumes are less zero (i.e. Year 30 - Year 25 < 0), indicating these brain MRI parameters decreased for more than 25% and 50% of the study population, respectively (N=187 individuals measured at both years).

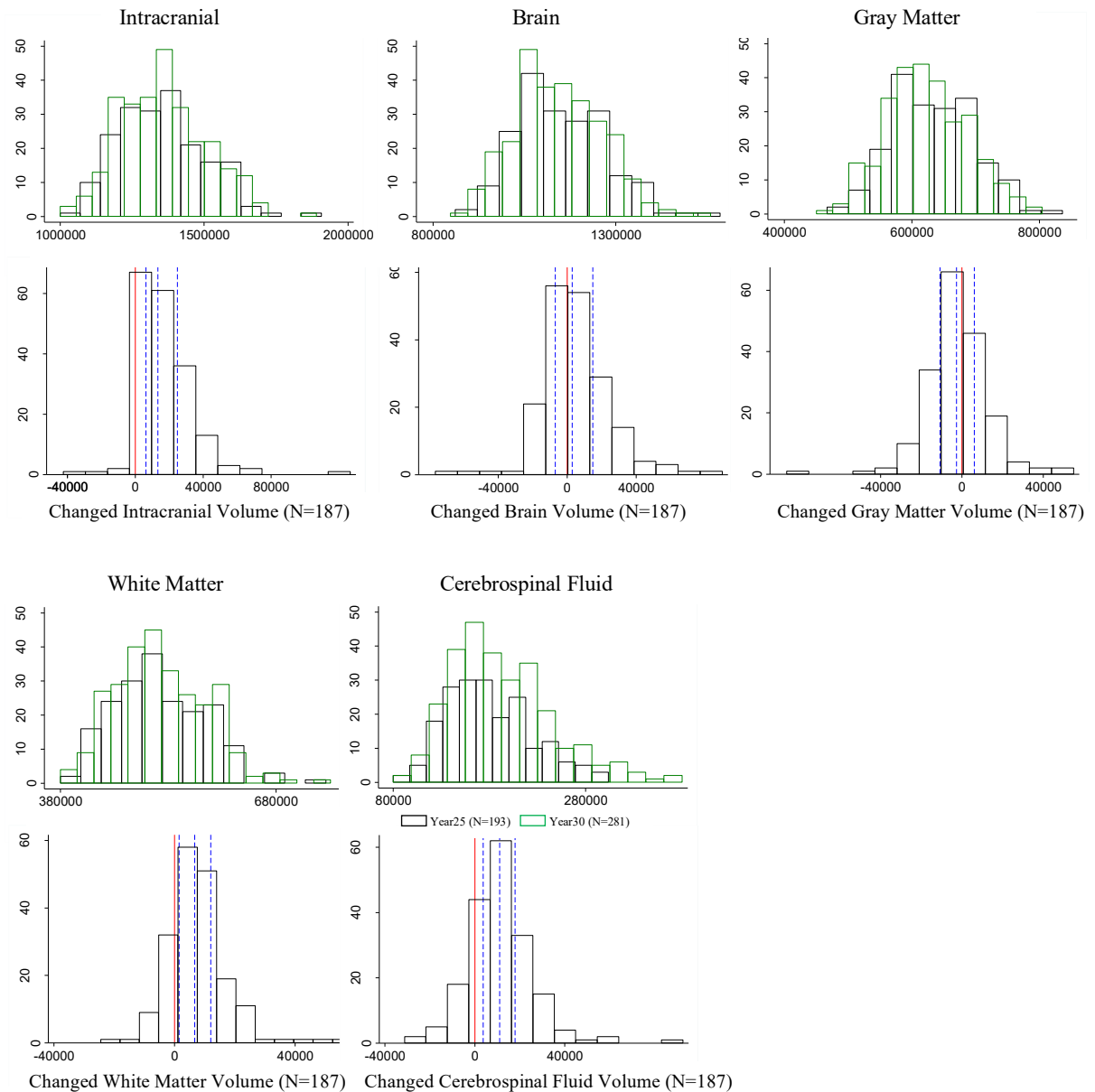


Figure 17 Total volumes (cc³) of intracranial (ICV), brain, white matter, gray matter, and cerebrospinal fluid; the blue lines indicate the 25th, median, and 75th percentile, respectively

Cerebrospinal fluid is a clear, colourless liquid which contains small amounts of protein [114]. Many studies have found an increase in cerebrospinal fluid tau protein (CSF tau) in AD (brain changes caused by brain plaques and tangles). Cerebrospinal fluid tau protein which is reported to be a biochemical marker for AD was highlighted in a community-based follow up study [115]. It has been reported that abnormally phosphorylated forms of the microtubule-associated protein tau, which have been consistently found in the brains of AD patients and can be detected in cerebrospinal fluid [116]. However, an increase in CSF tau has also been found by some studies in VaD, brain changes caused by blocked or damaged blood vessels in brain, although not in all [117]. Blennow (2004) explained these discrepant findings may depend on the patient characteristics and diagnostic criteria used [118]. High total tau (T-tau) in clinically diagnosed VaD cases may be caused by accompanying AD pathology in VaD cases, which is difficult to identify clinically [118].

Abnormal tissue volume

The normal and abnormal tissues were characterized in the CARDIA Study based on automated analysis of signal intensity, and assigned as 92 anatomical regions of interest (ROIs) covering both hemispheres [119]. An automated computer algorithm segmented MRI structural images of supratentorial (i.e. above the tentorium of the brain) brain tissue into GM (it serves to process information in the brain), WM (it plays an essential role in communication between the brain regions and spinal cord) and cerebrospinal fluid [111].

An automatic algorithm detects abnormal tissues by comparison with the mean blood pool intensity and labels the corresponding vertices in the 3D meshes [120, 121]. GM and WM were further characterized as normal and abnormal tissue based on automated analysis of signal intensity and assigned as 92 ROIs covering both hemispheres [111]. These 92 anatomic regions were derived from an expert-delineated atlas [122], which was used as a brain template to which the MRI brain measures were co-registered [111].

Histograms of abnormal tissue volumes measured at Year 25 and Year 30, and their changes are presented. Fiftieth percentiles of change in abnormal brain tissue (-142.6 cc^3); abnormal GM tissue (-34.8 cc^3); abnormal WM tissue (-95.4 cc^3); and abnormal cerebrospinal fluid tissue (-60 cc^3) volumes are less than zero (i.e. $\text{Year 30} - \text{Year 25} < 0$), indicating these brain abnormal tissue volumes decreased for more than 50% of the study population ($N=186$ individuals measured at both years) (Figure 18).

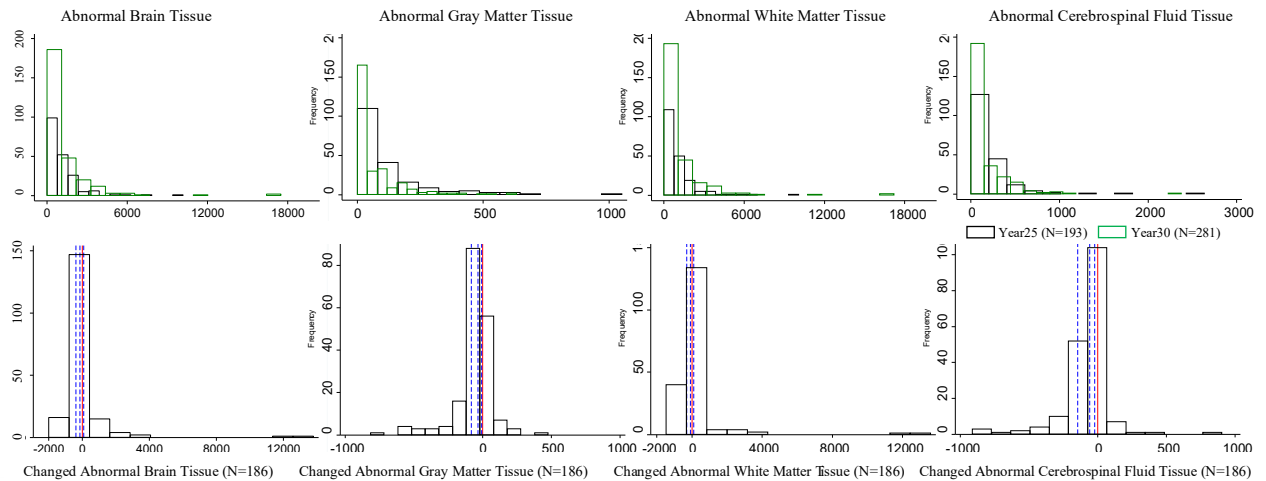


Figure 18 Abnormal tissue volume (cc³); the blue lines indicate the 25th, median, and 75th percentile, respectively

Fractional anisotropy

FA, a measure of the degree of diffusion directionality, is a scalar measure of anisotropy in diffusion tensor (DT) MRI, which is a sensitive technique that provides useful in vivo maps of deep-tissue organization (Figure 19) [1-3]. Consistent with a number of previous reports, Giorgio, *et al.* (2010) and Salat *et al.* (2005) both reported widespread linear negative associations of FA with age, and they suggest the loss of directionality in WM fibres becomes apparent during the transition from young to middle adulthood (41-59 years). In addition, schizophrenia is reportedly associated with reduced FA values in frontal regions (which correlates with cognitive and motor deficits in patients [123]) and in the right occipital WM, and in the left posterior hippocampus have been reported repeatedly in adult-onset schizophrenia [124, 125].

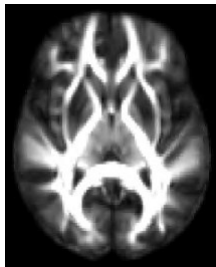


Figure 19 FA is a summary measure of microstructural integrity; under the terms of the Creative Commons Attribution 4.0 International License, which permits unrestricted use, distribution, and redistribution in any medium, provided that the original author and source are credited [1-3]

A value of zero FA means that diffusion is equally restricted in all directions. The histograms of (corrected mean) FA measured at Year 25 and Year 30, and their changes are presented. Seventy-five percentiles of change in FA brain (-0.0047); GM (-0.005); WM (-0.0009); and cerebrospinal fluid (-0.005) are less than zero (i.e. Year 30 - Year 25 < 0), indicating the degree of anisotropies of these brain parameters decreased for more than 75% of the study population (N=183 individuals measured at both years) (Figure 20).

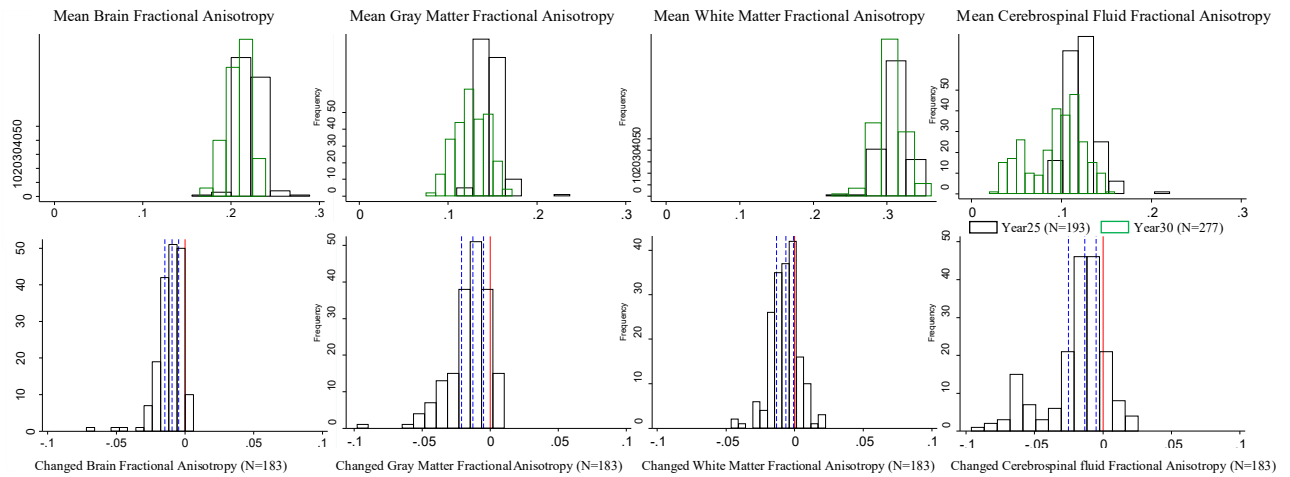


Figure 20 (corrected mean) Fractional anisotropy (normalized); the blue lines indicate the 25th, median, and 75th percentile, respectively

Cerebral blood flow

In the general population, lower cerebral perfusion is associated with accelerated cognitive decline and increased risk of dementia. In a population-based study (follow-up 6.9 years of participants median age 61.3 years), lower cerebral perfusion at baseline was associated with accelerated decline in global cognition, particularly in memory and executive function [126]. CBF, the blood supply to the brain in a given period of time, can be used as biomarker for Parkinson's disease. There is no definitive validated biomarker for Parkinson's disease at this time, it is considered that positron emission tomography (PET) imaging (which are often combined with computerized tomography (CT) or MRI scans) measuring cerebral glucose metabolism and CBF may have some potential application for the differential diagnosis of parkinsonian syndromes [127].

The histograms of CBF measured at Year 25 and Year 30, and their changes are presented (Figure 21). Fiftieth percentiles of change in CBF of brain ($-2.8 \text{ mm}^3/100\text{gms/min}$); GM ($-3.0 \text{ mm}^3/100\text{gms/min}$); WM ($-2.5 \text{ mm}^3/100\text{gms/min}$) are less than zero (i.e. $\text{Year 30} - \text{Year 25} < 0$), indicating these CBF parameters decreased by more than 50% of the study population ($N=109$ individuals measured at both years).

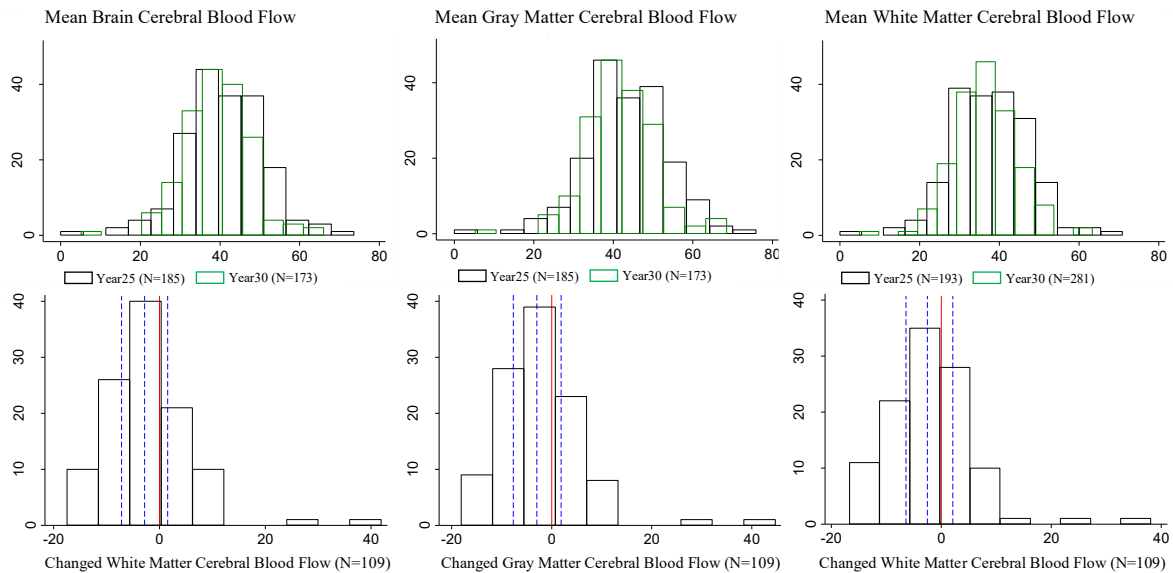


Figure 21 (Mean) Cerebral blood flow (mm³/100gms/min); the blue lines indicate the 25th, median, and 75th percentile, respectively

Brain structure involved in the successive stages of memory processing

The hippocampus, medial frontal cortex (MFC) and anterior cingulate gyrus (ACG) are known for the successive stages of memory processing for encoding, storage, and retrieval [128-130]. The hippocampus (i.e. hippocampus proper, cornu ammonis, Ammon's horn) is a convex elevation, approximately 5 cm long, within the hippocampal gyrus (a GM cortical region of the brain that surrounds the hippocampus and is part of the limbic system) inside the inferior or temporal horn of the lateral ventricle (Figure 22) [114, 131]. In her book, Watchman explained (2017) that the hippocampus is needed for retrieval of memories, but the retrieval of older memories may depend on it less. According to Watchman, this could explain the phenomenon of someone in the earlier stages of AD (with a damaged hippocampus but an intact cortex) remembering a childhood holiday but at the same time, struggling to remember the activities they have done during the day [132].

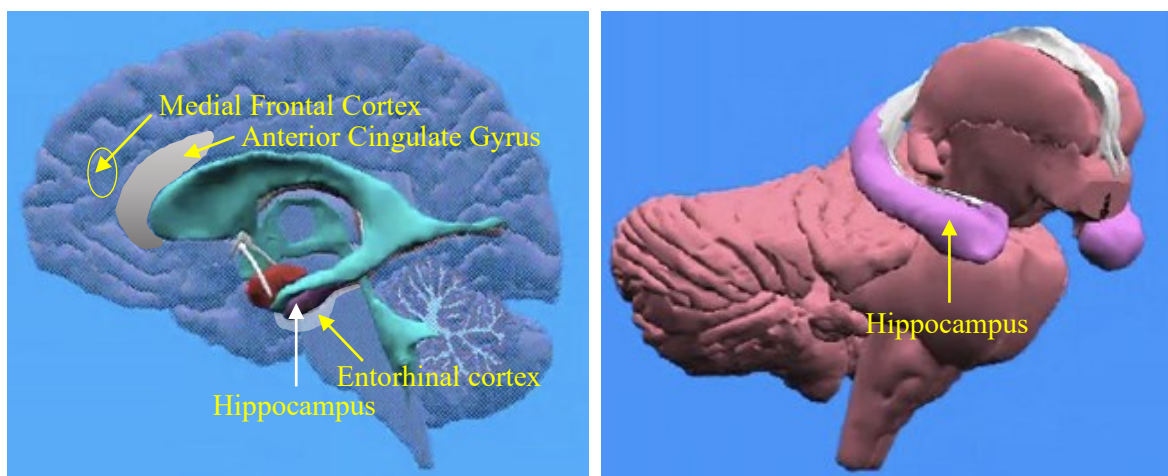


Figure 22 The hippocampus, medial frontal cortex, anterior cingulate cortex and entorhinal cortex seen inside a translucent cerebral hemisphere (left) and the hippocampus (right); images by the author of this dissertation [133]

The histograms relating to the hippocampus measured at Years 25 and 30, and their changes (i.e. Year 30 - Year 25) are presented. More than 50% of the study population showed a decrease in hippocampus volume; fiftieth percentiles of those (N=187 at both years) with changed volumes are right (10.5 cc³) and left (19.1 cc³). More than 50% of the study population had a decrease in abnormal tissue volume in hippocampus (right -1.95 cc³ and left -0.95 cc³ in those N=186 at both years). FAs in hippocampus decreased by more than 75% of the study population (right -0.0006 and left -0.003 in those N=183 at both years). More than 25% of the study population showed increased CBF in hippocampus; seventy-five percentiles of those (N=109 at both years) with changed values are right (1.99 mm³/100gms/min) and left (2.7 mm³/100gms/min) (Figure 23).

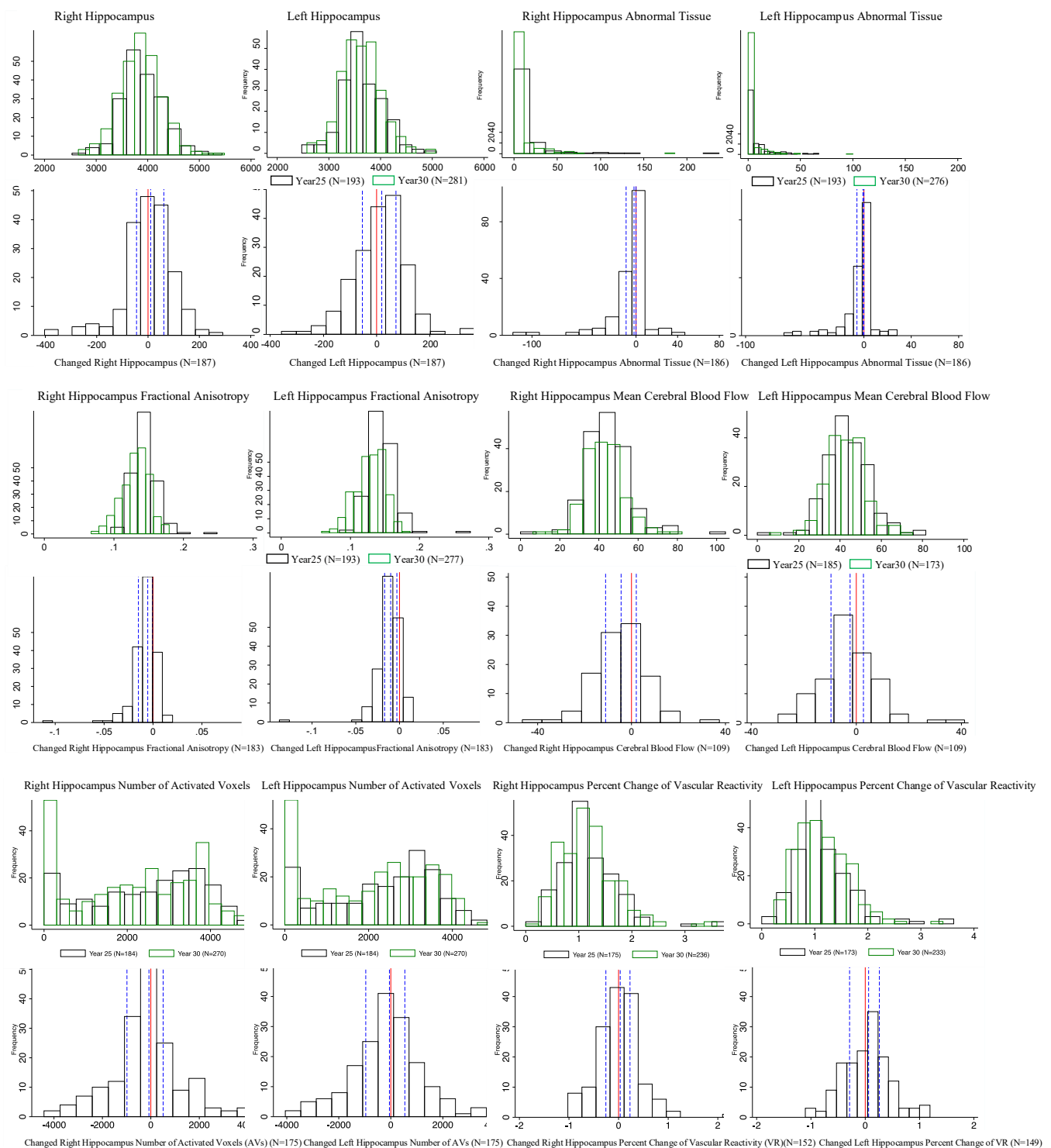


Figure 23 Hippocampus – Total volumes (cc^3), abnormal tissue volume (cc^3), fractional anisotropy (normalized), (mean) cerebral blood flow ($\text{mm}^3/100\text{gms}/\text{min}$), number of activated voxels, and (mean) percent change of vascular reactivity; the blue lines indicate the 25th, median, and 75th percentile, respectively

Histograms showing the MFC measured at Years 25 and 30, and their changes (i.e. Year 30 - Year 25) are presented in Figure 24. More than 50% of the study population showed a decrease in MFC volume; fiftieth percentiles of those ($N=187$ at both years) with changed volumes are right (-16.2 cc^3) and left (-21.9 cc^3). There were no significant changes in abnormal tissue volume in MFC; quartiles of those ($N=186$ at both years) with changed volumes are zeros (cc^3). In addition, more than 50% of the study population showed decreases in the number of activated voxels within MFC (right -90 and left -71 in those $N=175$ at both years) and percent change of vascular reactivity in MFC (right -0.06 and left -0.12 in those $N=142-4$ at both years), respectively.

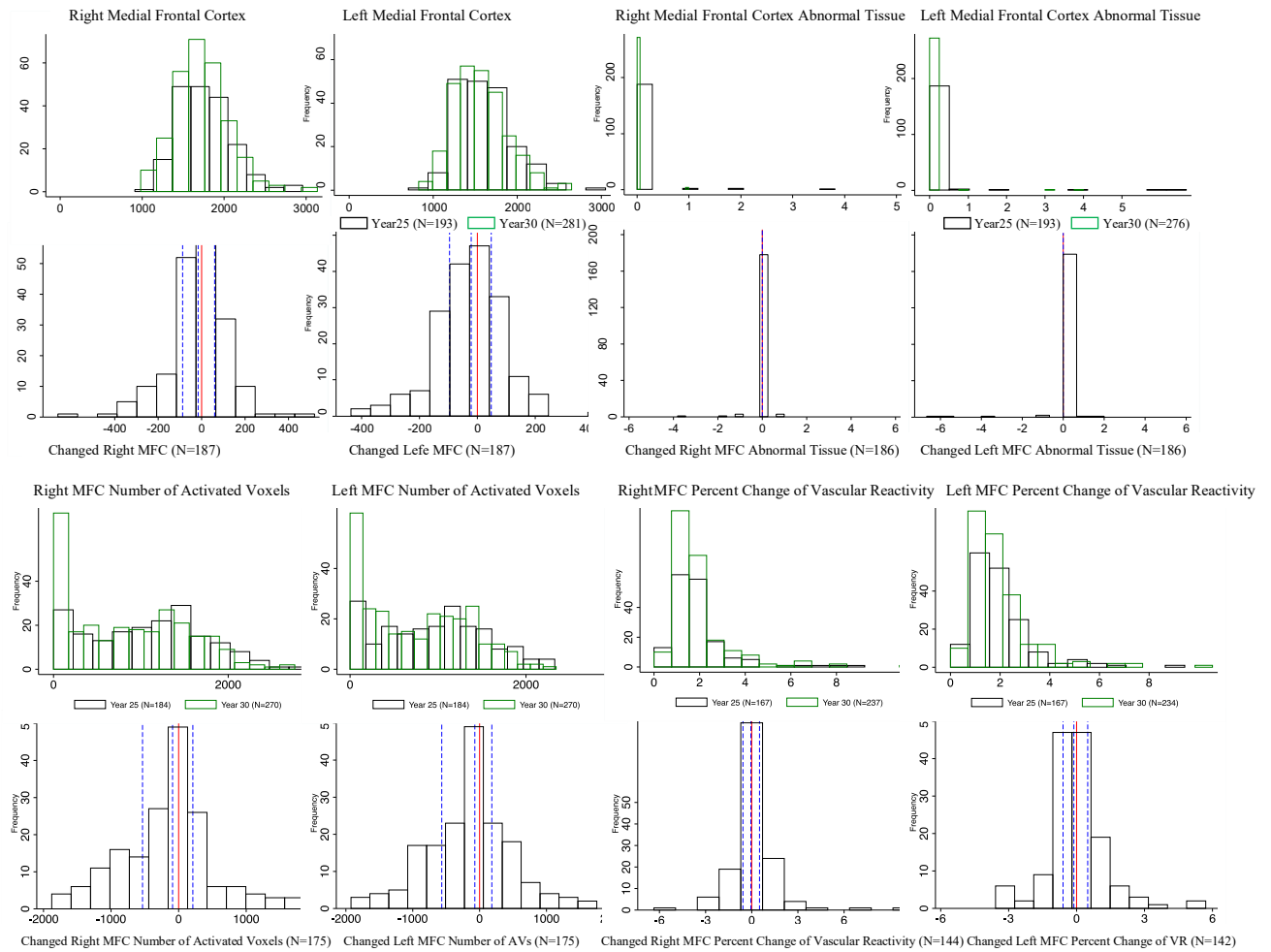


Figure 24 Medial frontal cortex – Total volumes (cc^3), abnormal tissue volume (cc^3), number of activated voxels, and (mean) percent change of vascular reactivity; the blue lines indicate the 25th, median, and 75th percentile, respectively

Histograms detailing the ACG measured at Years 25 and 30, and their changes (i.e. Year 30 - Year 25) are presented in Figure 25. More than 50% of the study population showed a decrease in ACG volume; fiftieth percentiles of those (N=187 at both years) with change in ACG volumes are right (-44.8 cc³) and left (-40 cc³). More than 25% of the study population had a decrease in abnormal tissue volume in ACG (right -4.8 cc³ and left -0.95 cc³ in those N=186 at both years). In addition, more than 50% of the study population showed decreases in the number of activated voxels within ACG (right -241 and left -282 in those N=175 at both years) and percent change of vascular reactivity in ACG (right and left -0.12 in those N=148 at both years), respectively.

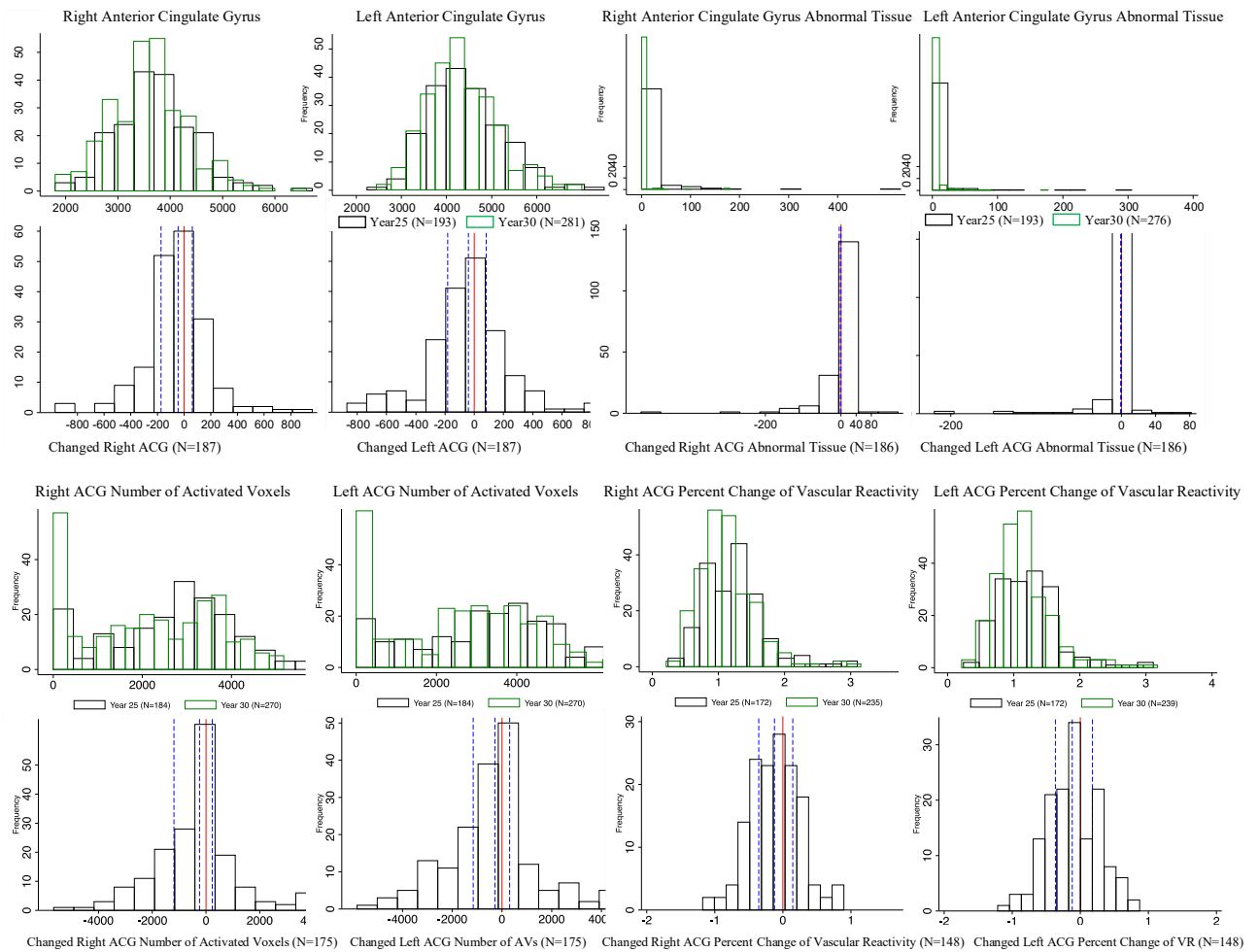


Figure 25 Anterior cingulate cortex – Total volumes (cc³), abnormal tissue volume (cc³), number of activated voxels, and (mean) percent change of vascular reactivity; the blue lines indicate the 25th, median, and 75th percentile, respectively

Brain structure as early markers of Alzheimer's disease

AD can be characterized by changes in the tau protein and in the cleaved fragments of the amyloid precursor protein (APP) [134].

In the present study, the (medial) temporal lobe volume and entorhinal area were additionally investigated as early markers of AD along with the hippocampus, given that these smaller brain structures are susceptible to neurofibrillary (abnormal accumulations of tau protein changes and atrophy in the early stages of the disease (i.e. Braak Stages in AD I and II, when the neurofibrillary tangles are limited to the trans-entorhinal region of the brain) [111, 135]. When the frontal and/or temporal lobes are damaged in this way, this causes the symptoms of frontotemporal dementia (FTD). These include changes in personality and behaviour, and difficulties with language. These symptoms are different from the memory loss often associated with more common types of dementia, such as AD [136].

The medial temporal lobes are known to influence performance in laboratory memory tasks and to be affected by aging and early-stage AD [137]. It is known that the cortex of the medial temporal lobe subserves fundamental mnemonic (aiding or designed to aid the memory) functions, providing critical input from heteromodal association cerebral cortices to the hippocampal formation (a compound structure in the medial temporal lobe) and receiving reciprocal afferents from the hippocampal formation [138, 139]. The medial temporal lobe volume is a composite structure which includes the entorhinal cortex, hippocampus proper, parahippocampal gyrus, and perirhinal cortex [111]. Visser *et al.* (2002) reported that the qualitative rating of medial temporal lobe atrophy is a good alternative to the volumetry of the hippocampus, in order to detect patients at high risk of AD among those with minor cognitive impairment in a study with participants aged 50 and over (N=31) [140].

Histograms relating to the temporal lobe measured at Years 25 and 30, and their changes (i.e. Year 30 - Year 25) are presented in Figure 26. More than 25% of the study population showed a decrease in temporal lobe volume (right -266.1 and left -277.0 in those N=187 at both years). More than 50% of the study population had a decrease in abnormal tissue volume in temporal lobe (right -4.3 cc³ and left -5.7 cc³ in those N=186 at both years).

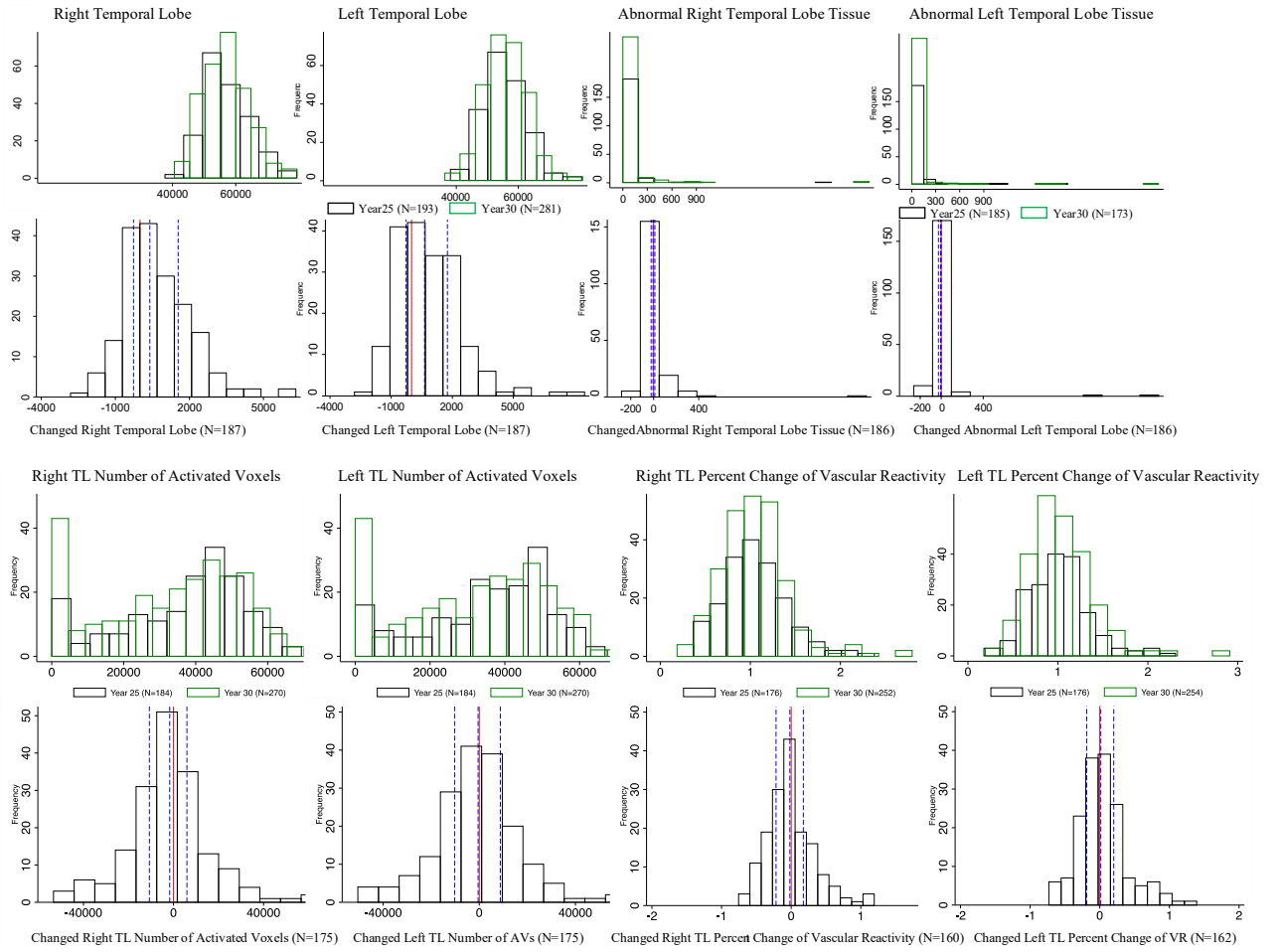


Figure 26 Temporal lobe – Total volumes (cc^3), Abnormal tissue volume (cc^3); the blue lines indicate the 25th, median, and 75th percentile, respectively

The entorhinal cortex operates as a centre in a widespread network for memory, navigation and the perception of time is located in the medial temporal lobe. Evidence from clinical and animal studies suggests that the entorhinal cortex plays an essential role within the fronto-temporal cortical memory network with the hippocampus, medial prefrontal cortex, and neocortex (the top layer of the cerebral hemispheres which is part of the cerebral cortex) [114, 141]. Go'mez-Isla *et al.* (1996) reported that the entorhinal cortex plays a crucial role as a gateway connecting the neocortex and the hippocampal formation [142]. Age-related loss of entorhinal volume has been reported and it has been implicated in the early stages of AD [134, 143]. Hoesen *et al.* (1991) stated that the entorhinal cortex (Brodmann's area 28) of the ventromedial temporal lobe is an invariant focus of pathology in all cases of AD [144].

A more recent review highlighted atrophy (shrivel) in the entorhinal cortex reflects the early pathological changes of AD and can be a strong predictor of early-stage AD (over the hippocampus) in individuals with mild cognitive impairment (MCI) [145]. The paper reported that the right EC in patients with mild AD exhibited greater changes than the left side [145]. The histograms of entorhinal area measured at Years 25 and 30, and their changes (i.e. Year 30 - Year 25) are presented in Figure 27. More than 75% of the study population showed an increase in entorhinal area volume

(right 187.9 cc³ and left 184.1 cc³ in those N=187 at both years). There were no significant changes in abnormal tissue volume in entorhinal area; quartiles of those (N=186 at both years) with changed volumes are zeros (cc³).

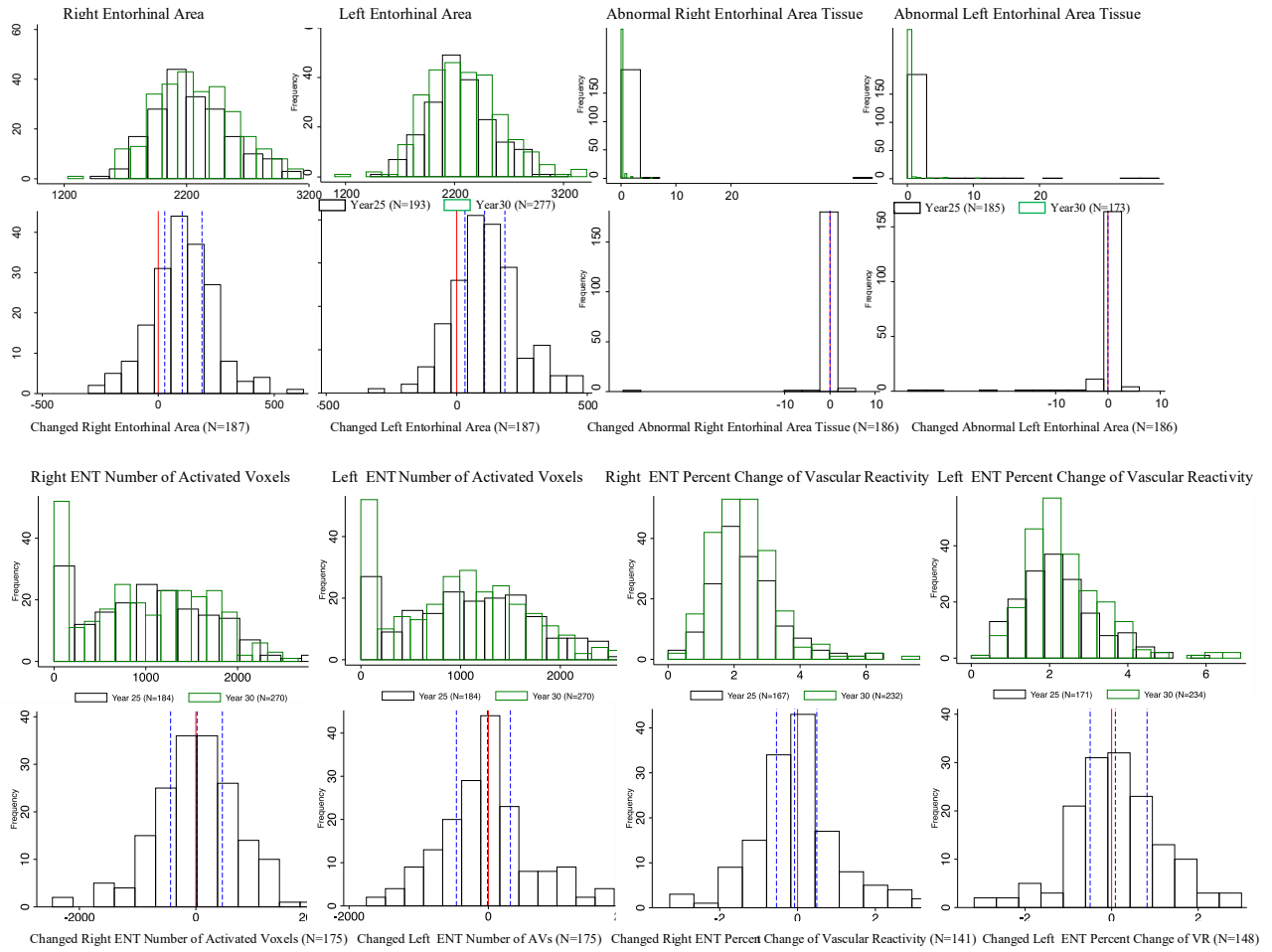


Figure 27 Entorhinal area – Total volumes (cc³), Abnormal tissue volume (cc³); the blue lines indicate the 25th, median, and 75th percentile, respectively

Total intracranial and brain volumes measured at Years 25 and 30 showed large variabilities. Abnormal tissue volume of temporal lobe WM right and left measured at both years show large ranges (min to max) and standard deviations (i.e. measures of variability) (Appendix Table 7).

The changed brain MRI parameters (i.e. Year 30 - Year 25) are significant by the Wilcoxon signed-rank test (i.e. equality tests on matched data). Most of changed brain MRI parameters measured for the number of activated voxels and percent change of vascular reactivity are significant by the Wilcoxon signed-rank test (i.e. equality tests on matched data). The changed number of activated voxels within temporal lobe WM and within right entorhinal area were not statistically significant (Table 9).

Table 9 The changed brain MRI parameters (i.e. Year 30 - Year 25)

		Change (Y30 - Y25) with participants have both Y25 and Y30					Wilcoxon signed-rank test			
		N	Mean	SD	Min	Max	positive	negative	zero	Prob > z
Total Volume (cc ³)	Intracranial	187	16644.52	16636.64	-42448.52	126647.00	174	13	0	<0.001
	Brain	187	5119.52	19740.13	-76476.91	89922.91	106	81	0	<0.001
	Gray Matter	187	-2265.67	15691.49	-86223.00	54931.64	76	111	0	0.011
	White Matter	187	7385.19	10156.89	-24486.48	58830.26	150	37	0	<0.001
	Cerebrospinal Fluid	187	11457.55	14878.29	-31271.86	92587.50	153	30	0	<0.001
	Right Hippocampus	187	4.20	97.78	-403.40	291.82	58	129	0	<0.001
	Left Hippocampus	187	11.86	99.19	-357.63	369.07	67	120	0	<0.001
	Right Medial Frontal Cortex	187	-21.22	138.20	-661.85	519.75	50	137	0	<0.001
	Left Medial Frontal Cortex	187	-27.86	121.38	-436.78	450.13	44	143	0	<0.001
	Right Anterior Cingulate Gyrus	187	-60.57	240.84	-967.03	960.35	51	135	1	<0.001
	Left Anterior Cingulate Gyrus	187	-49.32	272.86	-870.70	886.00	56	131	0	<0.001
	Temporal Lobe WM Right	187	691.11	1406.92	-2818.10	6307.60	118	69	0	<0.001
	Temporal Lobe WM Left	187	837.48	1578.27	-2804.76	8508.68	126	61	0	<0.001
	Right Entorhinal Area	187	104.02	139.22	-302.31	625.61	127	60	0	<0.001
	Left Entorhinal Area	187	111.47	130.32	-339.51	484.47	131	56	0	<0.001
Abnormal Tissue Volume (cc ³)	Brain	186	48.74	1492.17	-1996.99	13787.27	50	136	0	<0.001
	Gray Matter	186	-66.45	145.73	-817.30	477.00	17	169	0	<0.001
	White Matter	186	115.19	1446.94	-1477.24	13595.58	56	130	0	<0.001
	Cerebrospinal Fluid	186	-90.18	175.03	-913.62	904.00	12	174	0	<0.001
	Right Hippocampus	186	-7.42	20.49	-118.26	47.00	0	186	0	<0.001
	Left Hippocampus	186	-5.15	13.47	-67.71	28.00	0	186	0	<0.001
	Right Medial Frontal Cortex	186	-0.03	0.36	-3.81	1.00	0	186	0	<0.001
	Left Medial Frontal Cortex	186	-0.08	0.73	-6.68	2.00	0	186	0	<0.001
	Right Anterior Cingulate Gyrus	186	-13.51	53.40	-528.34	150.68	2	184	0	<0.001
	Left Anterior Cingulate Gyrus	186	-5.88	29.85	-219.35	82.00	2	184	0	<0.001
	Temporal Lobe WM Right	186	16.17	160.56	-282.29	1884.46	26	160	0	<0.001
	Temporal Lobe WM Left	186	7.25	189.94	-260.35	2060.89	12	173	1	<0.001
	Right Entorhinal Area	186	-0.23	3.44	-45.78	5.72	0	186	0	<0.001
	Left Entorhinal Area	186	-0.77	4.55	-38.15	6.00	0	186	0	<0.001

(Table Continued)

		Change (Y30 - Y25) with participants have both Y25 and Y30					Wilcoxon signed-rank test			
		N	Mean	SD	Min	Max	positive	negative	zero	Prob > z
(Corrected Mean)	Brain	183	-0.01	0.01	-0.07	0.01	0	183	0	<0.001
Fractional Anisotropy (Normalized)	Gray Matter	183	-0.02	0.02	-0.10	0.01	0	183	0	<0.001
	White Matter	183	-0.01	0.01	-0.05	0.02	0	183	0	<0.001
	Cerebrospinal fluid	183	-0.02	0.02	-0.10	0.03	0	183	0	<0.001
	Right Hippocampus	183	-0.01	0.01	-0.11	0.02	0	183	0	<0.001
	Left Hippocampus	183	-0.01	0.01	-0.14	0.02	0	183	0	<0.001
	Right Medial Frontal Cortex	183	-0.01	0.02	-0.07	0.03	0	183	0	<0.001
	Left Medial Frontal Cortex	183	-0.02	0.02	-0.06	0.03	0	183	0	<0.001
	Right Anterior Cingulate Gyrus	183	-0.01	0.02	-0.10	0.04	0	183	0	<0.001
	Left Anterior Cingulate Gyrus	183	-0.01	0.02	-0.07	0.03	0	183	0	<0.001
	Temporal Lobe WM Right	183	0.00	0.01	-0.04	0.04	0	183	0	<0.001
	Temporal Lobe WM Left	183	-0.01	0.02	-0.05	0.04	0	183	0	<0.001
	Right Entorhinal Area	183	-0.01	0.02	-0.11	0.05	0	183	0	<0.001
	Left Entorhinal Area	183	-0.01	0.02	-0.11	0.03	0	183	0	<0.001
(Mean) Cerebral Blood Flow (mm ³ /100gms/min)	Brain	109	-2.18	8.12	-17.47	41.87	0	109	0	<0.001
	Gray Matter	109	-2.34	8.68	-18.09	44.71	0	109	0	<0.001
	White Matter	109	-1.94	7.66	-16.72	38.14	0	109	0	<0.001
	Right Hippocampus	109	-4.40	11.07	-46.57	37.51	0	109	0	<0.001
	Left Hippocampus	109	-2.96	10.56	-29.78	41.13	0	109	0	<0.001
	Right Medial Frontal Cortex	109	-6.64	19.24	-90.33	56.81	1	108	0	<0.001
	Left Medial Frontal Cortex	109	-3.94	18.05	-68.46	44.30	0	109	0	<0.001
	Right Anterior Cingulate Gyrus	109	-8.54	16.24	-78.90	42.53	0	109	0	<0.001
	Left Anterior Cingulate Gyrus	109	-6.70	14.43	-59.89	41.54	0	109	0	<0.001
	Temporal Lobe WM Right	109	-3.05	8.39	-19.09	40.22	0	109	0	<0.001
	Temporal Lobe WM Left	109	-1.68	8.22	-17.25	36.39	0	109	0	<0.001
	Right Entorhinal Area	109	-4.35	12.39	-34.56	31.47	0	109	0	<0.001
	Left Entorhinal Area	109	-6.73	14.44	-38.25	40.31	0	109	0	<0.001

(Table Continued)

		Change (Y30 - Y25) with participants have both Y25 and Y30					Wilcoxon signed-rank test			
		N	Mean	SD	Min	Max	positive	negative	zero	Prob > z
Number of activated voxels	Right Hippocampus	175	-174.65	1468.29	-4412.00	4228.00	67	108	0	0.022
	Left Hippocampus	175	-175.77	1348.53	-4056.00	3674.00	79	96	0	0.036
	Right Medial Frontal Cortex	175	-171.34	686.20	-1879	1856	58	116	1	<0.001
	Left Medial Frontal Cortex	175	-150.02	640.02	-1924	1753	58	117	0	<0.001
	Right Anterior Cingulate Gyrus	175	-392.64	1524.64	-5669	4090	57	118	0	<0.001
	Left Anterior Cingulate Gyrus	175	-402.97	1771.96	-5828	4720	64	111	0	<0.001
	Temporal Lobe WM Right	175	-2029.45	18488.33	-53294	65872	73	102	0	0.054
	Temporal Lobe WM Left	175	-1138.31	17915.35	-50487	61102	86	89	0	0.370
	Right Entorhinal Area	175	-1.94	713.73	-2463	2143	85	90	0	0.402
	Left Entorhinal Area	175	-37.62	660.34	-1751	1807	73	102	0	0.044
(Mean) Percent change of vascular reactivity	Right Hippocampus	152	0.01	0.45	-1.02	2.38	0	152	0	<0.001
	Left Hippocampus	149	0.02	0.43	-1.10	1.19	0	149	0	<0.001
	Right Medial Frontal Cortex	144	0.09	1.58	-6.36	10.60	0	144	0	<0.001
	Left Medial Frontal Cortex	142	0.002	1.31	-3.58	5.70	0	142	0	<0.001
	Right Anterior Cingulate Gyrus	148	-0.11	0.37	-1.18	0.90	0	148	0	<0.001
	Left Anterior Cingulate Gyrus	148	-0.11	0.37	-1.21	0.87	0	148	0	<0.001
	Temporal Lobe WM Right	160	0.001	0.35	-0.76	1.19	0	160	0	<0.001
	Temporal Lobe WM Left	162	0.04	0.35	-0.73	1.39	0	162	0	<0.001
	Right Entorhinal Area	141	0.01	1.14	-3.29	3.56	0	141	0	<0.001
	Left Entorhinal Area	148	0.12	1.07	-3.18	3.00	0	148	0	<0.001

1-6. Multi-analytical platforms used in metabolic phenotyping at Year 30

High-resolution NMR spectroscopy is a quantitative non-destructive and non-invasive technique that provides detailed information on solution-state molecular structures and motion which enables the detection of a wide range of structurally diverse metabolites simultaneously at a particular time point [146, 147]. NMR spectroscopy exploits the ability of spin active nuclei to absorb and re-emit pulsed electromagnetic radiation of a characteristic frequency pattern when placed in a magnetic field. It is highly reproducible, and gives atom-centred connectivity information; however it is relatively insensitive (mitigated by use of cryoprobes and high magnetic fields) [148].

The urinary metabolite data in this study were acquired by high-resolution NMR spectroscopy which was quantified by a software tool (the NMR Bruker AVANCETM 600 MHz Spectrometer, Bruker Biospin GmbH) developed by Bruker Corporation. The amino acids and tryptophan pathway metabolites were measured by ACQUITY ultra-performance liquid chromatography (UPLC) unit interfaced with a Waters Xevo TQ-S Triple Quadrupole MS.

Method of NMR spectroscopy

Among 600 participants, urine samples were collected from 425 participants during June and December 2010 and from 160 participant during January and August 2011 (Year 25); 562 and 12 samples were collected before and after midday, respectively. The experiments were carried out in the MRC-NIHR National Phenome Centre. Samples were frozen after aliquots (which are directly from specimen collection cups) were completed⁵. The collected urine was spun down and aliquoted into 2 mL cryovials. Samples were stored frozen at -80°C until analysed, and otherwise kept at 4°C . Each urine sample was centrifuged at 12000g at 4°C for 5 minutes. Urine samples were prepared into well plates for NMR spectroscopy using a Bruker Sample Track system and a Liquid Handler 215 preparation robot or similar [146]. Eppendorf tubes containing 600 μL of sample were centrifuged at 12000g for 5 min at 4°C , and 540 μL of the supernatant was transferred into a SampleJet NMR tube along with 60 μL of urine buffer and mixed well. The final concentration of sodium azide in urine samples were 0.2 mM after preparation [146].

MS ‘weighs’ molecules by separating ions according to the mass to charge ratio and in high-resolution MS one can obtain the exact mass in atomic mass units (amu), and possible molecular formulae can be narrowed down with the masses of individual common elements [149]. Currently, UPLC–MS is used in general, which uses chromatographic columns packed with small particles (1.7 μm) to allow the use of ultra-high pressure elution with improved chromatographic separation and reproducibility [147, 148].

⁵ <https://www.cardia.dopm.uab.edu/images/more/pdf/Year25/CARDIA/Form51.pdf>

Method of UPLC–MS

Although it has much lower detection limits, MS is destructive and a more targeted approach which is often needed with prior separation of metabolites, using either chromatography or capillary electrophoresis [150]. MS approaches known to be less reproducible, more platform dependent and susceptible to variability [151]. As it is used as a molecular separation phase before MS detection, LC, particularly UPLC, provides rapid analysis and delivers excellent chromatographic resolution; it is being used increasingly for metabolic profiling [151, 152]. The UPLC–MS approaches and orthogonal acceleration time-of-flight MS systems for urine analysis were conducted for amino acids and tryptophan pathway metabolites [92]. The retention time precision of each chromatographic method was assessed using the method reference standards and internal standards within the QC samples of each study as exemplary molecular targets [92]. Urine represents time-averaged data stored in the bladder; it has the physiological characteristics including high enzyme activities in pathological states and a dynamically reactive matrix [151]. System robustness is a key factor in large scale phenotyping applications to reduce or correct batch effects (technical sources of variation). In regard to this matter, the utility of data correction techniques throughout each experiment, automatic control of the voltage applied to the MS detector was used. In addition, feature specific LC–MS data correction procedures such as curve fitting and locally estimated scatterplot smoothing (LOESS) regression were applied and expected to improve further data precision [92, 153, 154]; no obvious batch effects were observed in exploratory analysis of the data.

The normalisation process for the urinary metabolite data with verification were conducted in the next section. The number of missing metabolites, and how many missing cases would be used for the analyses were examined in the following sections.

2. Metabolomics

Metabolomics is the comprehensive study of metabolites present in a given biological system such as biofluid and tissue [155, 156]. It is also defined as the systematic study of the unique chemical fingerprints that specific cellular processes leave behind [157].

In this study metabolomics was used as a global biochemical approach to examine urinary metabolic concentration in middle-aged study population and correlate them to cardio-metabolic risk factors, and cognitive performance.

2-1. Normalisation of urinary metabolites at Year 30

Normalisation in bioscience denotes a method applied to biological datasets to correct for systematic technical errors and sometimes irrelevant biological differences in peak intensity differences and to stabilize variance [158]. The urinary metabolite data acquired by high-resolution NMR spectroscopy, quantified by a software tool developed by Bruker Corporation, also needs to be normalized to remove

the effect of dilution on the samples when the variance in the total integrated intensity is large. The urinary metabolic data by LC–MS should also be normalized to take account of dilution of urine samples and other effects. If the formats of the data are concentration of a metabolite or peak intensities, normalisation can be applied in the same way as NMR data. A conventional creatinine-based normalisation is one technique used to remove the effect of dilution, dividing all the concentrations of the metabolites by the level of creatinine. However, the creatinine itself can be affected by a number of patient characteristics such as age, race, gender, physical activity, muscle mass, and (ab)normal physiological functions [159]. Another approach is probabilistic quotient normalisation (PQN) and was developed specifically for metabolomics research to reduce unwanted peak effects [160]. The PQN relies on the assumption that at least 50% of the metabolites in the profile are affected by only the dilution of the samples, i.e. not affected by other factors [161]. However, it becomes difficult to meet the assumption as the number of metabolites becomes smaller. One way to check this assumption would be to examine the changes and the patterns of the total integrated intensity before and after the normalisation.

One of the methods to determine the optimal post-analytical normalisation approach for (untargeted) metabolomics is to calculate the relative standard deviation (RSD) of pooled quality control samples (QCs) [160]. As a measure of feature reproducibility across QCs, the RSD can be calculated by dividing the standard deviation (SD) of the peak area by the mean peak area across the QCs, i.e. $100 \times \frac{\text{SD of QCs}}{\text{Mean of QCs}}$. The ideal RSD is less than 30% [160].

Process of the probabilistic quotient normalisation

The 47 metabolites acquired by NMR spectroscopy (mmol/L) were normalized using the PQN method. The final stage data is zero-inflated by undetected (i.e. non-quantified) metabolite concentrations, and it seems to be difficult to meet the assumption of the PQN. Therefore, the high resolution one-dimensional (1D) NMR spectroscopy data which consists of the intensity values for each spectrum in each point, was used to calculate the ‘normalisation factor’ applied to the individual’s metabolite profile.

The 28 amino acids measured by LC–MS (μM) were also normalized using the PQN method. As the same subjects’ metabolites were identified by the NMR and LC–MS spectrometry, high resolution 1D NMR spectroscopy data can be used to calculate the ‘normalisation factor’ applied to the individual’s amino acids profile. The quantified final stage amino acids data can be used to calculate the ‘normalisation factor’ applied to the individual’s metabolite profile, unless it is not inflated by undetected metabolite concentrations and the number of metabolites is not small.

The process of the PQN using the 1D pre-saturation NMR spectroscopy data is as follows [162] and the detailed process are presented in the Appendix Table 8:

1. Defining a reference intensity for each metabolite, i.e. calculation of the median of each metabolite.
2. Calculation of the ratio for each metabolite, dividing the original metabolite value by the median of the metabolite.
3. Calculation of the median of these ratios per individual. This is a 'normalisation factor'.
4. Applying the 'normalisation factor' to the individual's profile, i.e. dividing the original metabolite value (mmol/L) by a 'normalisation factor' per individual, using the final stage metabolite data.

The process of the PQN using the final stage zero-inflated quantified data is the same process, applying a 'normalisation factor' from the final stage metabolite data to the individual's profile. In the next section, I investigate the range of normalisation methods, in order to select the optimal method for use in the study.

2-2. Normalisation verification

In order to check if at least 50% of the metabolites in the profile are affected by only the dilution of the samples, I investigated the changes and the patterns of the total integrated intensity by each normalisation method. I investigated score plots from the principal component analysis (PCA), and an individual's total metabolite intensities matched to the score plots by normalisation methods (Figure 28 and Figure 29). In addition, the RSD of the pooled QCs was calculated as a measure of feature reproducibility across QCs, as a reference of an optimal post-analytical normalisation method. The orthogonal varimax rotation was conducted after standardizing the data for PCA.

Figure 28 [A1, A2] shows score plots using the 1D NMR spectroscopy data without application of normalisation methods. Figure 28 [A1] shows the dispersed distribution of the scores of metabolites across the 1st and 2nd principal components. The QCs (red colour dots) were widely dispersed across the 1st principal component. Figure 28 [A2] is an individual's total metabolite intensities matched to the score plot. The light blue colour dots indicate the lowest and the purple colour dots indicate the highest intensities; ranges were from 0 to 34.99 for the metabolite concentrations. Among 47 metabolites, only 10 metabolites were included in PCA as they showed less than 50% missing cases (due to the insensitivity of NMR to detect all the metabolites) in the total study population (N=606).

Figure 28 [B1, B2] illustrates the NMR spectroscopy data, normalized by 'PQN-normalized creatinine' (using 1D NMR spectroscopy data). The metabolite scores were less dispersed across the 2nd principal components in comparison to the non-normalized data. Although the QCs were still widely dispersed across the 1st principal component, some subjects with high intensities (i.e. the purple colour dots) moved into the centre and were tightly clustered across the 1st principal component after the creatinine normalisation.

Figure 28 [C1, C2] shows the NMR data, normalized using PQN factors derived from the high resolution 1D NMR spectroscopy data. The metabolites were less dispersed across the 1st and 2nd principal components in comparison with the non-normalized data. The QCs were less dispersed in comparison with the non-normalized

data. Including some distinctive outliers, many subjects with high intensities (i.e. the purple colour dots) moved into the centre of the 1st and 2nd principal components.

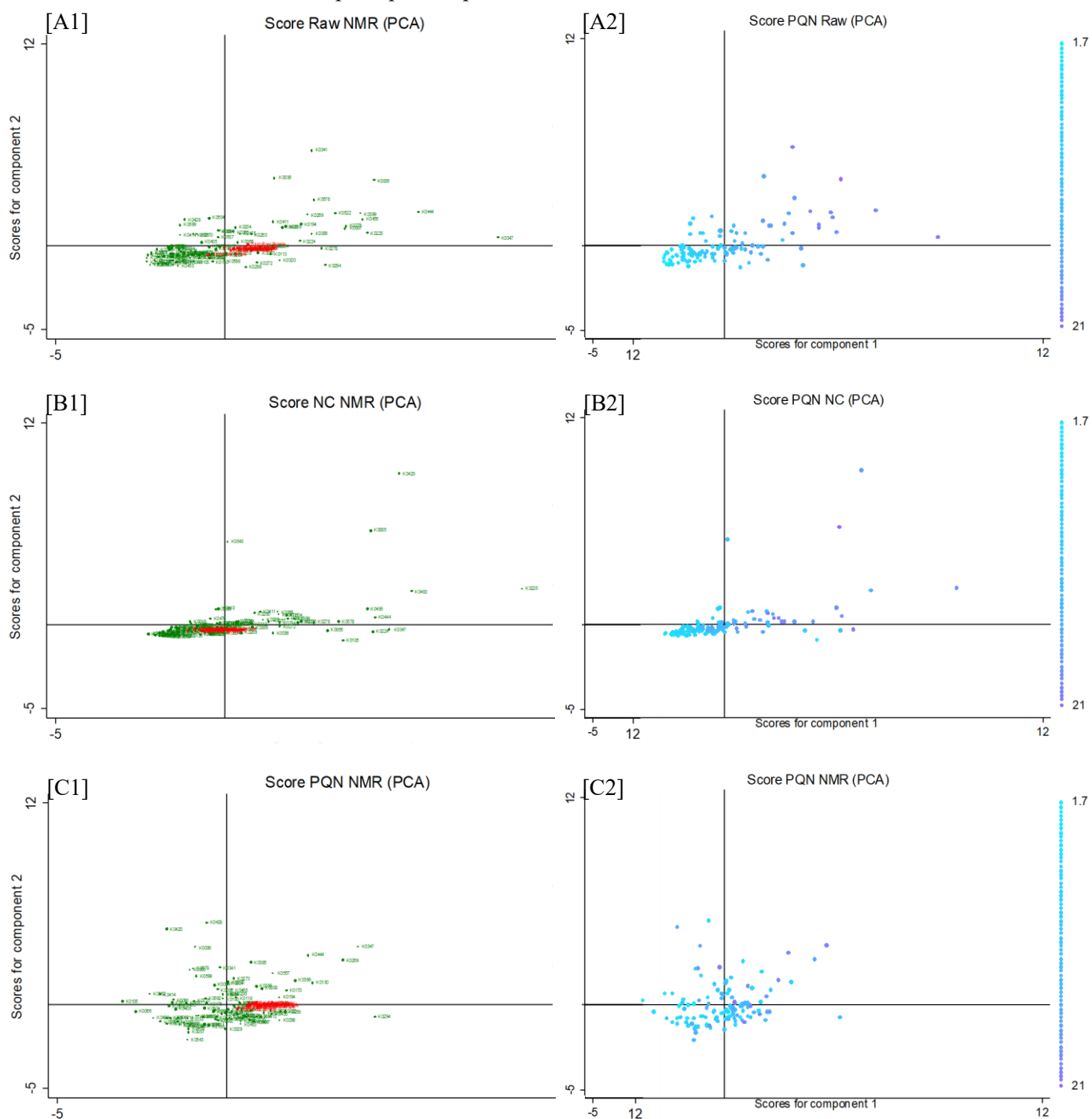


Figure 28 The NMR spectroscopy data - Score plots with [1] sample types (left), and [2] the total intensity (right).

The colour scales indicate the total intensity of individual concentration; ranges from 0 to 34.99.

[A1, A2] are the NMR data quantified by Bruker Corporation, without application of normalisation

[B1, B2] are the NMR spectroscopy data, normalized by PQN-normalized creatinine (using the 1D NMR spectroscopy data)

[C1, C2] are the PQN NMR data, normalized using the 1D NMR data

Note: Acetone was removed for the PCA as its value was not observed in the QCs

The SD of the total integrated intensity of the NMR metabolite concentrations was 5.83 (mean 9.26). Although variance in the total integrated intensity was not relatively large, normalisation was conducted to remove the probable effect of dilution. The RSD of total integrated 10 NMR metabolites in the QC samples were calculated in the Table 10. The PQN method [C] showed the smallest RSD.

Table 10 The relative standard deviation (RSD) of 10 NMR metabolites in the QC samples (N=15)

	RSD	Mean of QCs	SD of QCs
[A] Without application of normalisation	199.1	1.1	2.1
[B] Normalized by PQN-normalized creatinine	199.6	0.2	0.3
[C] PQN	170.2	1.9	3.2

Figure 29 [A1, A2] shows score plots of the 24 amino acids metabolite measured by LC–MS without application of normalisation methods. Among the 28 amino acids, 4 metabolites showed more than 50% of the Lower Limit of Quantification (LLOQ) in the total study population (N=606), and therefore 24 amino acids were included in PCA. Figure 29 [A1] shows dispersed distribution of the scores of metabolites across the 1st principal component. Figure 29 [A2] is an individual's total metabolite intensities matched to the score plot, ranges were from 0 to 17496.94 for the amino acid concentrations. The subjects with high intensities (i.e. the purple colour dots) were consistently located on the right part of the 1st principal component.

Figure 29 [B1, B2] illustrates score plots of the data, normalized by PQN-normalized creatinine (using the 1D NMR spectroscopy data). The metabolites were less dispersed across the 2nd principal component in comparison with the non-normalized data. Some subjects with high intensities moved into the centre and was tightly clustered across the 1st principal component after creatinine normalisation.

Figure 29 [C1, C2] shows score plots of the data, normalized using the final stage amino acids data which is consist of 28 metabolites (μM). The metabolite scores were cantered in the principal components in comparison with the non-normalized data and the creatinine normalisation data. Including some distinctive outliers, many subjects with high intensities moved into the centre of the 1st and 2nd principal components. Metabolites were noticeably located in the centre compactly including the QCs.

Figure 29 [D1, D2] shows score plots of the data, normalized using the 1D NMR spectroscopy data. The metabolite scores were dispersed across the 1st and 2nd principal components. Including some distinctive outliers, some subjects with high intensities moved into the centre and the 2nd principal component. The metabolite scores were rather sparsely located, but they were more dispersed into both the 1st and 2nd principal components.

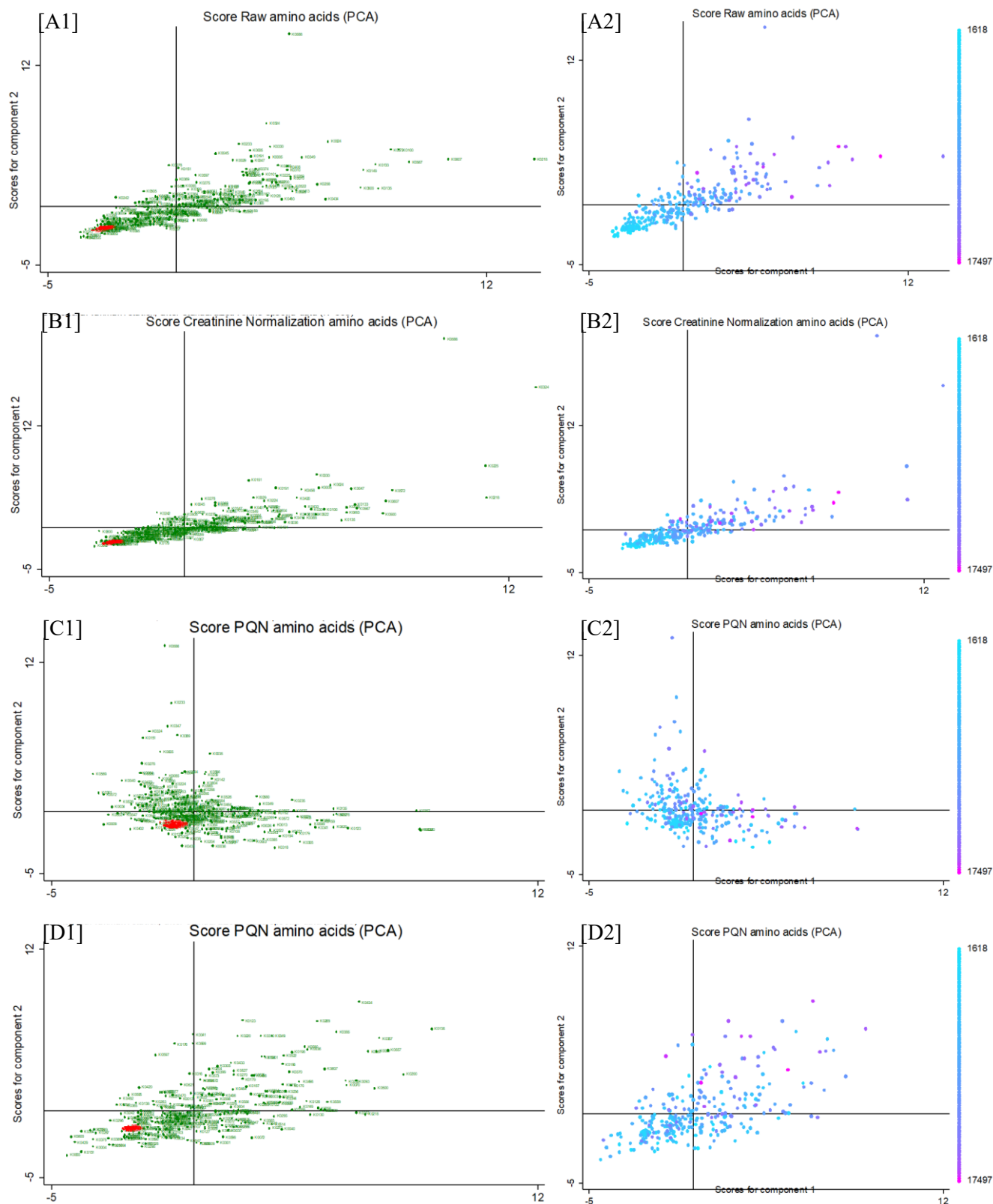


Figure 29 The amino acids by LC-MS data - Score plots from [1] PCA (left), and [2] the total intensity (right). The colour scales indicate the total intensity of each individual concentration; ranges from 0 to 17496.94.

[A1, A2] are the amino acids (μM) without application of normalisation

[B1, B2] are the amino acids, normalized by PQN-normalized creatinine

[C1, C2] are the PQN amino acids, normalized using the amino acids data consist of 28 metabolites (μM)

[D1, D2] are the PQN amino acids, normalized using the 1D NMR data

Note: Urine id K0233, K0541 in the [B] and [D] do not have creatinine values

The SD of the total integrated intensity of the amino acid concentrations was 2626.53 (mean 3807.41). Normalisation seemed to be needed for the disperse pattern of the raw data. The RSD of total integrated 24 amino acids in QC samples were calculated in Table 11. The PQN [C] has the smaller RSD, but the numbers were similar across the methods. There was no 1D NMR spectroscopy data for the QCs. For the PQN [C], the final stage amino acids data was used only for QCs normalisation. Therefore, the direct comparison of RSD for the PQN [C] and [D] was not possible.

Table 11 The relative standard deviation (RSD) of 24 amino acids in the QC samples for (N=77)

	RSD	Mean of QCs	SD of QCs
A Without application of normalisation	129.0	102.6	132.3
B Normalisation using the PQN creatinine	129.0	15.1	19.5
C PQN using the final stage amino acids data	128.7	117.4	151.1
D* PQN using the 1D NMR data	128.7	117.4	151.1

* The RSD was the same as C. There was no 1D NMR data for the QCs.

Therefore, the final stage amino acids data was used only for QCs normalisation in this case.

In summary, in terms of the normalisation method, the PQN method showed a smaller RSD for both assays (the NMR and LC–MS), although there was not much difference in the methods for the LC–MS data (Figure 29).

In the next section, I investigate the concordance on the identical metabolites in both assays with regard to selecting the optimal normalisation method. This investigation was particularly informative in terms of selecting PQN methods for the LC–MS data between the method of using the final stage amino acids data and the method of using the 1D NMR spectroscopy data.

Identical urinary metabolites assayed by both nuclear magnetic resonance (NMR) and liquid-chromatography mass spectrometry (LC–MS)

Alanine, glycine, valine, sarcosine and methionine were measured by both NMR and LC–MS. The measurements of these metabolites in the two assays should show good concordance. Alanine, glycine and valine showed relatively good concordance in both assays without the application of normalisation methods [1]. They also showed relatively good concordance in both assays after applying the PQN creatinine normalisation [2]. However, methionine was excluded in the correlation analyses as its concentration was all non-quantified by NMR. Sarcosine had few data points for correlations to be reliable. Among the PQN applied outcomes, normalisation using the 1D NMR spectroscopy data [4] showed the better concordance in comparison with the method using the final stage amino acids data [3] as shown in Table 12 and Figure 30. The PQN method [4] led to a decreased concordance for valine (i.e. Spearman's rank correlation coefficient (ρ) from 0.9 with the data without normalisation [1] to 0.64), but there were no large changes in alanine and glycine, after applying the method [4]. Valine by NMR exhibited 33.3% undetected concentration, and on the other hand, 0.5% undetected concentration by the LC–MS data. It seems the missing data percentage influences the PQN method [4].

Table 12 Spearman's rank correlation coefficient of metabolites measured by both NMR and LC–MS

		Alanine (LC–MS)				Glycine (LC–MS)				Valine (LC–MS)				Sarcosine (LC–MS)			
		[1]	[2]	[3]	[4]	[1]	[2]	[3]	[4]	[1]	[2]	[3]	[4]	[1]	[2]	[3]	[4]
Alanine (NMR)	Corr.	0.97	0.96	0.69	0.90	0.69	0.63	0.03	0.32	0.86	0.83	0.03	0.57	0.31	0.38	-0.26	0.05
	N	501				501				499				95			
Glycine (NMR)	Corr.	0.74	0.68	0.09	0.40	0.98	0.97	0.82	0.96	0.75	0.69	-0.25	0.34	0.23	0.25	-0.26	0.08
	N	546				546				544				96			
Valine (NMR)	Corr.	0.74	0.69	0.16	0.41	0.57	0.53	0.02	0.26	0.89	0.86	0.24	0.64	0.30	0.40	-0.38	-0.02
	N	402				402				401				90			
Sarcosine (NMR)	Corr.	0.55	0.66	-0.24	-0.16	0.40	0.44	0.11	0.00	0.55	0.62	-0.10	-0.12	-0.36	0.50	-0.10	-0.90
	N	23				23				23				5			

[1] The data without application of normalisation (with standardisation)

[2] Normalisation by the PQN Creatinine method (with standardisation)

[3] PQN using the 1D NMR data for NMR, and PQN using the final stage amino acids data for LC–MS (with standardisation)

[4] PQN, using the 1D NMR data for both NMR and LC–MS (with standardisation)

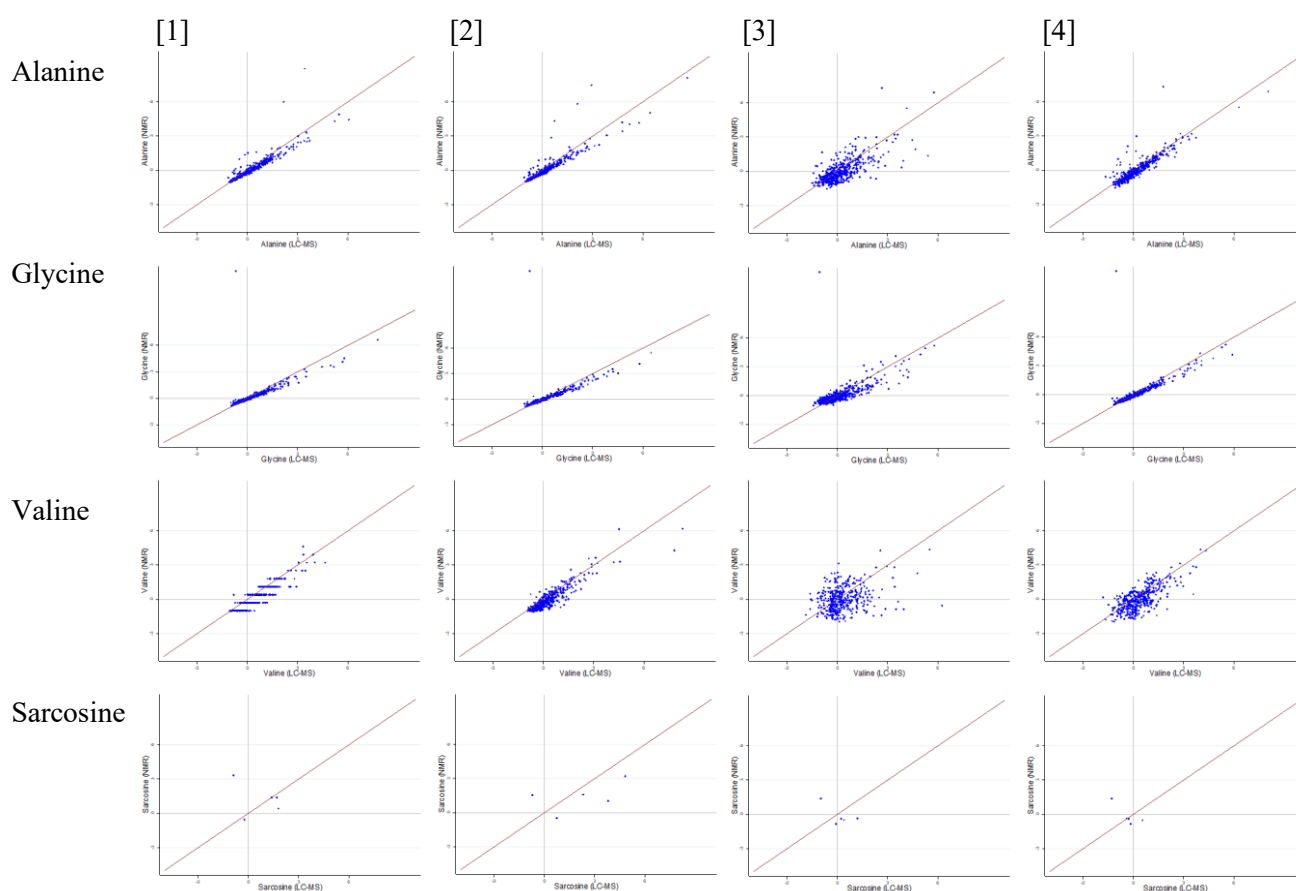


Figure 30 Metabolites measured by both NMR and LC–MS

[1] The data without application of normalisation (with standardisation)

[2] Normalisation by the PQN Creatinine (with standardisation)

[3] PQN, using the 1D NMR data for NMR, and PQN using the final stage amino acids data for LC–MS (with standardisation)

[4] PQN, using the 1D NMR data for both NMR and LC–MS (with standardisation)

PQN for the amino acids LC–MS data could be conducted using the final stage amino acids data itself, and the method noticeably led many subjects with high intensities to move to the centre of the 1st and 2nd principal components. The method seems to be effective to adjust the peak intensity differences and to stabilize variance including probable effect of dilution. Nevertheless, in terms of the concordance of the normalized metabolites

across both assays, the method of using the 1D NMR data for both NMR and LC–MS data was used for further analyses.

2-3. The NMR-based urinary metabolites

Targeted urinary metabolites by NMR and LC–MS were measured from about 600 participants in the present study. Spot urine samples were collected on the day participants visited the centres in the US. Many studies support that the method of spot urine collections was comparable for clinical end-points (outcomes), and it can be used in clinical practice or research as an alternative for the 24-hour urinary collections despite some limitations [163, 164]. In the present study, the high-resolution NMR spectroscopy data (mmol/L) of the 600 participants are available. Among 47 metabolites, 10 metabolites showed less than 50% missing cases in the total study population (N=600) due to the insensitivity to quantify the metabolites, ranges from no missing cases for creatinine to 37% of missing cases for acetone (Table 13). In addition, the Upset plot (an alternative to the Venn diagram) for missing cases in the 10 metabolites are presented in Appendix Figure 3. The upper triangle of a Spearman's rank correlation coefficient matrix is presented in Appendix Figure 4; glycine and citric acid (0.48), glycine and alanine (0.41), alanine and valine (0.42), acetic acid and succinic acid (0.41), creatinine and acetone (-0.44), creatinine and acetic acid (-0.38).

Table 13 The high-resolution NMR spectroscopy data (mmol/L)

	Missing N (%) [*]	Data quantified, without application of normalisation					PQN using the 1D NMR data			
		N	Mean	SD	Min	Max	Mean	SD	Min	Max
Creatinine	0 (0)	600	6.59	4.39	0.500	26.00	5.96	1.68	2.15	13.70
Citric acid	11 (1.8)	589	1.71	1.27	0.090	6.50	1.66	0.82	0.09	4.91
Glycine	56 (9.3)	544	0.63	0.72	0.040	11.00	0.60	0.50	0.09	6.89
Alanine	100 (16.7)	500	0.14	0.13	0.010	1.30	0.13	0.07	0.02	0.62
Acetic acid	127 (21.2)	473	0.06	0.12	0.010	1.40	0.06	0.10	0.02	1.59
Succinic acid	139 (23.2)	461	0.05	0.04	0.010	0.39	0.06	0.04	0.01	0.37
Oxaloacetic acid	192 (32)	408	0.27	0.49	0.010	7.60	0.23	0.30	0.05	4.76
Valine	200 (33.3)	400	0.02	0.01	0.010	0.09	0.02	0.01	0.01	0.05
N,N-Dimethylglycine	214 (35.7)	386	0.04	0.03	0.010	0.17	0.04	0.02	0.02	0.14
Acetone	222 (37)	378	0.01	0.01	0.010	0.16	0.02	0.01	0.01	0.14

* Missing N indicates less than the lower limit of quantitation (LLOQ) and therefore the data is recorded as zero.

* Missing cases $\geq$ 50%: D Glucose (52.3), Hippuric acid (59), Acetoacetic acid (66.2), Proline betaine (76.8), Dimethylamine (77.8), Tartaric acid (79.3), Guanidinoacetic acid (89.5), Allopurinol (90.3), Pyruvic acid (91.7), Trigonelline (91.8), Taurine (92.7), Fumaric acid (93), Caffeine (93), Allantoin(95), Trimethylamine (96), Methylnicotinamide (96), D Mannose (96), Sarcosine(96.2), Lactic acid (96.2), Furoylglycine (96.3), Aminobutyric acid (96.7), D Mandelic acid (97.8), D Mannitol (98), Methylhistidine (98.7), Benzoic acid (99.2), Hydroxybutyric acid (99.2), D Lactose (99.2), Oxoglutaric acid (99.5), Imidazole (99.7), Inosine (99.7), Arginine (99.8), D Galactose (99.8), Methionine (100), Methylsuccinic acid (100), Methyladenosine (100), Adenosine (100), MyoInositol (100)

Concentrations of non-normalized urinary metabolites by NMR which show right skewed distributions are reported in Figure 31. The metabolites in black ethnicity show relatively dispersed distributions.

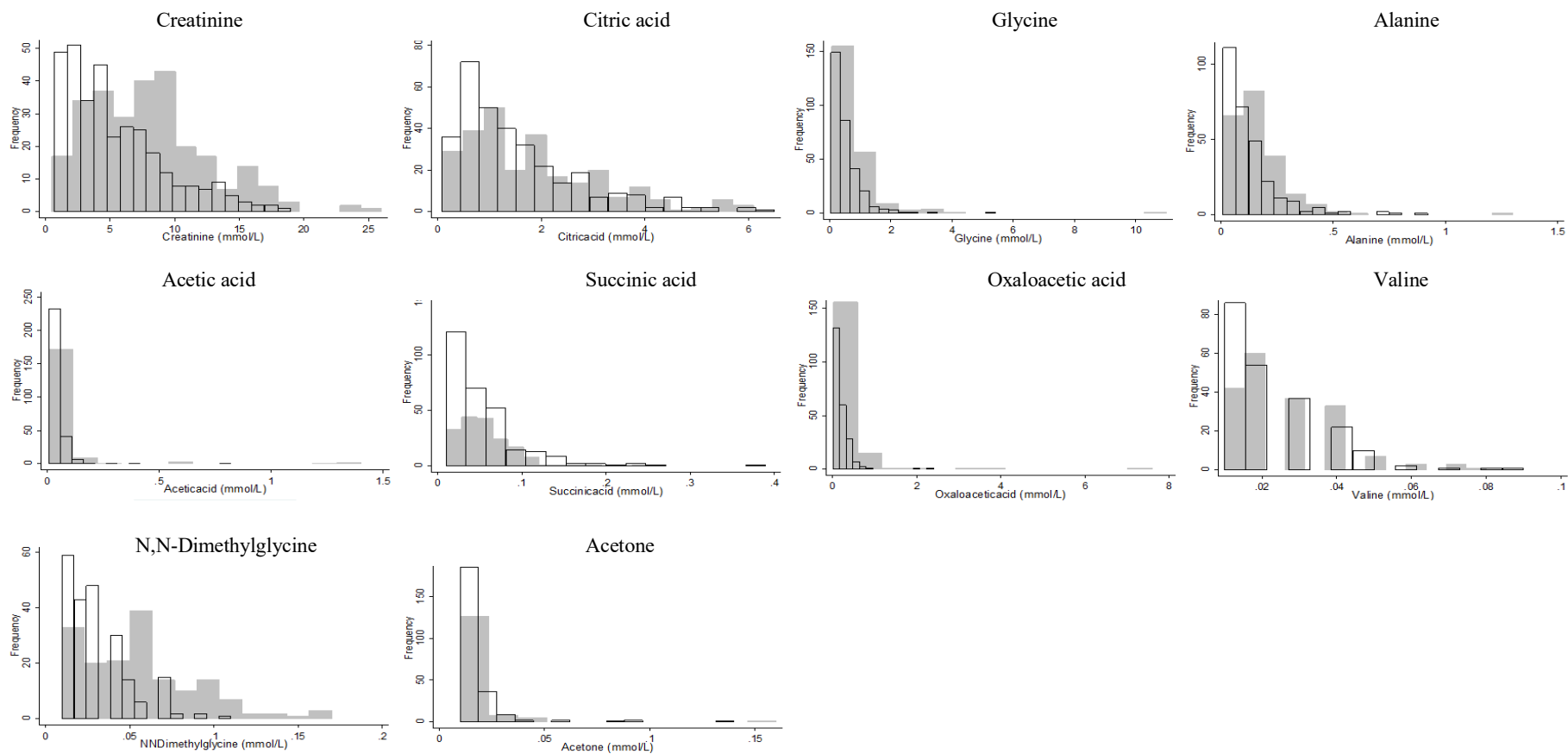


Figure 31 Concentrations of non-normalized urinary metabolites by NMR; colour indicates black (grey) and white (white) ethnicity

2-4. The LC–MS-based Amino Acids

Amino acids (μM) data of 599 participants by LC–MS are available at Year 30. The urinary amino acids are listed in Table 14. Initially, 28 amino acids were identified, and four metabolites showed more than 50% of missing cases in the study population (N (%)=600); alpha-aminobutyric acid (53.0), aspartic acid (53.2), hydroxyproline4 (72.0), sarcosine (82.7). Those four metabolites were excluded from the analyses. Among 24 amino acids, the number of individuals with up to 2 amino acids missing N (%) are 492 (82.0) or up to 9 amino acids missing are 598 (99.7); 19 missing is 1 (0.2) and 24 missing is 1 (0.2), respectively. The values indicated as below the limit of detection ($< \text{LOD}$) means we do not have the ability to quantify it accurately with the mass spectrometry assay; it doesn't necessarily mean it is not present in urine. The assay is validated to be linear over a quantification range up to $800\mu\text{M}$ for each amino acid. An extension of the linear regression plot was used to calculate the amino acids concentrations above the $800\mu\text{M}$, using the standard curve.

The histograms show the right skewed distributions of amino acids, and many cases of the concentrations are above the $800\mu\text{M}$. Normalisation seems to be required. The amino acids in black ethnicity show relatively more dispersed distributions (Figure 32). The visualised intersections of missing cases in the 24 amino acids are presented in Appendix Figure 5.

Table 14 Amino acids (μM) by LC–MS

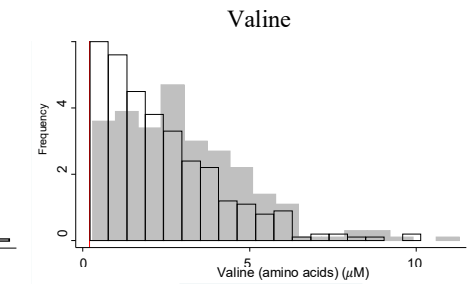
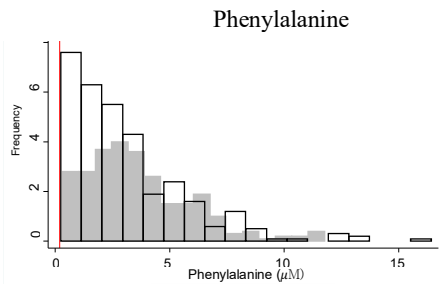
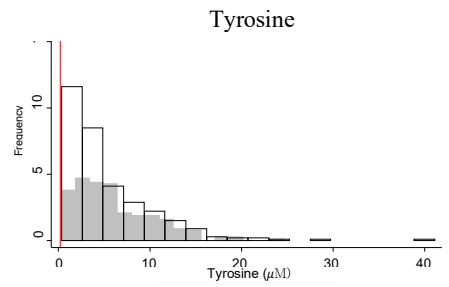
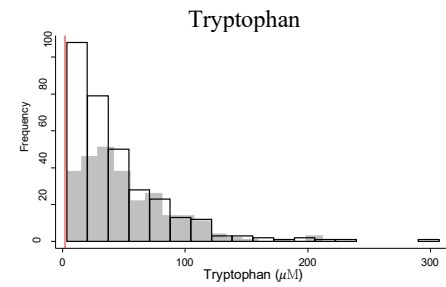
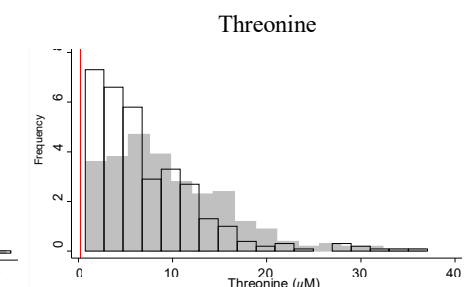
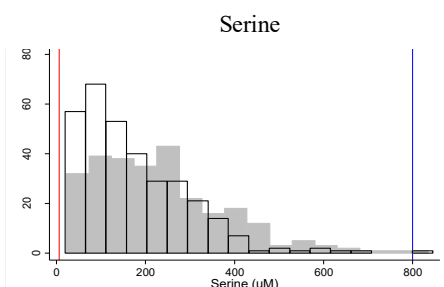
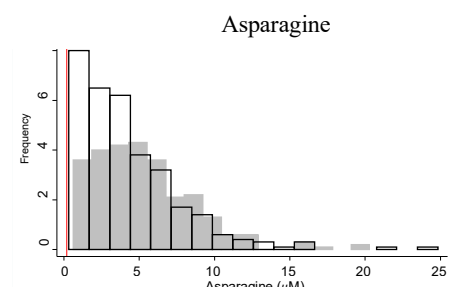
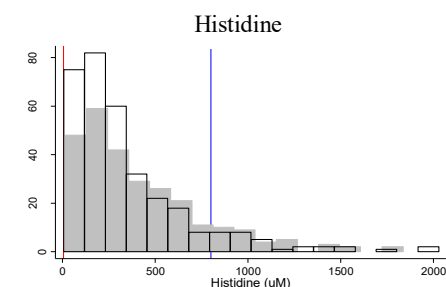
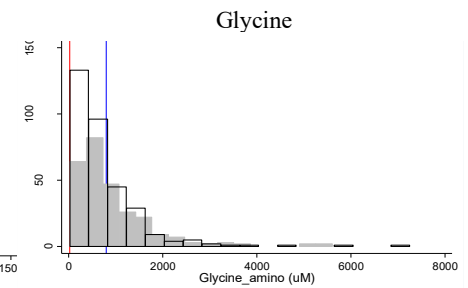
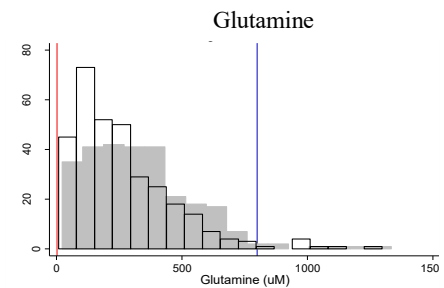
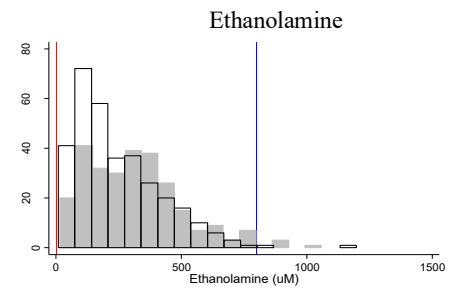
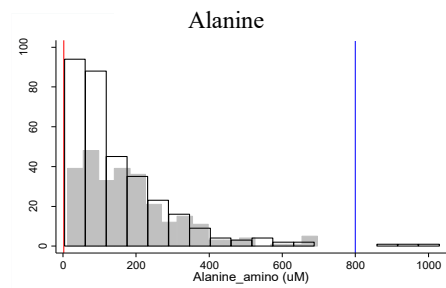
	Missing N (%)	2x <LOD*	Amino acids (μM), without application of normalisation					PQN using the 1D NMR data			
			N	Mean	SD	Min	Max	Mean	SD	Min	Max
Alanine	1 (0.2)	2	599	166.3	142.9	4.3	1029.5	152.0	88.4	6.3	862.9
Ethanolamine	1 (0.2)	2	599	275.3	182.1	9.5	1198.8	256.5	84.2	13.9	616.4
Glutamine	1 (0.2)	2	599	296.5	212.3	8.5	1335.2	278.9	120.7	12.3	933.1
Glycine	1 (0.2)	20	599	837.9	826.8	25.6	7251.6	813.1	619.1	37.3	4466.6
Histidine	1 (0.2)	6	599	376.9	335.4	6.9	2028.6	350.8	222.7	10.0	1490.2
Asparagine	2 (0.3)	2	598	47.4	34.0	3.2	248.5	44.4	19.6	9.7	188.2
Serine	2 (0.3)	6	598	201.7	139.1	19.5	845.1	189.7	74.2	13.7	523.3
Threonine	2 (0.3)	2	598	83.3	63.7	7.0	370.7	78.1	41.3	13.4	324.3
Tryptophan	2 (0.3)	2	598	47.1	39.1	2.5	307.0	42.5	21.4	7.3	161.4
Tyrosine	2 (0.3)	2	598	56.8	47.1	3.6	410.8	51.6	27.2	7.1	196.5
Phenylalanine	3 (0.5)	2	597	33.5	24.8	2.5	164.2	30.4	12.2	4.9	89.3
Valine	3 (0.5)	2	597	26.3	18.6	2.2	112.8	24.0	8.9	3.1	62.6
Hydroxylysine	3 (0.5)	20	597	217.1	151.5	20.3	946.3	196.8	63.3	34.4	617.7
Aminoadipic acid	3 (0.5)	2	597	33.5	27.5	2.0	204.2	31.6	18.9	1.1	202.2
Leucine	14 (2.3)	2	586	20.1	15.1	2.1	138.7	18.0	7.1	2.9	53.4
Lysine 244	14 (2.3)	2	586	81.6	142.6	2.9	1739.6	79.1	154.4	3.0	2728.4
1-methyl histidine	18 (3.0)	20	582	197.7	124.1	20.0	732.1	183.9	73.5	15.8	1115.6
Beta aminoisobutyric acid	26 (4.3)	6	574	134.9	272.5	6.1	2353.2	118.0	185.6	5.7	1484.6
Glutamic acid	28 (4.7)	2	572	12.2	9.5	2.0	102.0	11.2	5.4	1.7	39.3
3-Methylhistidine	28 (4.7)	20	572	666.5	781.7	20.1	6908.5	622.4	582.9	17.2	4134.9
Isoleucine	77 (12.8)	2	523	9.4	6.8	2.0	63.2	7.9	3.1	2.1	28.1
Cystathionine 282	87 (14.5)	2	513	20.6	51.4	2.0	1029.4	18.6	44.5	0.8	917.3
Methionine	140 (23.3)	2	460	7.0	4.2	2.0	31.4	5.6	1.8	1.1	12.1
Beta-alanine	297 (49.5)	6	303	10.8	8.6	4.0	56.0	9.4	6.9	1.4	42.8
Alpha-amino-n-butyric acid	318 (53.0)	6									
Aspartic acid	319 (53.2)	2									
4-hydroxyproline	432 (72.0)	20									
Sarcosine	496 (82.7)	2									

* 2x diluted limit of detection (LOD)

* The upper limit of quantitation (ULOQ) (i.e. validated concentration range) is $800 (\mu\text{M})$

* Metabolites in bold were also measured by LC–MS

The upper triangle of a Spearman's rank correlation coefficient matrix is presented in Appendix Figure 6; serine and glutamine (0.81), threonine and asparagine (0.83), threonine and glutamine (0.85), tryptophan and tyrosine (0.90), tryptophan and phenylalanine (0.89), phenylalanine and tyrosine (0.92), phenylalanine and leucine (0.82), phenylalanine and valine (0.81), isoleucine and leucine (0.91), isoleucine and valine (0.89), leucine and valine (0.93).



Black White

(Figure Continued)

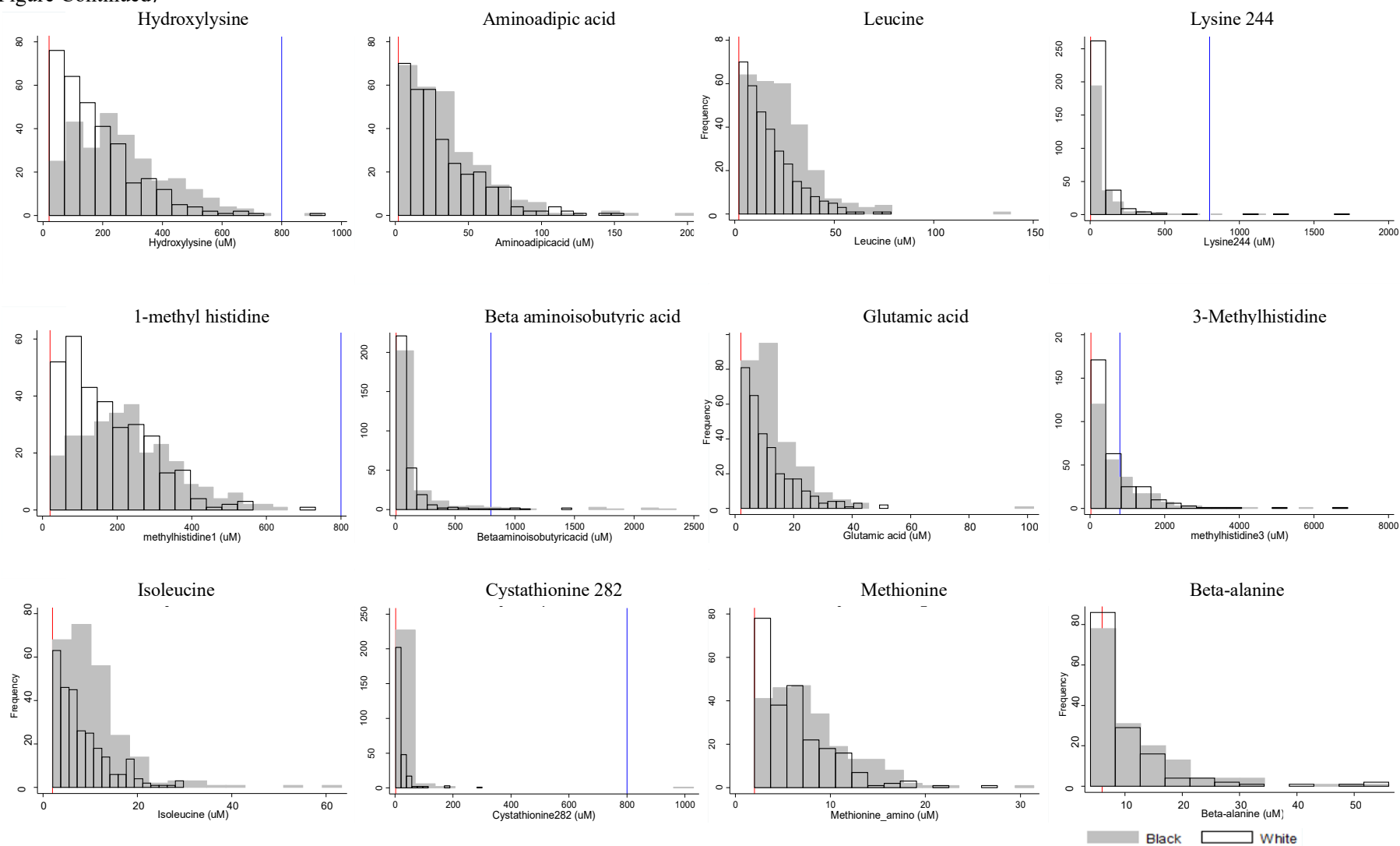


Figure 32 Concentrations of non-normalized urinary amino acids by LC-MS; red and blue lines indicate 2x<LOD and >ULOQ, respectively

2-5. The LC–MS-based tryptophan pathway metabolites

In total, fourteen targeted tryptophan metabolites measured by LC–MS were available. In total, 11 tryptophan metabolites were included for the analyses as three metabolites showed more than 50% missing cases (Table 15 and Figure 33). Among 11 tryptophan metabolites, the number of individuals with up to 2 metabolites missing N (%) are 473 (78.83) or up to 5 metabolites missing are 597 (99.5). The visualised intersections of missing cases in the 11 tryptophan metabolites are presented in Appendix Figure 7). The upper triangle of a Spearman's rank correlation coefficient matrix is presented in Appendix Figure 8; kynurenic acid and xanthurenic acid (0.54), tryptophan (positive) and kynurenine (0.55), tryptophan (positive) and 3-OH-kynurenine (0.47), tryptophan (positive) and picolinic acid (0.46), kynurenine and 3-OH-kynurenine (0.58).

Table 15 Tryptophan pathway metabolites (ng/mL) assayed by LC–MS

	Missing %	<L OD *	>ULO Q*	Tryptophan (ng/mL), without application of normalisation					PQN using the 1D NMR data			
				N	Mean	SD	Min	Max	Mean	SD	Min	Max
Picolinic acid	2 (0.3)	20	2000	598	167.5	120.8	20.1	1357.4	168.1	88.4	20.8	967.2
Neopterin	4 (0.7)	20	2000	596	338.8	260.8	28.0	1754.6	313.6	139.7	67.4	1456.2
Tryptophan (positive) ¹⁾	27 (4.5)	400	20000	573	7197.7	4765.0	466.0	19815.0	7105.8	3151.8	1477.1	20605.3
Kynurenine	31 (5.2)	20	2000	569	518.3	434.8	20.2	1957.6	507.8	356.1	51.7	2636.1
Dopamine	34 (5.7)	20	2000	566	252.2	189.1	20.8	1188.7	221.8	97.5	21.1	637.8
3-OH-kynurenine	60 (10)	20	2000	540	192.4	211.6	20.0	1640.4	169.8	148.2	18.0	1703.2
HAA (3-Hydroxyanthranilic acid)	105 (17.5)	80	2000	495	710.5	453.4	84.5	1980.7	749.8	483.7	71.6	3757.7
Indole-3-acetic acid	112 (18.7)	40	4000	488	1435.1	1008.7	81.7	3990.6	2026.8	2015.2	55.4	13491.0
Nicotinic riboside	126 (21)	20	2000	474	72.9	54.7	20.5	637.4	59.8	28.6	18.4	264.2
Kynurenic acid	198 (33)	20	2000	402	1020.0	501.6	100.5	2000.0	1351.8	451.3	204.0	3342.5
Xanthurenic acid	297 (49.5)	20	2000	303	1108.9	518.8	78.8	1976.1	1717.2	692.6	134.0	4250.0
Quinolinic acid	312 (52)	40	4000									
5-Hydroxyindoleacetic acid	361 (60.2)	40	4000									
Nicotinic acid	501 (83.5)	20	2000									

1) Detected in positive mode positive, which is a second measurement of tryptophan, and correlated with tryptophan negative mode

* Limit of detection (LOD)

* Upper limit of quantitation (ULOQ)

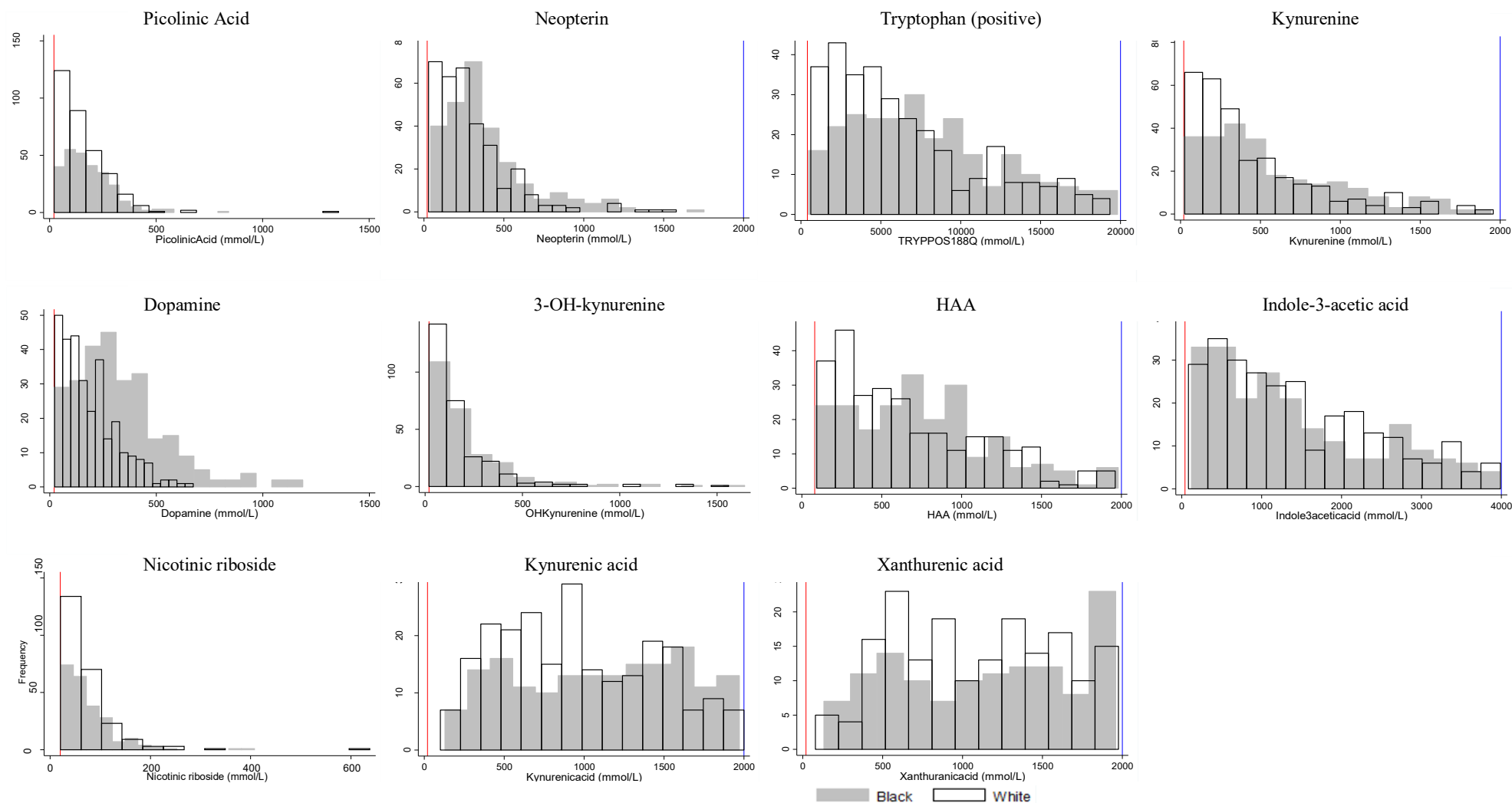


Figure 33 Concentrations of non-normalized urinary tryptophan pathway metabolites by LC-MS; red and blue lines indicate <LOD and >ULOQ, respectively

2-6. Metabolic syndrome with the metabolites

One hundred sixty-three (26.90%) participants met the metabolic syndrome criteria. This was further broken down into black ethnicity (N=104, 37.7%) and white ethnicity (N=59, 17.9%) ($p < 0.001$, Fisher's exact); male (N=77, 28.2%) and female (N=86, 25.8%).

Metabolic syndrome with the NMR-based urinary metabolites

A biplot displays the study participants (marker symbols) and the relative positions of metabolites (arrows) (Figure 34). The participants having all 10 metabolites concentrations are projected to two dimensions such that the distance between the participants is approximately preserved, and the cosine of the angle between arrows approximates the correlation between metabolites [165]. Several protective metabolites which improve cognition and neuron functioning such as **valine**, **glycine**, **citric acid**, showed large component values in dimension 1 and 2 in the non-metabolic syndrome group [166]. On the other hand, these metabolites show smaller values in the metabolic syndrome group, and **succinic acid** with a participant shows greater negative value. In addition, citric acid, creatinine, valine and N,N-Dimethylglycine are positively correlated in the study cohort according to PCA (Appendix Figure 9).

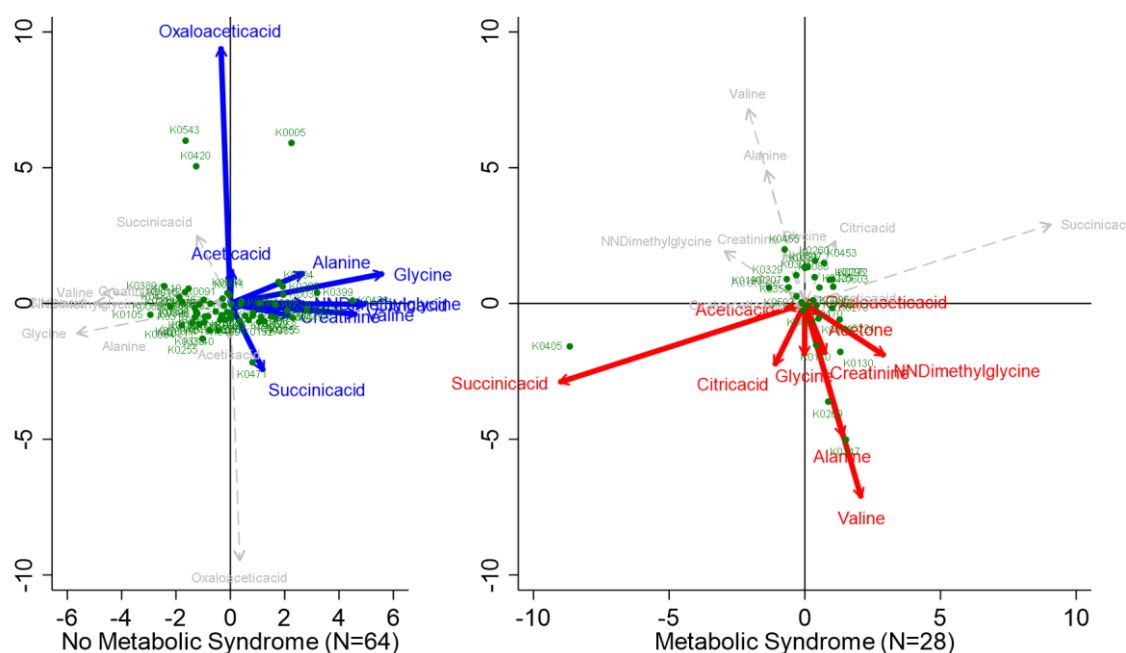


Figure 34 The biplot for (1) no metabolic syndrome and (2) metabolic syndrome (N=64 vs 28) group using the principal component analysis for the PQN-normalized NMR data; the grey arrows a reflection of negative directions

Metabolic syndrome with the LC-MS-based amino acids

The LC-MS-based amino acids are positively correlated in dimension 1 in the metabolic syndrome free group (phenylalanine, tryptophan, methionine, glutamic acid; asparagine, glutamine, threonine, and histidine show large values). On the other hand, twenty-four amino acids show large negative values in the metabolic syndrome group (Figure 35).

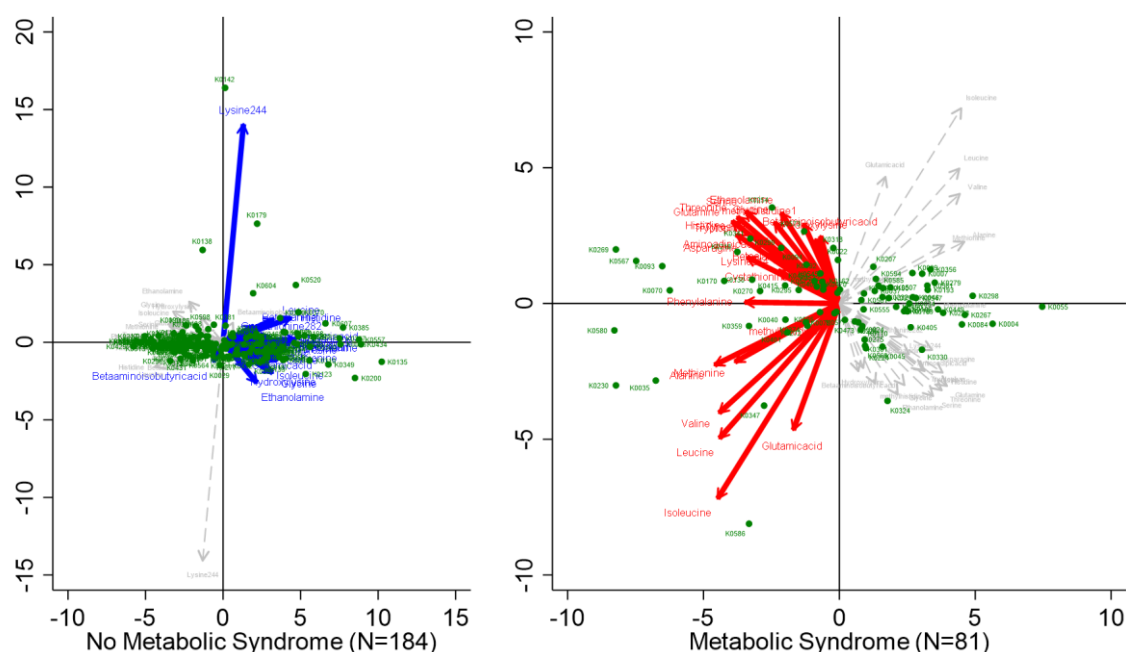


Figure 35 The biplot for (1) no metabolic syndrome and (2) metabolic syndrome (N=184 vs 81) group using the principal component analysis for the PQN-normalized amino acids

Metabolic syndrome with the LC–MS-based tryptophan pathway metabolites

The LC–MS-based tryptophan pathway metabolites are positively correlated in dimension 1 in the metabolic syndrome free group (except nicotinic riboside). Fourteen tryptophan pathway metabolites show large negative component values in the metabolic syndrome group. Further investigations are required for comprehensive and systematic associations with other exposures in the epidemiological study (Figure 36).

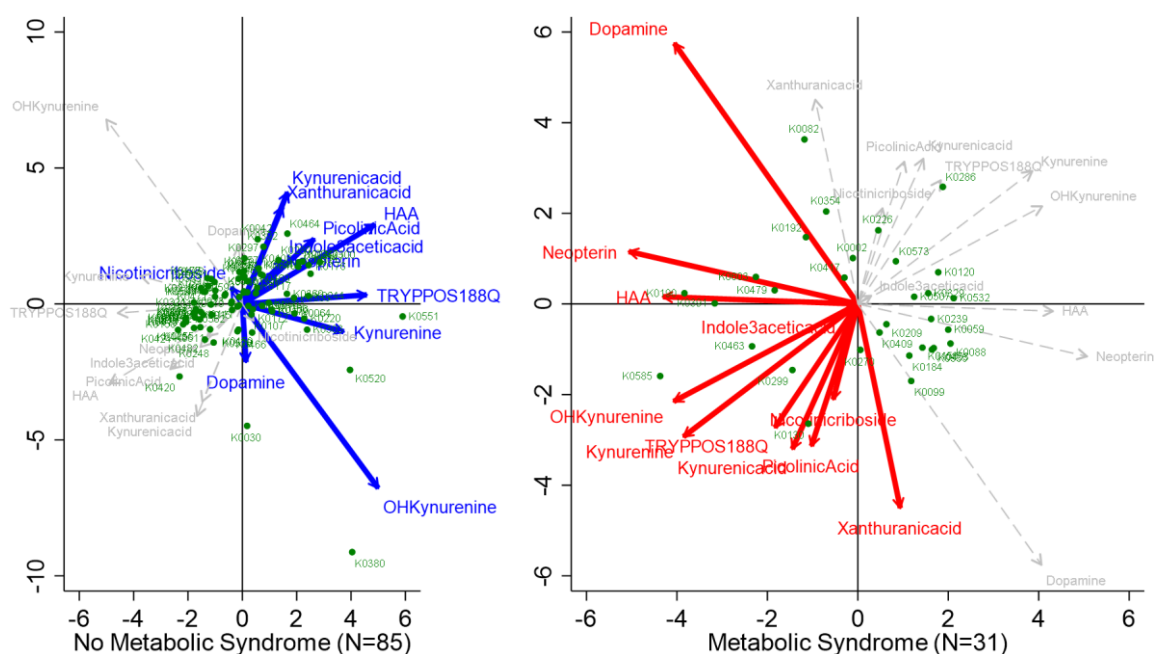


Figure 36 The biplot for (1) no metabolic syndrome and (2) metabolic syndrome (N=85 vs 31) group using the principal component analysis for the PQN-normalized tryptophan pathway metabolites

3. Statistical methods for analysing prognostic markers for brain health

The overall analytic methods used to address Objectives 1 to 5 are as follows:

The summation score of the four standardized (i.e. z-scores; $\frac{x_i - \bar{x}}{SD_x}$) cognitive tests (RAVLT, DSST, reversed Stroop, and MoCA) was generated as a composite cognitive score at Year 30; standardized RAVLT + standardized DSST – standardized Stroop + standardized MoCA. In addition, the summation score of the three standardized cognitive tests (RAVLT, DSST, and reversed Stroop) was generated as a composite cognitive score at Year 25; standardized RAVLT + standardized DSST – standardized Stroop. Composite cognitive measures based on the tests were used in the analysis with the objective of reducing floor and ceiling artifacts and other forms of measurement error [167].

Regarding Objectives 1 to 3 that I address in Chapter III, I conducted a mediation analysis (which is comprised of three sets of regression, Baron and Kenny's (1986) test [89]) to investigate whether the standardized metabolites and the standardized brain MRI parameters are mediators regarding cognition on the cardio-metabolic risk factors, respectively. I fitted a cardio-metabolic risk-specific multivariable-adjusted regression model regarding cognition on (given) each cardio-metabolic risk factor (i.e. SBP/DBP, HDL-C/LDL-C/triglycerides, glucose/insulin/eGFR, BMI/WC, 10-Year Framingham Risk) measured at Year 30. For each NMR, amino acids, tryptophan pathway metabolites and brain MRI parameters, a path analysis using multivariate linear structural equation modelling (SEM) was used to investigate mediators (i.e. metabolites or brain MRI parameters) of the adjusted composite cognitive score (i.e. the summation of the standardized four cognitive tests) on the cardio-metabolic factor measured at Year 30. Significant confounding factors and 'other exposures' (i.e. clinical, socio-demographic and lifestyle exposures) were investigated for the associations. In SEM, the relationships between the latent constructs are represented by regression or path coefficients between the factors [168]. A cardio-metabolic risk factor which showed more mediators was selected among the same category to be included in each total multivariate regression path diagram for metabolites and brain MRI parameters. The bootstrapped (500 bootstrap replications) standard errors and confidence intervals were obtained in the total multivariate regression path diagram to increase the statistical power and reproducibility of the study design [169].

The brain MRI parameters were transformed into z-scores, to allow comparisons across measures on a standardized one-SD increment. Total intracranial volume (ICV) was additionally adjusted for in all MRI model analyses to control for variations in voxel size over time [170].

Next, in Chapter IV, I conducted repeated-measures regression such as Bayesian multilevel models to examine the degeneration of the cognitive outcome on each brain MRI parameter for each cardio-metabolic factor measured at Year 25 and Year 30.

In Chapter V, to provide an integrated map (i.e. profile of the markers) of cardio-metabolic risk-specific brain health, I conducted network analysis and tree-based classification models, and examined profiles of urinary metabolites and brain MRI parameters, using significant other exposures and prognostic markers (i.e. urinary metabolites and brain MRI parameters) which were identified from Objectives 1 to 4.

Lastly, in Chapter VI, all the information from Objectives 1 to 5 (i.e. risk factors, physiological markers for the longitudinal information) was used to examine the possibility of developing a novel risk assessment tool of brain health (i.e. cognition or dementia). The risk scores using the significant variables will be discussed.

Chapter III. Predictions of cognition at Year 30

1. Predictions of cognition based on urinary metabolites

The hypothesis of the present study regarding urinary metabolites is described in Chapter I, Section 4; that is, that the urinary metabolites are mediators regarding cognition with a cardio-metabolic risk. In this chapter, the total path diagrams of SEM analysis were conducted to examine the NMR-and LC–MS-based urinary metabolites as mediators regarding adjusted composite cognition measured at Year 30 (Objectives 1 to 3). Mediation models concern the effects that connect indirectly through other variables, considering a response variable affected by an explanatory variable mediated through a mediator [171]. A mediation analysis (which is comprised of three sets of regression: $X \rightarrow Y$, $X \rightarrow M$, and $X + M \rightarrow Y$, Baron and Kenny's (1986) test [89]) was conducted to investigate whether the urinary metabolites are mediations of the relationship between cardio-metabolic risk factors and cognition (Figure 37 and Figure 38). The direction of associations indicates the dependent variable (to which arrows are pointing $\rightarrow$) and independent variables (having no arrows pointing to them) [172].

1-1. Mediations on cardio-metabolic risk factors

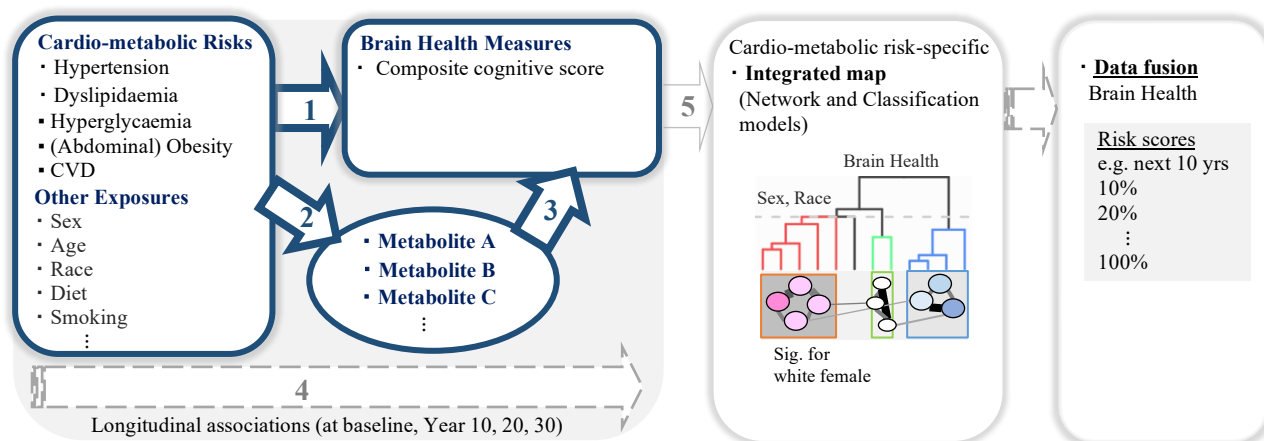


Figure 37 Schematic of the study highlighting the work focusses on Objectives 1 to 3. In the present section, urinary metabolites are hypothesised as mediators regarding cognition on a cardio-metabolic risk.

Application of systematic and comprehensive associations is the main interest of the analytic methods in this study, using robust models which can also deal with non-normalized data and measurement errors. I display the mediation analysis (i.e. SEM) for the **adjusted composite cognitive score with a cardio-metabolic risk factor** as a first step to build 1-2. in Figure 38.

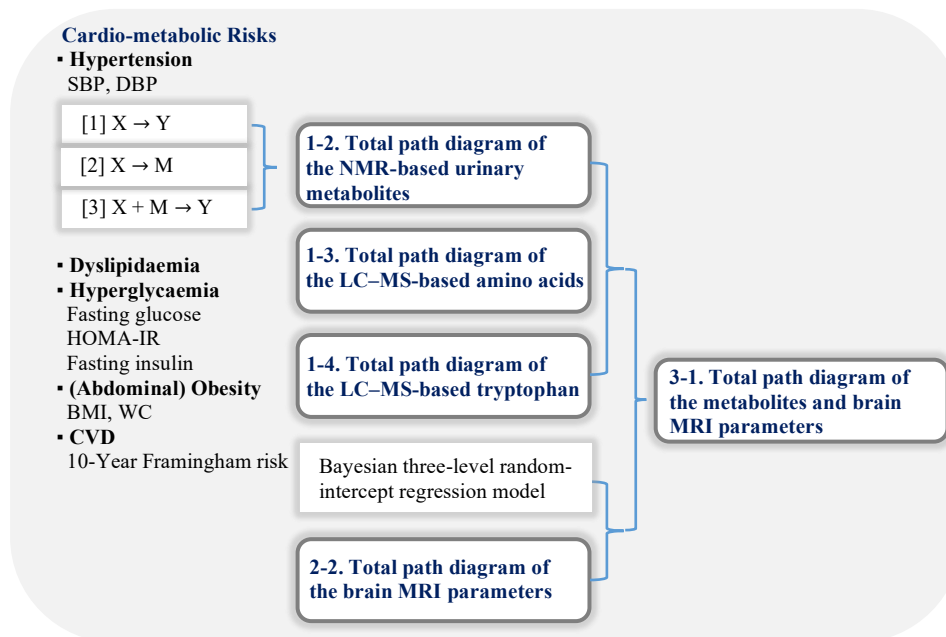


Figure 38 Workflow used in this chapter: predictions of cognition analysis at Year 30. A mediation analysis is comprised of three sets of regressions: $X \rightarrow Y$, $X \rightarrow M$, and $X + M \rightarrow Y$, Baron and Kenny's (1986) [89]; X: cardio-metabolic risk factor, M: metabolite, and Y: cognition. Bayesian three-level random-intercept regression model were explored to examine significant brain MRI parameters

Firstly, I investigated whether a urinary metabolite (M) is a mediator of the association between a cardio-metabolic risk factor (X) and the adjusted composite cognitive outcome (Y). In addition, I investigated significant exposures and confounding factors in the association. Here, a cardio-metabolic risk factor SBP was used to demonstrate a path diagram of the NMR-based urinary metabolites in the univariate and multiple structural equation models.

[1] Prediction of cognition from a cardio-metabolic risk factor; $X \rightarrow Y$

The cardio-metabolic risk factor SBP was significant regarding cognition in the univariate structural equation model (Table 16).

Table 16 Structural equation model for adjusted composite cognition on a cardio-metabolic risk factor SBP

Structural equation model		Number of observations = 588					
Estimation method = ml							
Log likelihood = -3964.5597							
		Observed Coef.	OIM Std. Err.	z	$P > z $	Normal-based [95% CIs]	
Adj. composite Cognition ←	SBP	-0.04	0.01	-5.56	<0.001	-0.06	-0.03
	constant	5.30	0.96	5.54	<0.001	3.43	7.18
var (e.Adj. composite Cognition)		9.58	0.56			8.55	10.75

OIM: observed information matrix, LR test of model vs saturated: $\chi^2(0) = 0.00$, Prob > $\chi^2 =$.

(The saturated model (i.e. the standard for comparison with the models) has a chi-square of zero with zero degrees of freedom)

While SBP itself was a significant factor regarding cognition, sex, education, HBP medication, sleep quality, and family income were significant exposures regarding cognition in the multiple structural equation model. Race was a confounding factor. SBP was significantly associated regarding cognition ($p=0.03$), but after adjusting for race, SBP is no longer significant ($p=0.79$). Therefore, race is a confounding factor of the association between SBP and the cognitive score (Table 17).

Table 17 Structural equation model for **adjusted composite cognition on a cardio-metabolic risk factor SBP** along with significant exposures

Structural equation model
Estimation method = ml
Log likelihood = -7826.3779

Number of observations = 580

		Observed Coef.	OIM Std. Err.	z	P > z	Normal-based [95% CIs]	
Adj. composite Cognition ←	SBP	-0.01	0.01	-2.16	0.03*	-0.03	-0.001
	Sex	1.07	0.20	5.26	<0.001	0.67	1.46
	Education	0.29	0.04	6.64	<0.001	0.21	0.38
	HBP medication	-1.05	0.23	-4.63	<0.001	-1.49	-0.60
	Sleep quality	-0.29	0.12	-2.41	0.02	-0.53	-0.06
	Family Income	0.47	0.05	9.54	<0.001	0.38	0.57
	constant	-5.93	1.11	-5.36	<0.001	-8.10	-3.76
var (e.Adj. composite Cognition)		5.75	0.34			5.12	6.45

OIM: observed information matrix, LR test of model vs saturated: $\chi^2(0) = 0.00$, Prob > $\chi^2 =$.

*After adjusting for race, SBP is no longer significant ($p=0.79$)

[2] Prediction of a metabolite from a cardio-metabolic risk factor; $X \rightarrow M$

A univariate structural equation model was constructed to examine the association between SBP (i.e. a cardio-metabolic risk factor) and each metabolite in turn. Maximum likelihood estimation was used for estimating the parameters of a model. SBP was significant regarding the standardized concentrations of creatinine, citric acid, valine, and N,N-Dimethylglycine, respectively, among the 10 NMR metabolites (Table 18).

Table 18 Summary of each structural equation model for the NMR-measured standardized concentrations of PQN-normalized creatinine, citric acid, valine, N,N-Dimethylglycine with SBP, respectively

Number of observations	Log likelihood			Observed Coef.	OIM Std. Err.	z	P > z	Normal-based [95% CIs]	
600	-3365.9599	Creatinine ←	SBP	0.01	0.003	2.85	0.004	0.002	0.01
589	-3304.5993	Citric acid ←	SBP	-0.01	0.003	-3.15	0.002	-0.01	-0.003
400	-2239.4267	Valine ←	SBP	0.01	0.003	2.14	0.03	0.001	0.01
386	-2181.0502	N,N-Dimethylglycine ←	SBP	0.01	0.003	2.77	0.01	0.002	0.01

OIM: observed information matrix, LR test of model vs saturated: $\chi^2(0) = 0.00$, Prob > $\chi^2 =$.

[3] Prediction of cognition from a cardio-metabolic risk factor and a metabolite, where a cardio-metabolic risk factor is significant regarding a metabolite; $\mathbf{X} + \mathbf{M} \rightarrow \mathbf{Y}$, where $\mathbf{X} \rightarrow \mathbf{M}$

The investigated exposures and PQN applied NMR metabolite concentrations, which were significant in the association between SBP and the cognitive outcome ([1] and [2]) were included in the next stage multiple structural equation model. The confounding factor, race, which makes SBP no longer significant regarding cognition ($p=0.48$), was not adjusted in the model. SBP was significant regarding the standardized concentrations of creatinine and N,N-Dimethylglycine, respectively, in the model. Citric acid and valine were mediators of the association between SBP and cognition, where the significant exposures were adjusted in the model. Citric acid and valine make SBP no longer significantly associated regarding cognition when they are adjusted in the model (without citric acid and valine, $p=0.04$), where SBP significantly associated regarding citric acid and valine, respectively. Therefore, citric acid and valine were considered mediators (Table 19).

Table 19 Structural equation model for adjusted composite cognition on a cardio-metabolic risk factor SBP with significant exposures and the PQN-normalized NMR metabolites

Structural equation model
Estimation method = ml
Log likelihood = -6027.9972

Number of observations = 372

		Observed Coef.	OIM Std. Err.	z	P > z	Normal-based [95% CIs]	
Adj. composite Cognition ←	Creatinine	-0.15	0.16	-0.91	0.36	-0.46	0.17
	N,N-Dimethylglycine	-0.31	0.14	-2.20	0.03	-0.58	-0.03
	SBP	-0.02	0.01	-2.07	0.04*	-0.03	0.00
	Sex	0.95	0.27	3.51	<0.001	0.42	1.48
	Education	0.28	0.06	4.78	<0.001	0.16	0.39
	HBP medication	-1.25	0.28	-4.56	<0.001	-1.79	-0.71
	Sleep quality	-0.31	0.15	-2.10	0.04	-0.60	-0.02
	Family income	0.48	0.06	7.48	<0.001	0.35	0.60
	constant	-5.21	1.39	-3.73	<0.001	-7.94	-2.48
Creatinine ←	SBP	0.01	0.003	2.47	0.01	0.001	0.01
	constant	-1.04	0.34	-3.08	0.002	-1.70	-0.38
N,N-Dimethylglycine ←	SBP	0.01	0.003	2.50	0.01	0.002	0.01
	constant	-0.93	0.37	-2.49	0.01	-1.66	-0.20
var (e.Adj. composite Cognition)		5.65	0.41			4.89	6.52
var (e.Creatinine)		0.80	0.06			0.69	0.92
var (e.N,N-Dimethylglycine)		0.98	0.07			0.85	1.13

OIM: observed information matrix, LR test of model vs saturated: $\chi^2(11) = 132.55$, Prob > $\chi^2 = 0.0000$

(In comparing the model with the saturated model, there was a difference of 11 degrees of freedom)

*After adjusting for race, SBP is no longer significant ($p=0.48$); race is a confounding factor

*After adjusting for citric acid and valine, SBP is no longer significant ($p=0.20$); citric acid and valine are mediators

1-2. Total path diagram of the NMR-based urinary metabolites

I conducted the same process for all the other cardio-metabolic risk factors (as shown above with SBP), Table 20 provides a summary of all those models. I investigated whether the urinary metabolites are mediators regarding the adjusted composite cognitive outcome on a cardio-metabolic risk factor (SBP, DBP, fasting glucose, HOMA-IR, fasting insulin, BMI, WC, and 10-Year Framingham Risk). Those cardio-metabolic risk factors which were not significant for cognition were not included; HDL-C, LDL-C and Triglycerides (Dyslipidaemia) and eGFR (Hyperglycaemia) were not significant for cognition, respectively ($p > 0.05$). To reduce the complexity of the model (i.e. overfitting), an ensemble model strategy was adopted in the present study; each SEM model per cardio-metabolic risk factor was conducted and based on statistical significance, a list of metabolites was selected to predict an overall outcome (Table 20). Cross-validation would be another method to access the issue; the variances and residual variance are calculated using predicted values through cross validation, which provides a certain level of control for the potential overfitting in the training process. (Model assessments based on cross validation for the metabolites were conducted in Chapter V, Section 3.) [173].

SEMs have been built with each cardio-metabolic factor (hypertension, dyslipidaemia, hyperglycaemia, (abdominal) obesity, CVD) and significant exposures regarding adjusted composite cognition (Table 20). SEM encompasses a broad array of models from linear regression to measurement models to simultaneous equations [174]. Firstly, I examined which (PQN applied) urinary metabolites by NMR are mediators on each cardio-metabolic risk factor, and which exposures are confounding factors in each association (Table 20). Two mediators, citric acid and valine were found for SBP. SBP was selected to be included in the total NMR model, as a hypertension factor as DBP has no mediator. No factors of dyslipidaemia (i.e. HDL-C, LDL-C, triglyceride) were significant for the adjusted composite cognitive score. Among the hyperglycaemia factors (i.e. fasting glucose, HOMA-IR, eGFR, fasting insulin), only eGFR was not significant for the adjusted composite cognitive score. Fasting glucose has a mediator, glycine; glycine, alanine, valine, and N,N-Dimethylglycine were significant regarding fasting glucose, respectively. When glycine was adjusted in the model with other significant exposures, fasting glucose was no longer significantly associated with cognition. Glycine was considered as a mediator in the SEM. Fasting insulin, which has three mediators (i.e. alanine, valine, and acetone) can be selected to be included in the total NMR model (Appendix Table 9 and Appendix Figure 18). However, the information including fasting glucose was applied to the total NMR model (in preference to fasting insulin) (Figure 39) as fasting glucose was selected for a hyperglycaemia factor in the amino acids, tryptophan, and brain MRI parameter models with a number of mediators (Figure 40, Figure 41, Figure 48). The (abdominal) obesity factors, BMI and WC have both a mediator, i.e. valine. The 10-Year Framingham Risk factor did not have a mediator (Table 20).

Table 20 Significant exposures, confounding factors, and NMR metabolite mediators in the structural equation models predicting the adjusted composite cognitive score, for each cardio-metabolic factor; each dot represents a mark of mediator

		Hypertension		Hyperglycaemia			(Abdominal) Obesity		CVD
		SBP	DBP	Fasting glucose	HOMA -IR	Fasting insulin	BMI	WC	10-Year Framingham Risk
Significant exposures	Race				•	•		•	•
	Sex	•	•		•	•	•		
	Age			•	•	•	•	•	
	Education	•	•		•	•	•	•	•
	CES-D				•	•			
	Smoking						•		
	HBP medication	•	•	•					
	Depression medication				•	•	•		
	Sleep quality	•	•				•		
	Family Income	•	•	•	•	•	•	•	•
Confounding factors	Conf Race	•	•	•			•		
	Conf Sex			•				•	•
	Conf Education			•					
Significant metabolites	Creatinine	•	•		•	•	•	•	•
	Citric acid	•	•			•			•
	Glycine			•					•
	Alanine		•	•	•	•	•	•	•
	Acetic acid								
	Succinic acid								•
	Oxaloacetic acid								
	Valine	•		•	•	•	•	•	•
	N,N-Dimethylglycine	•		•	•			•	•
	Acetone				•	•	•	•	•
Mediators	Citric acid	•							
	Glycine			•					
	Alanine				•	•			
	Succinic acid								
	Valine	•				•	•	•	
	N,N-Dimethylglycine				•				
	Acetone				•	•			

Note: eGFR (hyperglycaemia) was not significant for Adj. composite Cognition ($p=0.05$)

HDL-C, LDL-C and Triglycerides (dyslipidaemia) were not significant for Adj. composite Cognition ($p=0.05$)

Once each model per cardio-metabolic risk factor has been built, only ‘significant’ cardio-metabolic factors, exposures, mediators, and confounding factors were included in one final nonrecursive structural equation model which has correlated errors; the SEMs with measurement errors in metabolites from NMR or LC-MS were conducted (Figure 39–41). This approach refers to the base estimators from each model per cardio-metabolic risk factor, and it was chosen as a way of preserving statistical power (over other model fitting methods). In addition, only one cardio-metabolic risk factor ‘per category’ was included in the final model, with preference to one with ‘more mediators’ in each association. Selecting a cardio-metabolic risk factor with more mediators will be helpful to identify and explain the mechanism that underlies an observed relationship. For instance, SBP was chosen to be included in the final model instead of DBP as SBP has more mediators in each

association in the category of hypertension. Metabolites were standardized (i.e. z-scores) in the models for additional benefits such as reducing multicollinearity, the likelihood function is to be maximized with variables in the z-scores faster and importantly, being able to include those from originally different units all together in the final model [175]. The bootstrapped (500 bootstrap replications) standard errors and confidence intervals were obtained by using the *vce(bootstrap)* option in Stata/IC 16.0 (StataCorp LLC); the method would offer more reliable findings in studies with a small sample size [176].

The total multivariate regression path diagram using the NMR metabolite concentrations can be summarised as follows:

Adj. Composite Cognition

$$= \alpha_1 + \text{SBP } \beta_1 + \text{Fasting glucose } \beta_2 + \text{WC } \beta_3 + \text{Framingham risk } \beta_4 + \text{Age } \beta_5 + \dots \\ + \text{Family income } \beta_{14} + \text{Citric acid } \beta_{15} + \text{Valine } \beta_{16} + \text{Glycine } \beta_{17} + e_5$$

where, $(\text{SBP}, \dots, \text{Family Income}, e_5) \sim i.i.d. \text{ with mean } \mu \text{ and covariance } \Sigma$

$$\text{Citric acid} = \alpha_1 + \text{SBP } \gamma_1 + e_1$$

$$\text{Valine} = \alpha_1 + \text{SBP } \delta_1 + \text{WC } \delta_2 + e_2$$

$$\text{Glycine} = \alpha_1 + \text{Fasting glucose } \zeta_1 + e_3$$

where, $(\text{SBP}, \text{Fasting glucose}, \text{WC}, e_1, e_2, e_3, e_4) \sim i.i.d. \text{ with mean } \mu \text{ and covariance } \Sigma$

The total path diagram of the NMR metabolites concentrations is presented in Figure 39 and Table 21. The risk factors or exposures in red indicate those which were statistically significant ($p < 0.05$) in the total mode. The risk factors in blue indicate those which were significantly associated with a mediator. The variables in grey colours are significant in each cardio-metabolic model, but no longer significant in the final models; they were adjusted in the final nonrecursive structural equation model.

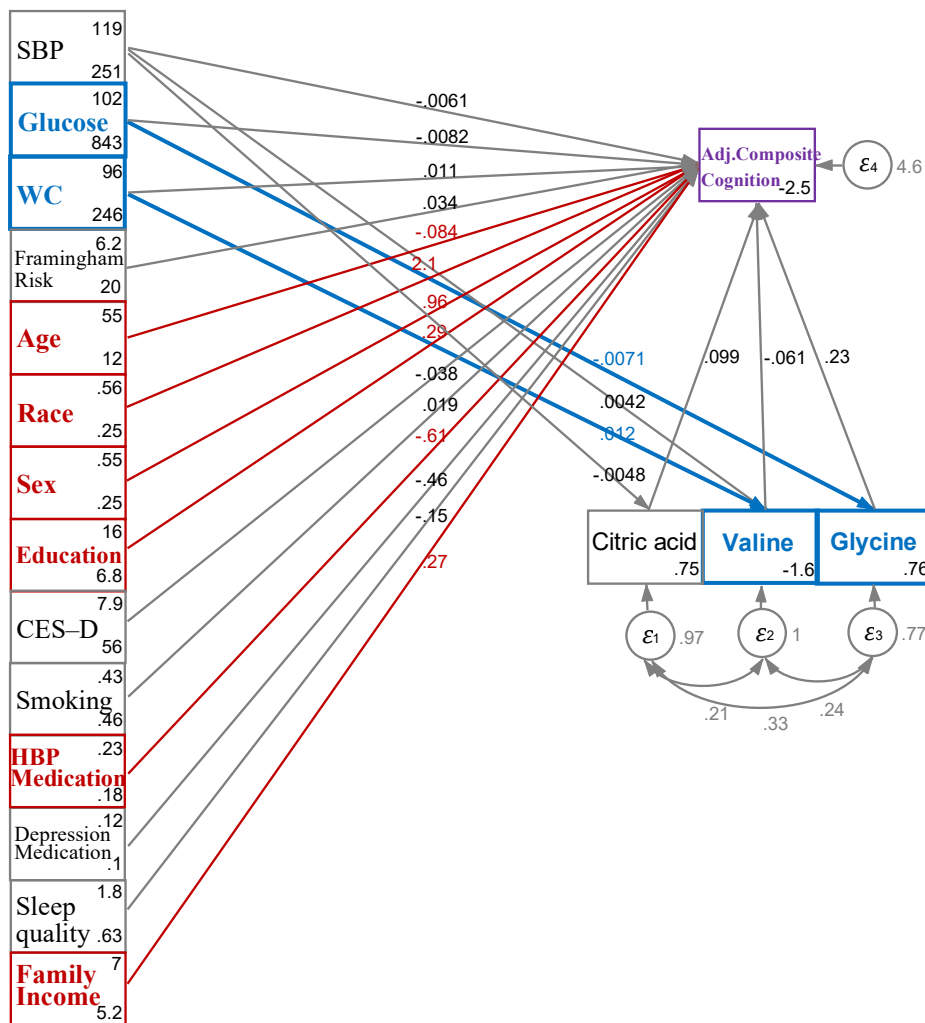


Figure 39 Total multivariate regression path diagram for the adjusted composite cognitive score on cardio-metabolic risk factors (SBP, fasting glucose, WC, and Framingham risk) with NMR metabolites (N=346)

Note: The variables inside the rectangles are observed variables in this path diagram. The figures inside the rectangles are mean and variance of the variables. For instance, 119 and 251 indicate mean and variance of SBP, respectively. Paths are direct relationships between variables. Estimated path coefficients are represented by straight arrows. Curved arrows specify that error terms covary.

In the total path diagram of the metabolites measured by NMR model which included (1) WC (instead of BMI), and (2) fasting glucose (instead of fasting insulin), age, race, sex, education (up to Year 30), HBP medication and family income were significant ($p < 0.05$) regarding the adjusted composite cognitive score. Fasting glucose was significant regarding **glycine** (Coef. -0.007). WC was significant regarding **valine** (Coef. 0.012) (Table 21).

Table 21 Structural equation model for the adjusted composite cognitive score on cardio-metabolic risk factors (SBP, fasting glucose, WC, and Framingham risk) with the NMR metabolite concentrations

Structural equation model
Log likelihood = -12504.675

Number of observations = 346
Replications = 500

		Observed Coef.	Bootstrap Std. Err.	z	P > z	Normal-based [95% CIs]	
Adj. composite Cognition ←	Citric acid	0.099	0.14	0.71	0.48	-0.18	0.37
	Valine	-0.061	0.13	-0.48	0.63	-0.31	0.19
	Glycine	0.232	0.13	1.76	0.08	-0.03	0.49
	SBP	-0.006	0.01	-0.72	0.47	-0.02	0.01
	Glucose	-0.008	0.01	-1.36	0.17	-0.02	0.00
	WC	0.011	0.01	1.20	0.23	-0.01	0.03
	Framingham risk	0.034	0.03	1.08	0.28	-0.03	0.10
	Age	-0.084	0.04	-2.39	0.02	-0.15	-0.02
	Race	2.124	0.31	6.89	<0.001	1.52	2.73
	Sex	0.958	0.28	3.41	0.001	0.41	1.51
	Education	0.286	0.06	4.53	<0.001	0.16	0.41
	CES-D	-0.038	0.02	-1.56	0.12	-0.08	0.01
	Smoking	0.019	0.22	0.09	0.93	-0.42	0.46
	HBP medication	-0.612	0.30	-2.02	0.04	-1.21	-0.02
	Depression medication	-0.456	0.50	-0.91	0.36	-1.43	0.52
	Sleep quality	-0.145	0.17	-0.88	0.38	-0.47	0.18
	Family income	0.273	0.07	3.85	<0.001	0.13	0.41
	constant	-2.518	2.46	-1.02	0.31	-7.35	2.31
Citric acid ←	SBP	-0.005	0.004	-1.34	0.18	-0.0118	0.0022
	constant	0.747	0.43	1.74	0.08	-0.0940	1.5873
Valine ←	SBP	0.004	0.004	1.19	0.24	-0.0027	0.0112
	WC	0.012	0.003	3.48	0.001	0.0052	0.0188
	constant	-1.623	0.49	-3.31	0.001	-2.5841	-0.6611
Glycine ←	Fasting glucose	-0.007	0.001	-4.81	<0.001	-0.0100	-0.0042
	constant	0.759	0.17	4.41	<0.001	0.4213	1.0965
var (e.Adj. composite Cognition)		4.607	0.64			3.51	6.04
var (e.Citric acid)		0.968	0.08			0.82	1.14
var (e.Valine)		1.010	0.10			0.84	1.22
var (e.Glycine)		0.770	0.14			0.54	1.10
cov (e.Citric acid, e.Valine)		0.205	0.05	4.02	<0.001	0.11	0.31
cov (e.Citric acid, e.Glycine)		0.330	0.06	5.60	<0.001	0.21	0.45
cov (e.Valine, e.Glycine)		0.238	0.06	4.08	<0.001	0.12	0.35

1-3. Total path diagram of the LC–MS-based amino acids

The same analytic approach was applied to the models for amino acids measured by LC–MS.

The (PQN applied) amino acids measured by LC–MS which are mediators on each cardio-metabolic risk factor, and the exposures which are confounding factors in each association were examined. SBP had a mediator, methionine, whereas DBP had no mediator. None of the factors of dyslipidaemia were significant for the adjusted composite cognitive score. Among the hyperglycaemia factors (i.e. fasting glucose, HOMA-IR, eGFR, fasting insulin), fasting glucose has mediators glycine, isoleucine, leucine, phenylalanine, tryptophan, valine, and aminoadipic acid. The (abdominal) obesity factors, WC has a mediator serine, and BMI has a mediator hydroxylysine. The 10-Year Framingham Risk has three mediators, isoleucine, methionine, and phenylalanine as shown in Table 22.

Table 22 Significant exposures, confounding factors, and amino acids mediators by LC–MS in the structural equation models by each cardio-metabolic factor

		Hypertension		Hyperglycaemia			(Abdominal) Obesity		CVD
		SBP	DBP	Fasting glucose	HOMA-IR	Fasting insulin	BMI	WC	10-Year Framingham Risk
Significant exposures	Race				•	•		•	•
	Sex	•	•		•	•	•		
	Age			•	•	•	•	•	
	Education	•	•	•	•	•	•	•	•
	CES–D				•	•			
	Smoking						•		
	HBP medication	•	•						
	Depression medication				•	•	•		
	Sleep quality	•	•				•		
	Family Income	•	•	•	•	•	•	•	•
Confounding factors	Conf Race	•	•	•			•		
	Conf Sex			•				•	•
	Conf Education								
	Conf Age								•
Significant metabolites	Alanine (amino acids)	•	•	•	•	•	•	•	•
	Asparagine			•			•	•	
	Ethanolamine					•			
	Glutamic acid	•							•
	Glutamine			•					
	Glycine (amino acids)			•				•	•
	Histidine							•	
	Isoleucine	•	•	•	•	•	•	•	•
	Leucine		•	•	•	•	•	•	•
	Lysine244								
	Methionine (amino acids)	•		•	•	•	•	•	•
	Phenylalanine	•	•	•	•	•	•	•	•
	Serine			•	•	•		•	
	Threonine	•	•						
	Tryptophan	•		•	•	•	•	•	•
	Tyrosine	•	•	•	•	•	•	•	•
	Valine (amino acids)	•	•	•	•	•	•	•	•
	Beta aminoisobutyric acid								
	Cystathionine282 methylhistidine3						•		
	1-Methylhistidine		•					•	
	Hydroxylysine	•	•				•	•	•
	Aminoadipic acid			•	•	•		•	•
	Beta-alanine								
Mediators	Glycine (amino acids)			•					
	Isoleucine			•					•
	Leucine			•					
	Methionine (amino acids)	•							•
	Phenylalanine			•					•
	Serine							•	
	Tryptophan			•					
	Threonine								
	Valine (amino acids)			•					
	methylhistidine3								
	Hydroxylysine						•		
	Aminoadipic acid			•					
	Beta-alanine								

Note: eGFR (Hyperglycaemia) was not significant for Adj. composite Cognition ($p=0.05$)

HDL–C, LDL–C and Triglycerides (Dyslipidaemia) were not significant for Adj. composite Cognition ($p=0.05$)

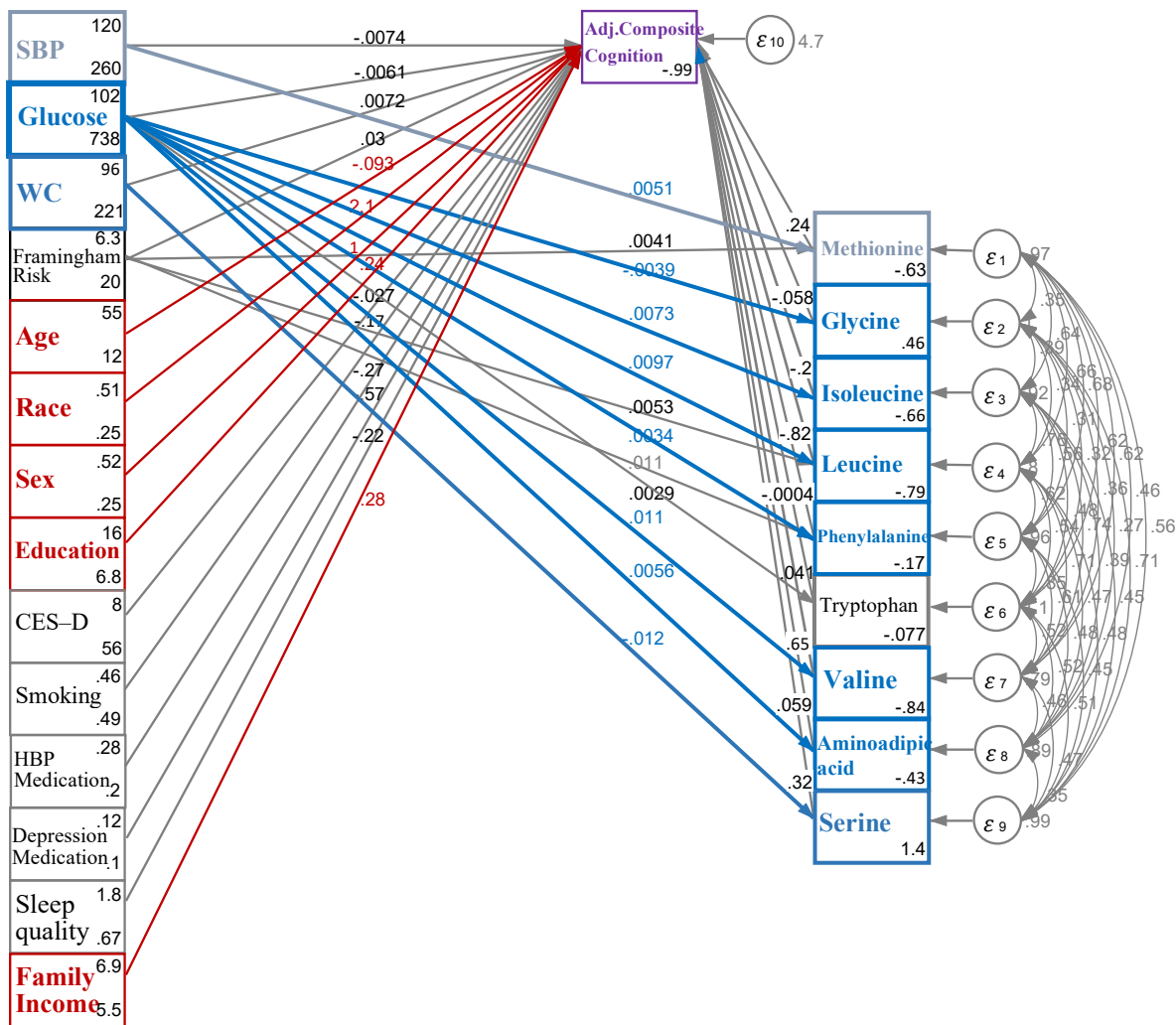


Figure 40 Total multivariate regression path diagram for the adjusted composite cognitive score on cardio-metabolic risk factors (SBP, Glucose, WC, and Framingham risk) with amino acids (N=420)

The total multivariate regression path diagram for amino acids measured by LC–MS is presented in Figure 40 and Table 23. In this total path diagram for amino acids, age, race, sex, education (up to Year 30), and family income were significant ($p < 0.05$) regarding the adjusted composite cognitive score.

SBP was significant regarding methionine (Coef. 0.0051). Fasting glucose was negatively associated with each glycine (Coef. -0.0039), serine (Coef. -0.012), and positively associated with isoleucine (Coef. 0.0073), leucine (Coef. 0.0097), phenylalanine (Coef. 0.0034), valine (Coef. 0.011), and amino adipic acid (Coef. 0.0056). WC was significant regarding serine in the total amino acids model. An example of fasting glucose, a mediator and a confounding factor in the model can be explained as follows: fasting glucose (explanatory variable) is no longer significant in predicting the cognitive score (the response variable) when glycine is included in the model, where fasting glucose (explanatory variable) significantly affects glycine. Therefore, Glycine is a mediator.

Glucose alone significantly ($p < 0.001$) affects the adjusted composite cognitive score, but after adjusting for sex, or race (fitted with education) glucose is no longer significant. Therefore, sex, or race (fitted with education) are respectively confounding factors of the association between glucose and cognition.

To investigate effects of anti-hypertensive medications to the urinary metabolites, I conducted the sensitivity analysis by not including HBP (i.e. a significant exposure among the medications). Without adjusting for HBP, leucine (Coef. -0.89) and valine (Coef. 0.72) were significant, respectively ($p=0.032$) regarding the adjusted composite cognitive score (Table 23); there was no significant effect of HBP to the urinary metabolites in the NMR and tryptophan models.

Table 23 Structural equation model for the adjusted composite cognitive score on cardio-metabolic risk factors (SBP, Glucose, BMI, and Framingham risk) with the amino acids by LC-MS

Structural equation model				Number of observations		=	420
Log likelihood = -17099.202				Replications		=	500
		Observed Coef.	Bootstrap Std. Err.	z	<i>P</i> > z	Normal-based [95% CIs]	
Adj. composite Cognition ←	Methionine	0.243	0.18	1.38	0.17	-0.10	0.59
	Glycine	-0.058	0.16	-0.36	0.72	-0.37	0.26
	Isoleucine	-0.204	0.33	-0.61	0.54	-0.86	0.45
	Leucine	-0.821	0.40	-2.07	0.04	-1.60	-0.05
	Phenylalanine	0.000	0.31	0.00	1.00	-0.60	0.60
	Tryptophan	-0.041	0.26	-0.16	0.87	-0.54	0.46
	Valine	0.647	0.33	1.98	0.05	0.01	1.29
	Aminoadipic acid	0.059	0.15	0.39	0.69	-0.24	0.35
	Serine	0.321	0.19	1.69	0.09	-0.05	0.69
	SBP	-0.007	0.01	-0.90	0.37	-0.02	0.01
	Glucose	-0.006	0.01	-1.03	0.31	-0.02	0.01
	WC	0.007	0.01	0.73	0.47	-0.01	0.03
	Framingham risk	0.030	0.03	0.91	0.36	-0.03	0.09
	Age	-0.093	0.03	-2.71	0.01	-0.16	-0.03
	Race	2.104	0.29	7.13	< 0.001	1.53	2.68
	Sex	1.034	0.27	3.82	< 0.001	0.50	1.57
	Education	0.240	0.06	4.24	< 0.001	0.13	0.35
	CES-D	-0.027	0.02	-1.15	0.25	-0.07	0.02
	Smoking	-0.173	0.18	-0.96	0.34	-0.53	0.18
	HBP medication	-0.270	0.26	-1.02	0.31	-0.79	0.25
	Depression medication	-0.573	0.38	-1.50	0.14	-1.32	0.18
	Sleep quality	-0.216	0.16	-1.33	0.19	-0.53	0.10
	Family income	0.284	0.07	4.33	< 0.001	0.16	0.41
	constant	-0.994	2.55	-0.39	0.70	-6.00	4.01
Methionine ←	SBP	0.005	0.002	2.41	0.02	0.001	0.009
	Framingham risk	0.004	0.01	0.40	0.69	-0.02	0.02
	constant	-0.632	0.24	-2.68	0.01	-1.10	-0.17
Glycine ←	Fasting glucose	-0.0039	0.00	-3.72	< 0.001	-0.006	-0.002
	constant	0.460	0.12	3.75	< 0.001	0.22	0.70
Isoleucine ←	Fasting glucose	0.0073	0.00	4.75	< 0.001	0.004	0.01
	constant	-0.659	0.16	-4.06	< 0.001	-0.98	-0.34
Leucine ←	Fasting glucose	0.0097	0.00	5.64	< 0.001	0.006	0.013
	Framingham risk	0.005	0.00	1.45	0.15	0.00	0.01
	constant	-0.792	0.17	-4.55	< 0.001	-1.13	-0.45
Phenylalanine ←	Fasting glucose	0.0034	0.00	2.33	0.02	0.001	0.006
	Framingham risk	0.011	0.01	1.81	0.07	-0.001	0.02
	constant	-0.168	0.14	-1.19	0.24	-0.45	0.11

(Table Continued)

Tryptophan ←	Fasting glucose constant	0.003 -0.077	0.002 0.16	1.83 -0.47	0.07 0.64	-0.0002 -0.40	0.006 0.24
Valine ←	Fasting glucose constant	0.011 -0.840	0.002 0.17	6.66 -5.06	< 0.001 < 0.001	0.008 -1.17	0.014 -0.51
Amino adipic acid ←	Fasting glucose constant	0.006 -0.427	0.002 0.23	2.55 -1.88	0.01 0.06	0.001 -0.87	0.01 0.02
Serine ←	WC constant	-0.012 1.366	0.002 0.21	-6.21 6.59	< 0.001 < 0.001	-0.02 0.96	-0.01 1.77
var (e.Adj. composite Cognition)		4.698	0.52			3.78	5.84
var (e.Methionine)		0.966	0.08			0.83	1.13
var (e.Glycine)		1.046	0.15			0.79	1.38
var (e.Isoleucine)		0.922	0.12			0.71	1.20
var (e.Leucine)		0.805	0.08			0.67	0.97
var (e.Phenylalanine)		0.962	0.10			0.79	1.18
var (e.Tryptophan)		1.059	0.11			0.86	1.31
var (e.Valine)		0.790	0.07			0.67	0.94
var (e.Amino adipic acid)		0.889	0.09			0.72	1.09
var (e.Serine)		0.987	0.10			0.82	1.19
cov (e.Methionine, e.Glycine)		0.355	0.06	5.91	< 0.001	0.24	0.47
cov (e.Methionine, e.Isoleucine)		0.639	0.07	8.55	< 0.001	0.49	0.79
cov (e.Methionine, e.Leucine)		0.660	0.07	9.89	< 0.001	0.53	0.79
cov (e.Methionine, e.Phenylalanine)		0.680	0.08	9.06	< 0.001	0.53	0.83
cov (e.Methionine, e.Tryptophan)		0.622	0.07	8.50	< 0.001	0.48	0.77
cov (e.Methionine, e.Valine)		0.617	0.06	9.86	< 0.001	0.49	0.74
cov (e.Methionine, e.Amino adipic acid)		0.461	0.06	8.12	< 0.001	0.35	0.57
cov (e.Methionine, e.Serine)		0.562	0.06	9.30	< 0.001	0.44	0.68
cov (e.Glycine, e.Isoleucine)		0.390	0.06	6.51	< 0.001	0.27	0.51
cov (e.Glycine, e.Leucine)		0.339	0.05	6.43	< 0.001	0.24	0.44
cov (e.Glycine, e.Phenylalanine)		0.307	0.06	5.11	< 0.001	0.19	0.42
cov (e.Glycine, e.Tryptophan)		0.324	0.06	5.37	< 0.001	0.21	0.44
cov (e.Glycine, e.Valine)		0.361	0.05	7.25	< 0.001	0.26	0.46
cov (e.Glycine, e.Amino adipic acid)		0.265	0.05	4.92	< 0.001	0.16	0.37
cov (e.Glycine, e.Serine)		0.713	0.09	7.83	< 0.001	0.53	0.89
cov (e.Isoleucine, e.Leucine)		0.761	0.09	8.71	< 0.001	0.59	0.93
cov (e.Isoleucine, e.Phenylalanine)		0.558	0.06	8.60	< 0.001	0.43	0.69
cov (e.Isoleucine, e.Tryptophan)		0.476	0.06	7.92	< 0.001	0.36	0.59
cov (e.Isoleucine, e.Valine)		0.737	0.08	8.75	< 0.001	0.57	0.90
cov (e.Isoleucine, e.Amino adipic acid)		0.392	0.05	8.32	< 0.001	0.30	0.48
cov (e.Isoleucine, e.Serine)		0.450	0.06	7.24	< 0.001	0.33	0.57
cov (e.Leucine, e.Phenylalanine)		0.621	0.06	10.30	< 0.001	0.50	0.74
cov (e.Leucine, e.Tryptophan)		0.544	0.06	9.52	< 0.001	0.43	0.66
cov (e.Leucine, e.Valine)		0.715	0.07	10.48	< 0.001	0.58	0.85
cov (e.Leucine, e.Amino adipic acid)		0.467	0.05	8.98	< 0.001	0.37	0.57
cov (e.Leucine, e.Serine)		0.475	0.05	8.67	< 0.001	0.37	0.58
cov (e.Phenylalanine, e.Tryptophan)		0.848	0.09	9.12	< 0.001	0.67	1.03
cov (e.Phenylalanine, e.Valine)		0.611	0.06	10.17	< 0.001	0.49	0.73
cov (e.Phenylalanine, e.Amino adipic acid)		0.479	0.05	9.23	< 0.001	0.38	0.58
cov (e.Phenylalanine, e.Serine)		0.455	0.06	7.47	< 0.001	0.34	0.57
cov (e.Tryptophan, e.Valine)		0.524	0.06	9.31	< 0.001	0.41	0.63
cov (e.Tryptophan, e.Amino adipic acid)		0.521	0.05	9.49	< 0.001	0.41	0.63
cov (e.Tryptophan, e.Serine)		0.513	0.07	7.30	< 0.001	0.38	0.65
cov (e.Valine, e.Amino adipic acid)		0.461	0.05	10.09	< 0.001	0.37	0.55
cov (e.Valine, e.Serine)		0.472	0.05	8.70	< 0.001	0.37	0.58
cov (e.Amino adipic acid, e.Serine)		0.348	0.05	6.91	< 0.001	0.25	0.45

In the total amino acids model applying BMI, instead of WC, fasting glucose was no longer significant regarding **phenylalanine**. BMI was significant regarding **hydroxylysine** (Coef. 0.065). In addition, valine was significant regarding the cognitive score (Coef. 0.718). Other exposures (age, race, sex, education up to Year 30 and family income), SBP regarding methionine, and fasting glucose regarding each glycine, isoleucine, leucine, valine,

and amino adipic acid were consistently significant ($p < 0.05$) in the total amino acids model examining BMI. The information of the total amino acids model with WC and fasting glucose was applied in the total brain MRI parameter models to investigate the associations among the brain health measurements, cardio-metabolic factors, and the urinary metabolites at Year 30 in Section 3.

1-4. Total path diagram of the LC–MS-based tryptophan pathway metabolites

Tryptophan is an essential amino acid that is needed for the in vivo biosynthesis of proteins and is metabolically transformed to bioactive metabolites after consumption [177]. The same analytic approaches were applied to the models for tryptophan measured by LC–MS; tryptophan pathway metabolites which are mediators on each cardio-metabolic risk factor, and exposures which are confounding factors in each association have been examined.

Table 24 Significant exposures, confounding factors, and tryptophan mediators by LC–MS in the structural equation models by each cardio-metabolic factor

		Hypertension		Hyperglycaemia			(Abdominal) Obesity		CVD
		SBP	DBP	Fasting Glucose	HOMA-IR	Fasting insulin	BMI	WC	10-Year Framingham Risk
Significant exposures	Race				•	•		•	•
	Sex	•	•		•	•	•	•	
	Age			•	•	•	•	•	
	Education	•	•		•	•	•	•	•
	CES–D				•	•			
	Smoking*								
	HBP medication	•	•	•					•
	Depression medication*								
	Sleep quality						•		•
	Family income	•	•	•	•	•	•	•	•
Confounding factors	Conf Race	•	•	•			•		
	Conf Sex			•					•
	Conf Education			•					
	Conf Age								•
Significant metabolites	Picolinic acid							•	
	Neopterin						•		
	Tryptophan	•	•	•	•	•	•	•	•
	Kynurenine	•	•	•	•	•	•	•	•
	Dopamine	•	•				•		
	3-OH-kynurenine	•	•	•	•	•	•	•	•
	HAA								
	Indole-3-acetic acid	•	•	•	•	•	•	•	•
	Nicotinic riboside								
	Kynurenic acid								
Mediators	Xanthurenic acid								
	Tryptophan			•					
	Kynurenine			•	•				
	3-OH-kynurenine				•	•			
	HAA								
	Indole-3-acetic acid	•		•					
	Kynurenic acid								

* Smoking and depression medication were not significant in any structural equation models by each cardio-metabolic factor

Note: eGFR (Hyperglycaemia) was not significant for Adj. composite Cognition ($p=0.05$)

HDL–C, LDL–C and Triglycerides (Dyslipidaemia) were not significant for Adj. composite Cognition ($p=0.05$)

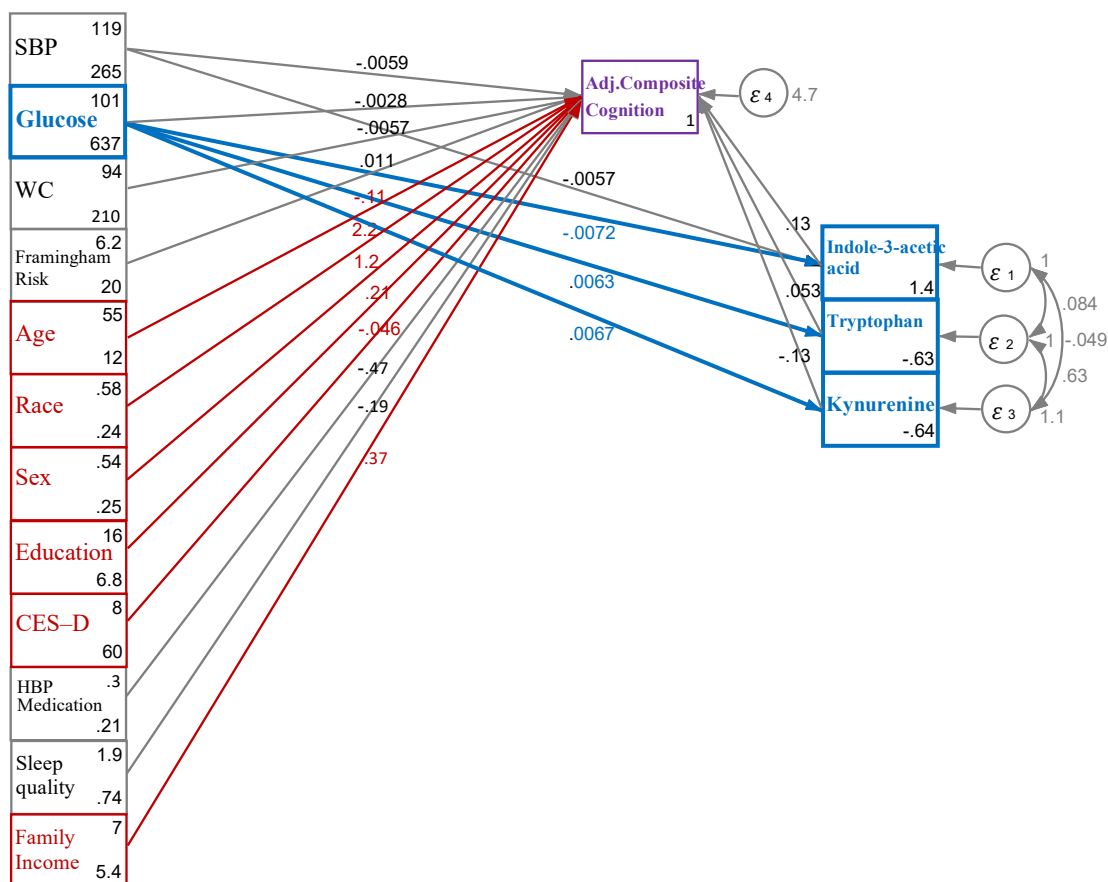


Figure 41 Total multivariate regression path diagram for the adjusted composite cognitive score on cardio-metabolic risk factors (SBP, fasting glucose, WC, and Framingham risk) with tryptophan (N=413)

Table 25 Structural equation model for the adjusted composite cognitive score on cardio-metabolic risk factors (SBP, Fasting glucose, BMI, and Framingham risk) with the tryptophan by LC-MS

Structural equation model
Log likelihood = -14466.635

Number of observations = 413
Replications = 500

		Observed Coef.	Bootstrap Std. Err.	z	P > z	Normal-based [95% CIs]	
Adj. composite Cognition ←	Indole-3-acetic acid	0.13	0.11	1.24	0.213	-0.08	0.35
	Tryptophan	0.053	0.14	0.37	0.711	-0.23	0.33
	Kynurenine	-0.13	0.13	-1.01	0.314	-0.38	0.12
	SBP	-0.01	0.01	-0.82	0.413	-0.02	0.01
	WC	-0.01	0.01	-0.63	0.530	-0.02	0.01
	Glucose	0.00	0.01	-0.47	0.641	-0.01	0.01
	Framingham risk	0.01	0.04	0.32	0.746	-0.06	0.08
	Age	-0.11	0.03	-3.56	< 0.001	-0.18	-0.05
	Race	2.16	0.29	7.34	< 0.001	1.58	2.74
	Sex	1.19	0.26	4.54	< 0.001	0.68	1.70
	Education	0.21	0.06	3.58	< 0.001	0.09	0.32

(Table Continued)

	CES-D	-0.046	0.02	-2.15	0.031	-0.09	-0.004
	HBP medication	-0.47	0.27	-1.73	0.083	-1.00	0.06
	Sleep quality	-0.19	0.16	-1.19	0.234	-0.51	0.13
	Family income	0.37	0.07	5.12	< 0.001	0.23	0.51
	constant	1.02	2.39	0.43	0.669	-3.66	5.71
Indole-3-acetic acid ←	SBP	-0.01	0.00	-1.42	0.16	-0.01	0.002
	Fasting glucose	-0.0072	0.00	-3.49	< 0.001	-0.01	-0.003
	constant	1.45	0.51	2.86	0.004	0.46	2.44
Tryptophan ←	Fasting glucose	0.0063	0.00	2.42	0.02	0.001	0.01
	constant	-0.63	0.26	-2.43	0.02	-1.14	-0.12
Kynurenine ←	Fasting glucose	0.0067	0.00	3.29	0.001	0.003	0.01
	constant	-0.64	0.21	-2.97	0.003	-1.06	-0.22
var (e.Adj. composite Cognition)		4.74	0.51			3.84	5.85
var (e.Indole-3-acetic acid)		1.00	0.14			0.75	1.32
var (e.Tryptophan)		1.00	0.09			0.84	1.19
var (e.Kynurenine)		1.08	0.13			0.85	1.38
cov (e.Indole-3-acetic acid, e.Tryptophan)		0.08	0.05	1.74	0.08	-0.01	0.18
cov (e.Indole-3-acetic acid, e.Kynurenine)		-0.05	0.05	-1.02	0.31	-0.14	0.05
cov (e.Tryptophan, e.Kynurenine)		0.63	0.07	9.26	< 0.001	0.49	0.76

1-5. Total path diagram of the metabolites – Summary and discussion

In this section, I summarise the mediators found in each model based on NMR measured metabolites, and LC–MS measured amino acids and tryptophan.

In the hypertension category, **methionine** showed positive associations between **SBP** (and cognition) in the amino acids model (Figure 40). A study indicated that a diet high in L–methionine triggered a decline in the cognitive capacities of wild-type mice and might contribute to the appearance of Alzheimer’s-like neurodegeneration [178]. For hyperglycaemia, **valine**, and **alanine** showed positive associations between **fasting insulin** and cognition in the NMR model (Figure 39). It is reported that valine is important for a healthy nervous system and cognitive functioning [179], and glycine is an amino acid that the body uses to create proteins, lessens daytime sleepiness and improves cognition [180-182]. In addition, valine and **amino adipic acid** showed positive associations between **fasting glucose** and cognition respectively, in the amino acids models.

Glenn *et al.* (2019) underlined the fact that protein or amino acid consumption is directly related to a number of disease-related risk factors as such low-muscle mass, sleep, stress, depression, and anxiety. Investigation of dietary intake with respect to protein or amino acid for the prevention of AD and related dementias has grown significantly in recent years [180].

Glycine showed negative associations between **fasting glucose** (and cognition) in the amino acids model (Figure 40). Glycine showed also a negative association with fasting glucose in the NMR model when fasting glucose was included instead of fasting insulin in the model. A study by Gannon *et al.* (2002) supported the hypothesis that oral glycine stimulates the secretion of a gut hormone that increases the effect of insulin on glucose removal from the circulation [183]. This means that ingestion of glycine reduces blood glucose levels,

and glycine is stimulating insulin secretion (which is benefit to diabetes patients) in humans [184]. Findings from many studies reported that insulin dysfunction (such as Type 2 diabetes) connects with Alzheimer-specific brain pathology [185-188].

Isoleucine, leucine and phenylalanine showed a positive association with **fasting glucose** respectively, in the amino acids model (Table 23). Humans require significant amounts of essential amino acids each day. Leucine is known to play several important roles in metabolic regulation as substrates (physical layers) for the synthesis of proteins [189, 190]. Layman (2003) explained that one of these roles includes being a direct link to the maintenance of glucose homeostasis (the balance of insulin and glucagon) by enhancing the recycling of glucose via the glucose-alanine cycle and a direct link to translational regulation of muscle protein synthesis through the insulin signalling cascade [191].

Tryptophan showed positive associations with **fasting glucose** (and cognition) in the tryptophan model (Figure 41). Tryptophan is an essential amino acid in humans, requiring that it be ingested [192]. Palego *et al.* (2016) underlined that a widely studied model of tryptophan uptake mechanism is the regulation of transport across the blood-brain barrier (BBB) (a physiologic barrier that restricts indiscriminate access of certain substances in the bloodstream to the brain) [193, 194]. It is known that insulin and other large neutral L-amino acids (i.e. valine, leucine, isoleucine, tyrosine, and phenylalanine) have been found to play an important role in the transportation of tryptophan across the BBB [195, 196].

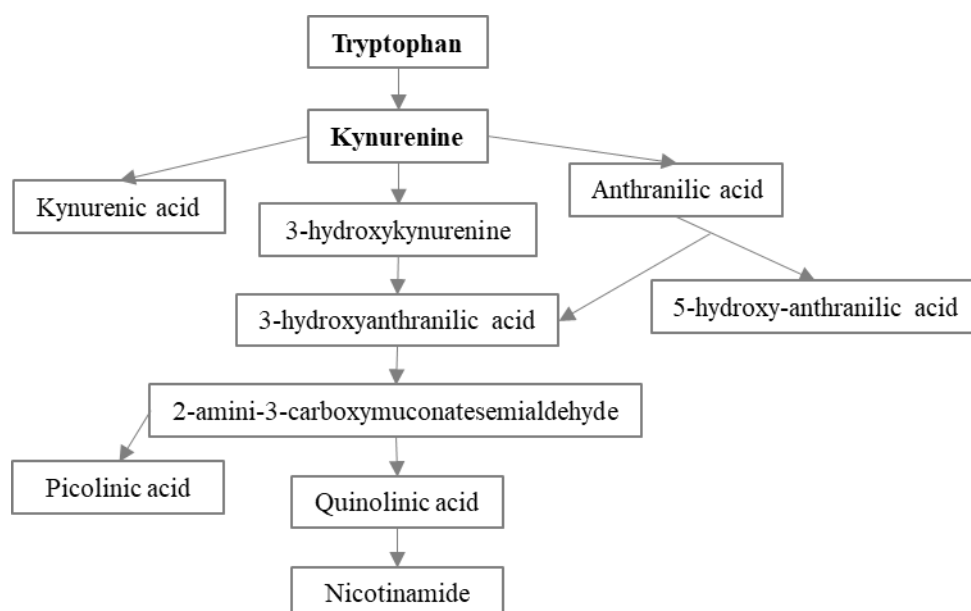


Figure 42 Modified diagram of the kynurenine pathway: during neuroinflammation, 95% of the dietary tryptophan is metabolized along this pathway within the brain [197]

Indole-3-acetic acid showed negative associations with both **SBP** and **fasting glucose** in the tryptophan model (Figure 41). Indole-3-acetic acid increased risk of impaired cognitive function in patients receiving hemodialysis

(a treatment to filter wastes and water from the blood) [198]. Another prospective pilot study at a 2-year follow-up, indole-3-acetic acid was correlated with anxiety (0.52 Spearman's ρ and $p=0.005$) and depression (0.39 Spearman's ρ and $p < 0.05$) in twenty-seven adults with chronic kidney disease [199].

The **kynurenine** concentration showed a positive association with fasting glucose (and a negative association regarding the composite cognitive score) in the tryptophan model (Table 24). The kynurenine levels are higher in those with major **depression** [200-202]. Myint *et al.* (2007) reported that the reduction in neuroprotective markers might play an important role in pathophysiology of major depression, in the study of investigating whether the neurodegeneration proposed major depression as a consequence of the imbalance between neuroprotective and neurodegenerative metabolites in the kynurenine pathway [203]. The imbalanced kynurenine metabolism in terms of neuroprotective and neurotoxic effects was demonstrated in major depression, and in drug-induced neuropsychiatric side effects such as interferon-treated depression [202]. For (abdominal) obesity, **serine** showed a negative association with WC in the amino acids model (Table 23). In the total amino acids model applied BMI, **hydroxylysine** showed a positive association with BMI: Coef. (95% CIs) 0.065 (0.05, 0.08).

All the information from the total models of NMR metabolites, amino acids, and tryptophan regarding the cognitive score, was merged into the brain MRI models. The following section will investigate the brain MRI parameters in the association between each cardio-metabolic risk factor and cognition.

2. Predictions of cognition with the brain magnetic resonance imaging (MRI)-derived parameters

Total ICV is defined as the volume within the cranium, including the brain, meninges and cerebrospinal fluid [170]. There are wide interindividual and sex-dependent variations in head size. In the study of normalisation of cerebral volumes by ICV, it was hypothesized that total ICV could be used to control for variations in voxel size over time, i.e. ICV normalisation [170]. In the regression analysis, the ICV parameter can be included to accounts for the individual head size. The image of brain atrophy (Figure 43), underlines that the ICV itself does not provide the information of brain aging, as an individual has different brain sizes. Brain aging is rather associated with WM and GM shrinkages in this example.

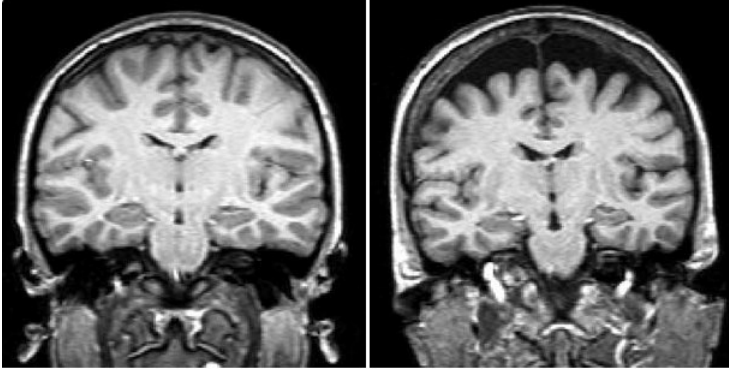


Figure 43 An image slice of a young man's brain (left) and an image of an elderly man's brain (right); both acquired near the midline position. The loss of brain cells is apparent in the elderly participant [204]. Reused with permission from the corresponding author and the journal: License Number 4983770658728.

The upper triangle of a partial Spearman's rank correlation coefficient matrix for ICV adjusted brain MRI parameters are presented in Appendix Figure 21; volumes of WM and temporal lobe WM right (0.7), WM and temporal lobe WM left (0.72), temporal lobe WM right and left (0.78), right and left hippocampus (0.85), total brain and GM (0.82), cerebrospinal fluid and total brain (-0.86), cerebrospinal fluid and GM (-0.72). In order to examine which brain MRI parameters are significant regarding the adjusted composite cognitive score (i.e. cognition) at Year 30, I initially conducted a Bayesian multilevel linear regression on the fixed effects with random intercepts for race and sex. (Race and sex are confounding factors in many associations in the structural equation models, Table 20, Table 22, Table 24). Accounting for the variability between sex of race would help to explain some of between-race variability. The reported estimates of posterior means and posterior standard deviations for the Bayesian model parameters are similar to the corresponding maximum-likelihood estimates and standard errors reported by multilevel mixed-effects linear regression [205].

The Bayesian three-level random-intercept regression model at Year 30 is as follows:

$$y = X\beta + Z\mathfrak{u} + \epsilon$$

where, Z is $n \times q$ design matrix and $u_j, j \in \{1, \dots, q\}$, are random-effects parameters. u_j 's are assigned a hierarchical prior, $u_j | \mu, \sigma_u^2 \sim i.i.d. N(\mu, \sigma_u^2)$ where μ and σ_u^2 are hyper-parameters.

The fixed effects included in each model of the adjusted composite cognitive score on a brain MRI parameter are total ICV in order to adjust each individual brain structure size, and four cardio-metabolic risk factors (i.e. SBP, fasting glucose, BMI, 10-Year Framingham Risk), and other exposures included in the structural equation models. Other exposures included in the model are lifestyle exposures (i.e. sleep quality, family income, depression score (CES-D), socio-demographic exposures (i.e. age, education up to Year 30, smoking status), and clinical exposures (i.e. medication for hypertension and for depression).

As an example, to examine if total brain volume (cc^3) is significant regarding the adjusted cognitive score, the total ICV was included in the model as a regressor of non-interest but an adjustment factor, along with other adjustments in the two level (i.e. race and sex) random effect models.

The 95% credible intervals (CrIs) contain 95% of the most plausible parameters of a particular dataset [206]. The brain MRI parameters which are supported by median and 95% CrIs from each Bayesian multilevel linear regression are original values of total brain volume (cc³); 0.000011 (0.000004, 0.000019), total GM volume (cc³); 0.000009 (0.000003, 0.000019), total WM volume (cc³); 0.000014 (0.000001, 0.000027), total cerebrospinal fluid volume (cc³); -0.000011 (-0.000019, -0.000004), corrected mean FA in right MFC (normalized); 13.57 (3.39, 23.58), Appendix Figure 22–25. When more brain MRI parameters are examined in the Bayesian multilevel linear regressions with the same adjusted variables, the following parameters were also significant: the total volume of the left side of GM and WM, percent change of vascular reactivity in the right hippocampus and the temporal lobe WM right, and the number of activated voxels within the right entorhinal area; and FA in the left entorhinal area was additionally significant when WC is included in the model, instead of BMI. These brain MRI parameters are significant regarding the adjusted cognitive score at Year 30, after adjustment for potential confounders. They are considered as essential brain MRI measurements to be included in the brain MRI models (in Section 3). As total brain volume (cc³) includes both GM and WM tissues, GM and WM tissues are proffered to be included in the model.

2-1. Brain MRI parameters on cardio-metabolic risk factors

Brain volume loss

It has been reported that inflammation, oxidative stress, and endothelial dysfunction (i.e. a non-obstructive coronary artery disease (CAD) in which the large blood vessels on the heart's surface constrict) increase the risk of AD, and these biomarkers are elevated in patients with AF [207]. The total brain volume, GM volume, and frontal GM volume are known as larger brain structures for which gradually declines at later stages of AD (i.e. Braak Stages V and VI, which there is extensive neocortical involvement such as sensory perception, cognition, generation of motor commands, spatial reasoning and language) (Figure 44) [17,18].

⁶ Standardized scales of total brain volume: 1.35 (0.44, 2.25), total GM volume; 0.59 (0.03, 1.17), total WM volume; 0.90 (0.07, 1.72), total cerebrospinal fluid volume; -0.59 (-0.99, -0.19), corrected mean FA in right MFC (normalized); 0.36 (0.09, 0.63).

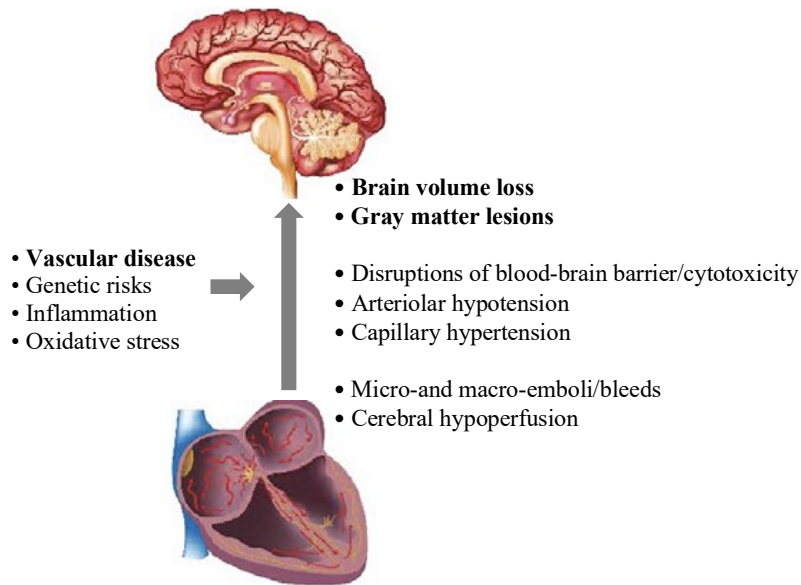


Figure 44 Risk of cognitive decline and dementia in patients with atrial fibrillation (AF) (modified image) [207]

Gray matter volume loss/ gray matter lesions

Giorgio *et al.* (2010) stated that there are a number of potential cellular processes that could explain these volume reductions in GM. GM volume reductions could be due to shrinkage of large neurons, or rarefaction in the GM microvasculature, resulting in a loss of neurons in middle and late adulthood [208-212]. Nonlinear (i.e. quadratic age effects on GM density) or more complex in associative frontal and parietal cortices and posterior temporal cortex, which continue myelination (the process of forming a myelin sheath around a nerve to allow nerve impulses to move more quickly) have been previously found in adulthood [213].

White matter volume/ white matter lesions

Ge *et al.* (2002) stated that the means of both GM and WM magnetization transfer ratio (MTR, an identifier of underlying structural changes in the brain) histograms demonstrated a quadratic relationship with age, showed a slight increase with age until middle adulthood [214].

Giorgio *et al.* (2010) reported nonlinear (quadratic) relationships between WM volume and age in the superior corona radiata (the innermost layer of the cells of the cumulus oophorus) bilaterally, and in the left superior longitudinal fascicle (a large bundle of association fibres in the WM of each cerebral hemisphere) [212]. It could be interpreted that both heavily myelinated fibres (e.g. corticospinal tracts, located in the posterior limb of the internal capsule) and thinly myelinated fibres (e.g. association fibres of the frontal lobe) are affected in the same way by the pathogenic mechanisms underlying the ageing process (Figure 45) [212].

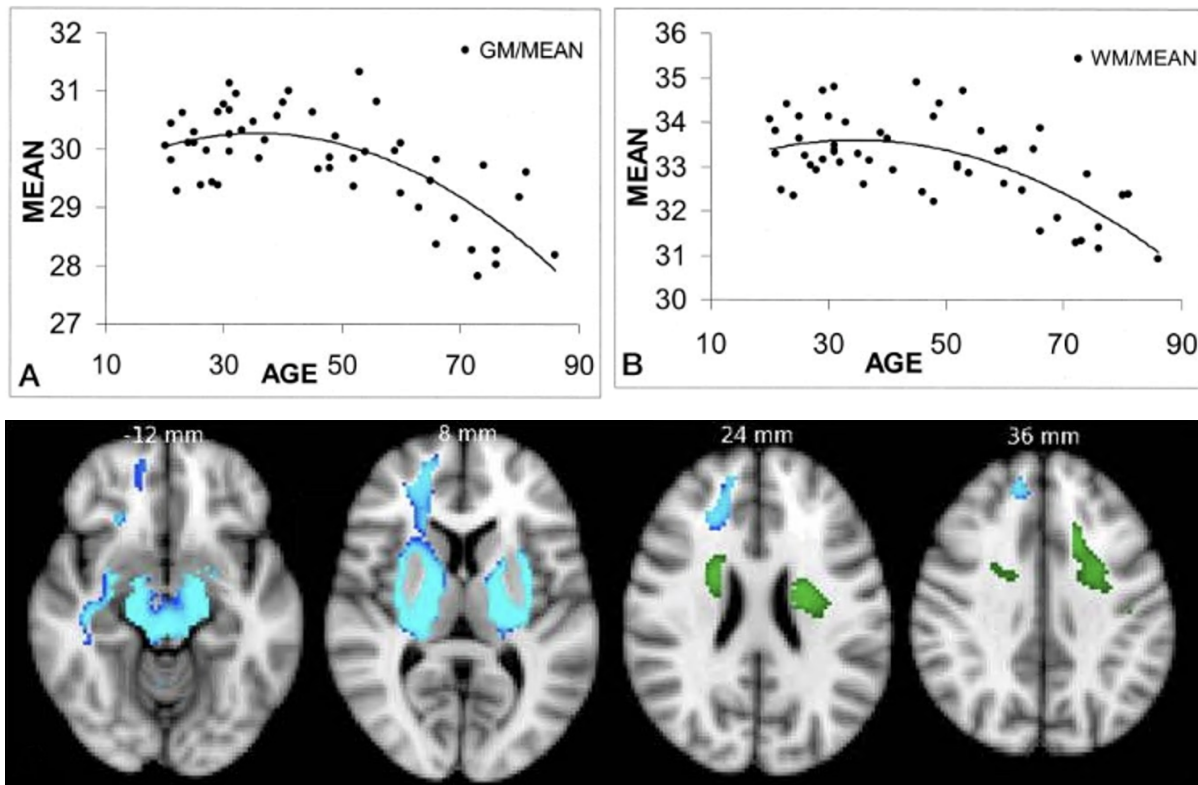


Figure 45 Gray matter (A) and white matter (B) magnetization transfer ratio histograms. The graph is from a paper by Ge *et al.* (2002) [214]. The below figure represents voxel-based morphometry style analysis of white matter changes with age. Voxel based morphometry (VBM) is a technique using MRI that allows the investigation of focal differences in brain anatomy, using the statistical approach of parametric mapping (Bal *et al.*, Handbook of Clinical Neurology, 2014). Coloured voxels show regions where white matter volume shows a significant linear (blue) or non-linear (green) relationship with age ($p < 0.05$, fully corrected for multiple comparisons across space). The image from a paper by Giorgio *et al.*, 2010 [212] [Permission is not required for this type of reuse: <https://creativecommons.org/licenses/by/3.0/>]

Giorgio *et al.* (2010) described that WM atrophy (decrease in size) could be a result of the ageing process by itself or a consequence of the effects of WML, which may be present in the neurologically asymptomatic elderly population, but not all the studies found association between age-related atrophy of WM fibres and the presence of WML [212, 215-217]. Age-related WML (areas of increased brightness on T2 weighted MRI images) are reported as a risk factor for stroke, cognitive decline, and dementia [35]. The research shows that CHD was associated with thinner cortex and lower total GM volume in people with midlife hypertension (and in people with declining blood pressure after midlife) (Figure 46) [53].

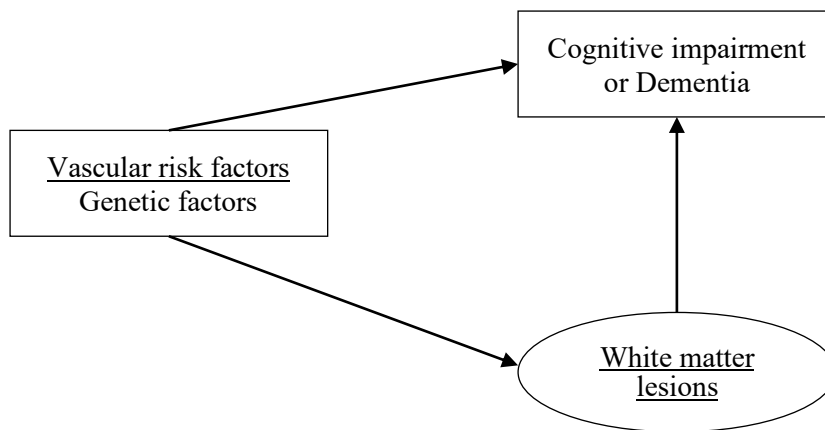


Figure 46 Simplified connections among white matter lesions and cognitive impairment (excluding hypertension and stroke; Sierra and Coca, 2006) [90]

2-2. Total path diagram of the brain MRI parameters

The hypothesis of the present study regarding brain MRI parameters is in Chapter I, Section 4, describes that a cardio-metabolic risk factor has a detrimental effect on a brain MRI parameter which is a marker of cognitive function. The present section will explain which brain MRI parameters are mediators of a cardio-metabolic risk factor, and which exposures are confounding factors in each association regarding the adjusted composite cognitive score measured at Year 30 (Table 26).

Concepts from the literature review concerning the connections among vascular risk factors, several brain MRI measures (such as brain/white/gray matter lesions or volume loss), and cognitive impairment or dementia were applied (Figure 44 and Figure 46). The adjusted composite cognitive score measured at Year 30 was consistently used for the examinations of midlife vascular risk factors and the brain health associations in the present study.

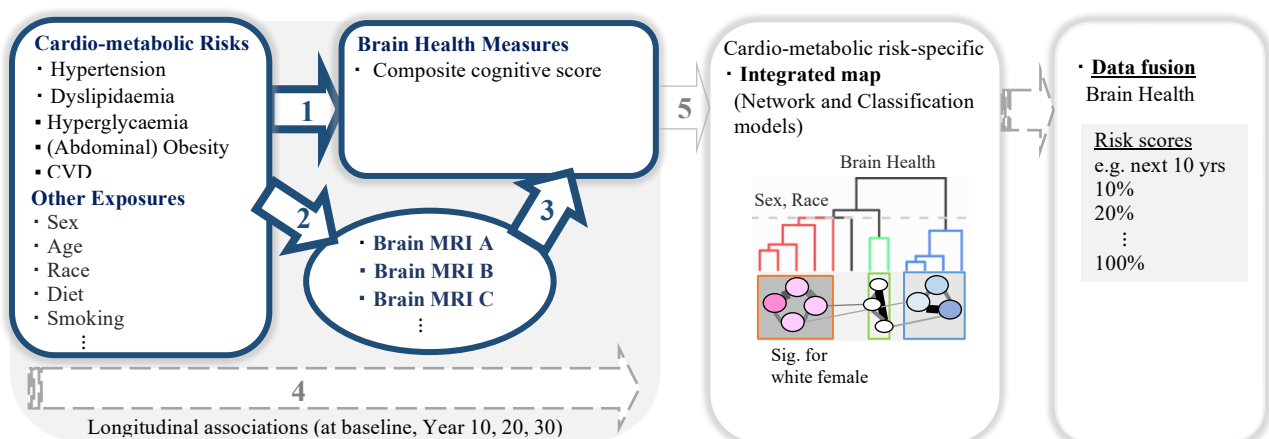


Figure 47 Schematic of the study highlighting the work focusses on Objectives 1 to 3. In the present section, brain MRI parameters are hypothesised as mediators regarding cognition on a cardio-metabolic risk. A mediation analysis is comprised of three sets of regressions: $X \rightarrow Y$, $X \rightarrow M$, and $X + M \rightarrow Y$, Baron and Kenny's (1986) [89]

The same analytic approaches used in the previous section were applied to examine the associations among the brain health measurements (cognition and the brain MRI parameters), cardio-metabolic factors at Year 30. SBP has nine mediators (such as total right MFC and ACG volume, abnormal brain and WM tissue volume, abnormal temporal lobe WM, and FA in WM and FA in temporal lobe WM). As DBP has fewer (five) mediators, SBP was selected to be included in the total brain MRI model as a hypertension factor. No factors of dyslipidaemia were significant for the adjusted composite cognitive score. In addition, eGFR was not significant (a hyperglycaemia factor) for the adjusted composite cognitive score (as it was stated in 1. Predictions of cognition based on urinary metabolites).

Among the hyperglycaemia factors (i.e. fasting glucose, HOMA-IR, eGFR, fasting insulin), fasting glucose has fourteen mediators including total brain volume measurements, CBF measurements, FA in ACG, and total temporal lobe WM volumes. The (abdominal) obesity factors, WC has sixteen mediators including total brain volume measurements and a number of CBF measurements, whereas BMI has six mediators. The 10-Year Framingham Risk has ten mediators such as a number of total brain volume measurements including temporal lobe WM and right entorhinal area volumes. Once **each model per cardio-metabolic risk factor has been built**, only ‘significant’ cardio-metabolic factors, exposures, mediators, and confounding factors were included in the final model (as described in Section 1). In addition, only one cardio-metabolic risk factor ‘per category’ was included in the final model, with preference for those possessing ‘more mediators’ in each association (Table 26). The standardized scales of the brain MRI parameters (i.e. z-score of MRI measurements) were applied in the models for additional benefits and to include those from originally different units all together in the final model. The bootstrapped (500 bootstrap replications) standard errors and confidence intervals were obtained.

Table 26 Significant exposures, confounding factors, and brain MRI parameter mediators in the structural equation models predicting the adjusted composite cognitive score, for each cardio-metabolic factor; each dot represents a mark of mediator

		Hypertension		Hyperglycaemia			(Abdominal) Obesity		CVD
		SBP	DBP	Fasting glucose	Homa-IR	Fasting insulin	BMI	WC	10-Year Framingham Risk
Significant exposures	Race	•	•		•	•	•	•	•
	Sex	•	•		•	•	•	•	
	Age								
	Education	•	•		•	•	•	•	•
	CES-D								
	Smoking								
	HBP medication	•	•						
	Depression medication						•		
Confounding factors	Sleep quality								
	Family income	•	•	•	•	•	•	•	•
	Conf Race	•	•	•			•		•
Significant brain MRI parameters	Conf Sex			•				•	•
	Conf Education			•					
	Total Intracranial Volume			•				•	•
	Total Brain Volume			•				•	•
	Total Cerebrospinal Fluid Volume			•	•			•	•
	Total White Matter Volume			•				•	•
	Total Gray Matter Volume			•				•	•
	Total Right Hippocampus Volume							•	

(Table Continued)

[illegible]

* ACG: Anterior cingulate gyrus

MFC: Medial frontal cortex

* eGFR (Hyperglycaemia) was not significant for Adj. composite Cognition ($p=0.05$).

* HDL-C, LDL-C and Triglycerides (Dyslipidaemia) were not significant for Adj. composite Cognition ($p=0.05$).

* CES-D, smoking, and sleep quality were not significant in any structural equation models by each cardio-metabolic factor.

The total path diagram of the brain MRI parameters is presented in Figure 48 and Appendix Table 10. The structural equation model is presented for the adjusted composite cognitive score on cardio-metabolic risk factors (SBP, Fasting glucose, WC, and Framingham risk) with brain MRI parameters (Log likelihood = -9919.213, N=155 with replications 500).

Race, sex and education up to Year 30 were significant regarding the adjusted composite cognitive score. There was a significant association ($p < 0.05$) between SBP and each standardized FA in WM (coefficient (Coef.) -0.014), FA in temporal lobe WM right (Coef. -0.017) and left (Coef. -0.014). Fasting glucose was found to be significant regarding the standardized total WM volume (Coef. 0.015) and total temporal lobe WM left (Coef. 0.016). WC was found to be significant regarding each standardized CBF measurement; left ACG (Coef. -0.013) and left entorhinal area (Coef. -0.015). The 10-Year Framingham Risk was significant regarding the standardized total right entorhinal area volume (Coef. 0.067) (Appendix Table 10).

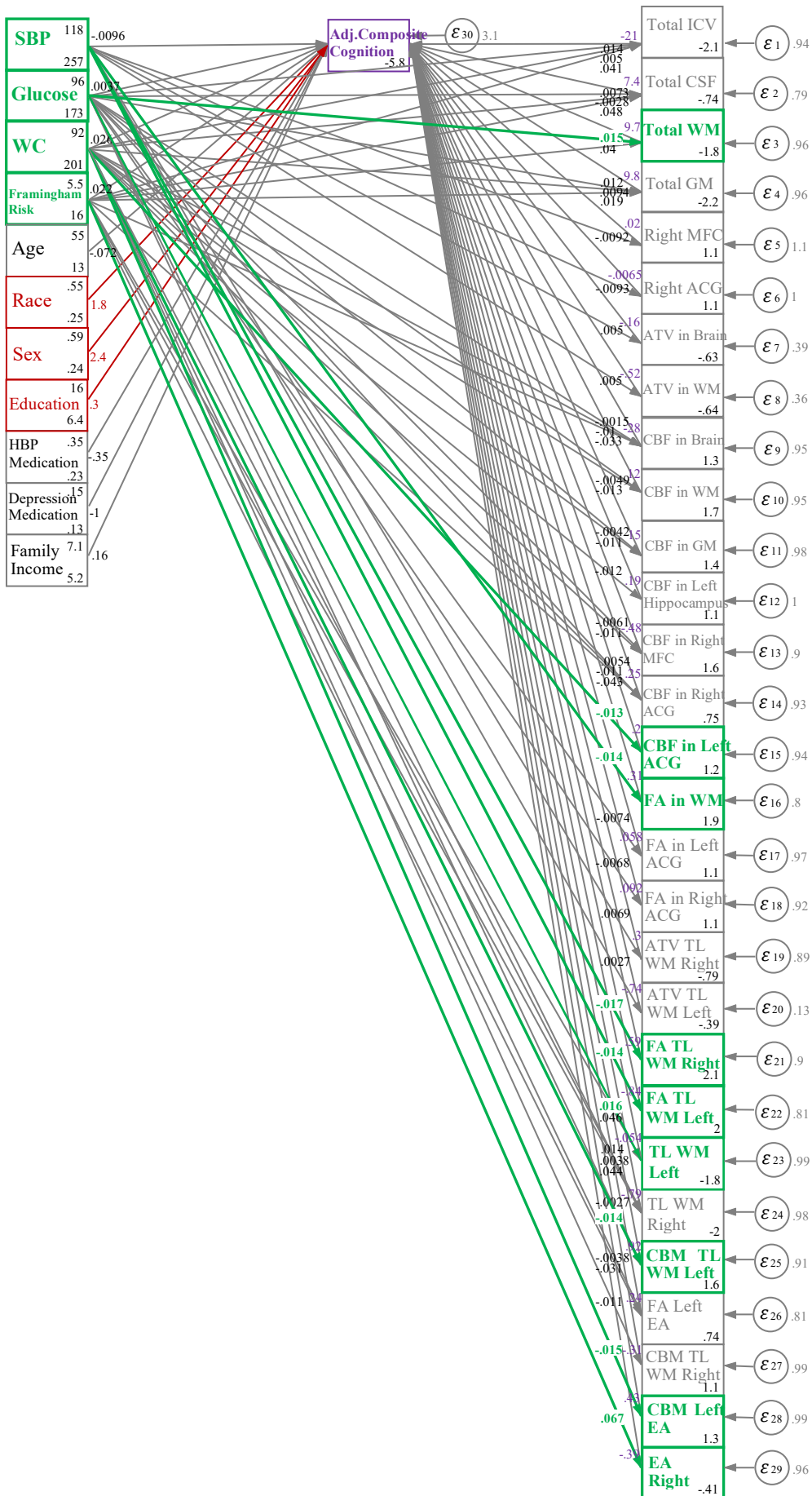


Figure 48 Total multivariate regression path diagram for adjusted composite cognitive score on cardio-metabolic risk factors (SBP, Glucose, WC, and Framingham risk) with brain MRI (N= 155)

2-3. Total path diagrams of the metabolites and brain MRI parameters – Summary and discussion

Naturally occurring amino acids can be classed on the basis of their structure and the general chemical characteristics of their side chains as follows; glycine, alanine, valine, leucine, and isoleucine (aliphatic side chains); serine (a hydroxyl or sulfur/selenium-containing group); and phenylalanine, tyrosine, and tryptophan (aromatic side chains) [218]. Quite a few metabolites in the classified groups were identified as mediators in the models; summary of the significant metabolites are presented in Table 27. The information of the associations between urinary metabolites and cardio-vascular risks (i.e. valine and WC; glycine and fasting glucose) in the total model with metabolites by NMR (Table 27 (1)) could be covered by the associations between valine/glycine and fasting glucose by the total amino acids model (Table 27 (2)). Methionine and SBP; a number of amino acids and fasting glucose; serine and WC are significantly associated, respectively, in the total amino acids model. The missing cases of glycine (N=544; 9.3 missing %) and valine (N=400; 33.3 missing %) by NMR are greater than the missing cases of glycine (N= 599; 0.17 missing %) and valine (N= 597; 0.5 missing %) by LC–MS.

Table 27 Significant metabolites in the structural equation models for the adjusted composite cognitive score on cardio-metabolic risk factors (Summary of Table 21, Table 23, Table 25)

(1) Significant metabolites measured by NMR spectroscopy on cardio-metabolic risk factors (SBP, fasting glucose, WC, and 10-Year Framingham Risk) for adjusted composite cognition in the SEM (N=346)

Urinary Metabolites	Cardio-vascular Risk	Observed Coef.	$P > z $	Normal-based [95% CIs]	
Valine ←	WC	0.012	0.001	0.005	0.019
Glycine ←	Fasting glucose	-0.007	<0.001	-0.010	-0.004

* Only significant paths from the total multivariate linear SEM are listed.

* Age, race, sex, education up to Year 30, HBP medication, and family income were significant regarding cognition.

(2) Significant amino acids measured by LC–MS on cardio-metabolic risk factors (SBP, fasting glucose, WC, and 10-Year Framingham Risk) for adjusted composite cognition in the SEM (N=420)

Urinary Metabolites	Cardio-vascular Risk	Observed Coef.	$P > z $	Normal-based [95% CIs]	
Methionine ←	SBP	0.005	0.016	0.001	0.009
Glycine ←	Fasting glucose	-0.004	<0.001	-0.006	-0.002
Isoleucine ←	Fasting glucose	0.007	<0.001	0.004	0.010
Leucine ←	Fasting glucose	0.010	<0.001	0.006	0.013
Phenylalanine ←	Fasting glucose	0.003	0.020	0.001	0.006
Valine ←	Fasting glucose	0.011	<0.001	0.008	0.014
Aminoadipic acid ←	Fasting glucose	0.006	0.011	0.001	0.010
Serine ←	WC	-0.012	<0.001	-0.016	-0.009

* Only significant paths from the total multivariate linear SEM are listed.

* Age, race, sex, education up to Year 30, and family income were significant regarding cognition.

(3) Significant tryptophan measured by LC–MS on cardio-metabolic risk factors (SBP, fasting glucose, WC, and 10-Year Framingham Risk) for adjusted composite cognition in the SEM (N=413)

Urinary Metabolites	Cardio-vascular Risk	Observed Coef.	$P > z $	Normal-based [95% CIs]	
Indole-3-acetic acid ←	Fasting glucose	-0.0072	<0.001	-0.011	-0.003
Tryptophan ←	Fasting glucose	0.0063	0.015	0.001	0.011
Kynurenine ←	Fasting glucose	0.0067	0.001	0.003	0.011

* Only significant paths from the total multivariate linear SEM are listed.

* Age, race, sex, education up to Year 30, CES–D, and family income were significant regarding cognition.

Furthermore, I investigated whether a cardio-metabolic risk factor has a detrimental effect on neurocognitive pathology using the brain MRI parameters (Table 28). The investigation was expanded to include the urinary metabolites (which can cause either neuroprotective or neurotoxic effects) in Chapter III, Section 3.

Table 28 Significant brain MRI parameters on cardio-metabolic risk factors (SBP, fasting glucose, WC, and 10-Year Framingham Risk) for adjusted composite cognition in the SEM (N=155)

Brain MRI parameters	Cardio-vascular Risk	Observed Coef.	$P > z $	Normal-based [95% CIs]	
Fractional Anisotropy in White Matter ←	SBP	-0.014	0.007	-0.024	-0.004
Fractional Anisotropy in Temporal Lobe WM Right ←	SBP	-0.017	0.003	-0.028	-0.006
Fractional Anisotropy in Temporal Lobe WM Left ←	SBP	-0.014	0.006	-0.024	-0.004
Total White Matter Volume ←	Fasting glucose	0.015	0.023	0.002	0.028
Total Temporal Lobe WM Left ←	Fasting glucose	0.016	0.024	0.002	0.029
Cerebral Blood Flow in Left Anterior Cingulate Gyrus ←	WC	-0.013	0.016	-0.024	-0.002
Cerebral Blood Flow in Temporal Lobe WM Left ←	WC	-0.014	0.027	-0.027	-0.002
Cerebral Blood Flow in Left Entorhinal Area ←	WC	-0.015	0.012	-0.026	-0.003
Total Right Entorhinal Area Volume ←	10-Year Framingham Risk	0.067	0.005	0.021	0.114

* Only significant paths from the total multivariate linear SEM are listed.

* Race, sex, and education up to Year 30 were significant regarding cognition.

3. Predictions of cognition from metabolites and brain MRI parameters

The present section describes the model predicting cognition from cardio-metabolic factors mediated by both urinary metabolites and the brain MRI parameters measured Year 30. All the significant associations between brain MRI parameters with a cardio-metabolic risk factor were included in the total model along with significant associations with the urinary metabolites (Table 29). The advantage of using SEM is that the indirect and total effects can be estimated from a single model, and the simple mediation model can be inserted into a larger model (even using latent variables) to measure any piece of the mediation model [219, 220].

In consideration of estimating the structural equation model with missing data, the information of the significant associations only from the total amino acids model and the total tryptophan model (Table 27 (2) and (3)) were

included; the significant metabolites measured by NMR, i.e. valine and glycine (Table 27 (1)) were included in the total amino acids model.

Total volumes of brain, white, gray and cerebrospinal fluid and in addition, (corrected) mean FA in the right MFC (normalized) were described as statistically/clinically significant parameters in Section 2. Among those measurements, total cerebrospinal fluid and total GM volume were also included in the model with urinary metabolites and brain MRI parameters although they were not significant in the total brain MRI parameter model (Table 28). (The total brain volume was not included in the model as its information can be covered by other brain measurements such as GM/WM volume. The (corrected) mean FA in right MFC (normalized) was also not included in the model, as it was not significant regarding the adjusted composite cognitive score in the total brain MRI parameter model, Table 28.)

3-1. Total path diagram of the metabolites and brain MRI parameters

The synthesized model of urinary metabolites and brain MRI parameters are presented in Figure 49 and Table 29. In the model, which alternatively used the variables merged the left and right side of the significant brain MRI parameters (i.e. summation of FA in temporal lobe WM right and left, CBF in ACG, CBF in temporal lobe WM, CBF in entorhinal area, total entorhinal area), the same significant variables were identified, and additionally the total entorhinal area was significant. However, in terms of maximizing the log likelihood, the model using both left and right brain MRI parameters was reported in the present study (-4573.57 vs -4600.66). Moreover, based on the research into the differences of the right and left-hemisphere cognitive functions (in stroke patients) [221], the outcomes using both right and left MRI parameters were mainly reported in the present study. Bootstrap estimations were obtained requesting 500 replications and convergence of the maximum likelihood estimators in the SEMs was examined (Table 29). Regarding overidentification (i.e. type I error) in small samples, multiple Monte-Carlo simulation studies have been made to evaluate the shown corrections [222]. In the present study, the Swain correction for a small sample size model with many indicators were conducted, providing a suite of corrections to the chi-square overidentification test (i.e. likelihood ratio test of fit) for SEMs; the Swain (1975) scaling factor which has been reported to work even to a sample size to estimated parameter ratio of 2:1, an empirical correction T (MLEC1), and the F-test as the small sample correction to the chi-square test presented (Table 29) [222-227]. Gelman *et al.* (2012) stressed that many of the issues of type I, type S (sign) or type M (magnitude) errors will disappear or at least be substantially ameliorated when viewed within a Bayesian multilevel modelling framework [228]. (The Bayesian multilevel models were employed to examine the degeneration of cognition at Years 25 and 30 in Chapter IV.)

Phenylalanine, aminoadipic acid, and tryptophan were mediators in the significant **positive** associations with fasting glucose in terms of hyperglycaemia, respectively, where fasting glucose itself showed a **negative** association regarding the cognitive score in the cohort (N=74). Indole-3-acetic acid is a mediator showing a

significant **negative** association with fasting glucose (Figure 49 and Table 29). This information is consistent with the outcomes in Section 1 as shown in Figure 41.

I investigated temporal lobes, and entorhinal area as early markers of AD in the present study (as described in Chapter II, Sub-section 1-5) [111, 144]. The total ICV is a mediator in the **positive** association with WC in terms of (abdominal) obesity (N=74). The measures of brain physiology, CBF in temporal lobe WM left, and CBF in left entorhinal area are mediators in the **negative** associations with WC, respectively (Figure 49 and Table 29). This outcome is supported by the review that obesity triggering VaD decreases not only blood supply to the brain, but also increases fat cells that damage the brain WM leading to loss of cognitive and intellectual behaviour [46].

Adipokines, cytokines secreted by adipose tissue, are known to contribute to the development of obesity-associated insulin resistance [229]. Although it is not only secreted from adipose tissue, adipokines are produced by adipocytes which express and secrete numerous peptide hormones and cytokines [230]. Whitmer (2007) suggested that adipokines, hormones, and inflammatory cytokines could explain the association between obesity and increased risk of dementia [231]. However, there are a number of limitations in the research into adipokine and dementia associations. One of the limitations would be anthropometric measures such as BMI or central adiposity is not considered to be a good estimate of the amount of adipose tissue during adulthood [232].

FA in the temporal lobe WM right and left, total WM and total right entorhinal area volumes, and CBF in the left ACG, which were examined to be significant with cardio-vascular risk factors in the brain MRI parameter model (Table 28), were no longer significant in the total model with the metabolites.

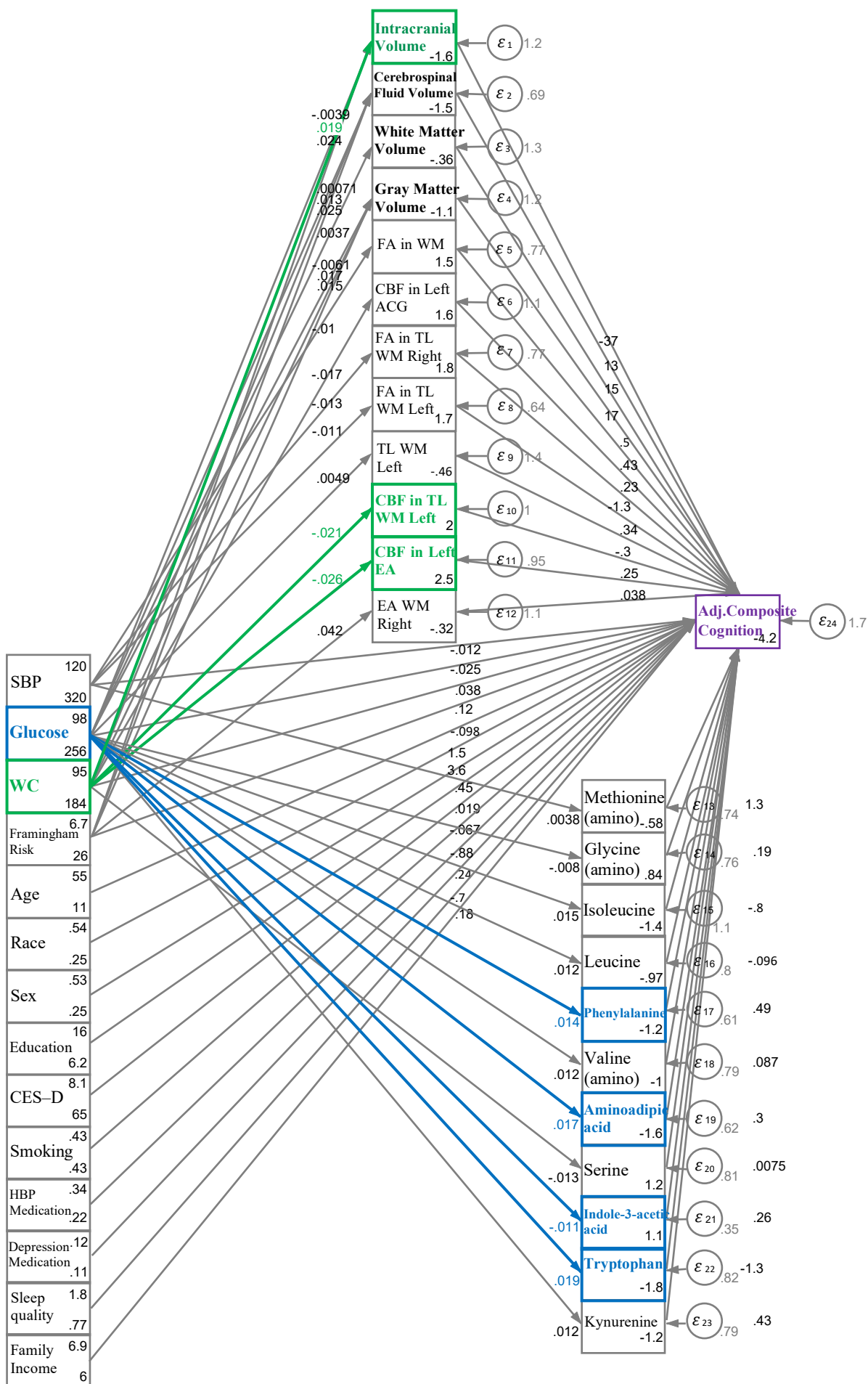


Figure 49 The total multivariate regression path diagram with significant metabolites, the brain MRI-derived parameters, and exposures (N= 74)

Table 29 The total multivariate regression path diagram with significant metabolites, the brain MRI parameters, and exposures (N=74)

Structural equation model
Log likelihood = -4573.5697

Number of observations = 74
Replications = 500

		Observed Coef.	Bootstrap Std. Err.	z	P > z	Normal-based [95% CIs]	
Adj. composite Cognition ←	Total Intracranial volume	-36.661	52.09	-0.70	0.48	-138.75	65.42
	Total Cerebrospinal Fluid volume	13.364	18.43	0.73	0.47	-22.76	49.49
	Total White Matter volume	15.318	22.54	0.68	0.50	-28.86	59.50
	Total Gray Matter volume	16.557	22.34	0.74	0.46	-27.24	60.35
	FA in White Matter	0.499	2.11	0.24	0.81	-3.64	4.64
	CBF in Left ACG	0.426	1.10	0.39	0.70	-1.72	2.58
	FA in Temporal Lobe WM Right	0.226	2.46	0.09	0.93	-4.59	5.04
	FA in Temporal Lobe WM Left	-1.349	2.08	-0.65	0.52	-5.43	2.73
	Total Temporal Lobe WM Left	0.340	2.13	0.16	0.87	-3.84	4.52
	CBF in Temporal Lobe WM Left	-0.303	1.28	-0.24	0.81	-2.81	2.20
	CBF in Left Entorhinal Area	0.253	1.17	0.22	0.83	-2.05	2.55
	Total Right Entorhinal Area	0.038	1.02	0.04	0.97	-1.95	2.03
	Methionine (amino acids)	1.280	1.30	0.98	0.33	-1.28	3.84
	Glycine (amino acids)	0.193	1.56	0.12	0.90	-2.86	3.25
	Isoleucine	-0.797	1.78	-0.45	0.66	-4.29	2.70
	Leucine	-0.096	2.06	-0.05	0.96	-4.13	3.94
	Phenylalanine	0.492	2.20	0.22	0.82	-3.81	4.80
	Valine (amino acids)	0.087	1.88	0.05	0.96	-3.61	3.78
	Aminoadipic acid	0.302	0.98	0.31	0.76	-1.63	2.23
	Serine	0.008	1.51	0.00	1.00	-2.95	2.96
	Indole-3-acetic acid	0.258	1.05	0.25	0.81	-1.80	2.32
	Tryptophan	-1.333	1.70	-0.78	0.43	-4.66	2.00
	Kynurenine	0.428	0.91	0.47	0.64	-1.36	2.22
	SBP	-0.012	0.05	-0.24	0.81	-0.11	0.09
	Fasting glucose	-0.025	0.05	-0.50	0.62	-0.12	0.07
	WC	0.038	0.07	0.57	0.57	-0.09	0.17
	10-Year Framingham Risk	0.117	0.22	0.54	0.59	-0.31	0.54
	Age	-0.098	0.21	-0.47	0.64	-0.50	0.31
	Race	1.525	1.98	0.77	0.44	-2.36	5.41
	Sex	3.601	2.19	1.64	0.10	-0.70	7.90
	Education	0.453	0.28	1.64	0.10	-0.09	0.99
	CES-D	0.019	0.10	0.20	0.84	-0.17	0.21
	Smoking	-0.067	1.16	-0.06	0.95	-2.34	2.21
	HBP medication	-0.881	1.59	-0.55	0.58	-4.00	2.24
	Depression medication	0.235	2.12	0.11	0.91	-3.92	4.39
	Sleep quality	-0.697	0.83	-0.84	0.40	-2.32	0.92
	Family income	0.176	0.27	0.64	0.52	-0.36	0.71
	constant	-4.250	18.83	-0.23	0.82	-41.15	32.66
Total Intracranial Volume ←	WC	0.019	0.01	2.33	0.02	0.003	0.04
	Fasting glucose	-0.004	0.01	-0.33	0.74	-0.03	0.02
	10-Year Framingham Risk	0.024	0.04	0.67	0.51	-0.05	0.09
	constant	-1.645	1.14	-1.44	0.15	-3.89	0.60
Total Cerebrospinal Fluid Volume ←	WC	0.013	0.01	1.86	0.06	-0.001	0.03
	Fasting glucose	0.001	0.01	0.07	0.94	-0.02	0.02
	10-Year Framingham Risk	0.025	0.03	0.97	0.33	-0.02	0.07
	constant	-1.524	0.97	-1.58	0.12	-3.42	0.37
Total White Matter Volume ←	Fasting glucose	0.004	0.01	0.39	0.70	-0.02	0.02
	constant	-0.362	0.92	-0.39	0.69	-2.17	1.44
Total Gray Matter Volume ←	WC	0.017	0.01	1.82	0.07	0.00	0.03
	Fasting glucose	-0.006	0.01	-0.53	0.60	-0.03	0.02
	10-Year Framingham Risk	0.015	0.04	0.39	0.70	-0.06	0.09
	constant	-1.124	1.18	-0.95	0.34	-3.45	1.20
FA in White Matter ←	SBP	-0.010	0.01	-1.68	0.09	-0.02	0.002
	constant	1.537	0.76	2.01	0.04	0.04	3.03
CBF in Left ACG ←	WC	-0.017	0.01	-1.95	0.05	-0.03	0.0001
	constant	1.557	0.76	2.04	0.04	0.06	3.05

(Table Continued)

FA in Temporal Lobe WM Right ←	SBP constant	-0.013 1.787	0.01 0.86	-1.89 2.08	0.06 0.04	-0.03 0.11	0.001 3.47
FA in Temporal Lobe WM Left ←	SBP constant	-0.011 1.667	0.01 0.71	-1.93 2.36	0.05 0.02	-0.02 0.28	0.0002 3.05
Total Temporal Lobe WM Left ←	Fasting glucose constant	0.005 -0.463	0.01 0.98	0.48 -0.47	0.63 0.64	-0.02 -2.39	0.03 1.47
CBF in Temporal Lobe WM Left ←	WC constant	-0.021 1.958	0.01 0.77	-2.42 2.55	0.02 0.01	-0.04 0.45	-0.004 3.46
CBF in Left Entorhinal Area ←	WC constant	-0.026 2.520	0.01 0.76	-3.25 3.31	0.001 0.001	-0.04 1.03	-0.01 4.01
Total Right Entorhinal Area ←	10-Year Framingham Risk constant	0.042 -0.323	0.02 0.17	2.00 -1.88	0.05 0.06	0.001 -0.66	0.08 0.01
Methionine (amino acids) ←	SBP constant	0.004 -0.580	0.005 0.60	0.76 -0.96	0.45 0.34	-0.01 -1.76	0.01 0.60
Glycine (amino acids) ←	Fasting glucose constant	-0.008 0.845	0.01 0.58	-1.41 1.45	0.16 0.15	-0.02 -0.30	0.00 1.99
Isoleucine ←	Fasting glucose constant	0.015 -1.370	0.01 0.76	1.81 -1.81	0.07 0.07	-0.001 -2.85	0.03 0.11
Leucine ←	Fasting glucose constant	0.012 -0.968	0.01 0.62	1.80 -1.57	0.07 0.12	-0.001 -2.18	0.02 0.24
Phenylalanine ←	Fasting glucose constant	0.014 -1.249	0.01 0.54	2.38 -2.31	0.02 0.02	0.002 -2.31	0.03 -0.19
Valine (amino acids) ←	Fasting glucose constant	0.012 -0.998	0.01 0.65	1.74 -1.53	0.08 0.13	0.00 -2.27	0.03 0.28
Aminoadipic acid ←	Fasting glucose constant	0.017 -1.615	0.01 0.76	2.22 -2.14	0.03 0.03	0.002 -3.10	0.03 -0.13
Serine ←	WC constant	-0.011 1.167	0.01 0.70	-1.51 1.66	0.13 0.10	-0.03 -0.21	0.00 2.55
Indole-3-acetic acid ←	Fasting glucose constant	-0.013 1.078	0.00 0.44	-3.05 2.45	0.002 0.01	-0.02 0.21	-0.005 1.94
Tryptophan ←	Fasting glucose constant	0.019 -1.796	0.01 0.57	3.07 -3.13	0.002 0.002	0.01 -2.92	0.03 -0.67
Kynurenine ←	Fasting glucose constant	0.012 -1.216	0.01 0.67	1.78 -1.81	0.08 0.07	-0.001 -2.54	0.03 0.10
var (e.Adj. composite Cognition)		1.700	0.28			1.24	2.34
var (e.Methionine (amino acids))		0.735	0.12			0.54	1.00
var (e.FA in Temporal Lobe WM Right)		0.770	0.12			0.57	1.04
var (e.FA in Temporal Lobe WM Left)		0.641	0.10			0.47	0.87
var (e.FA in White Matter)		0.774	0.13			0.56	1.08
var (e.CBF in Left ACG)		1.147	0.23			0.77	1.70
var (e.Total Intracranial Volume)		1.202	0.20			0.87	1.67
var (e.Total Cerebrospinal Fluid Volume)		0.691	0.11			0.51	0.93
var (e.Total Gray Matter Volume)		1.221	0.18			0.92	1.63
var (e.CBF in Temporal Lobe WM Left)		1.008	0.19			0.70	1.46
var (e.CBF in Left Entorhinal Area)		0.953	0.25			0.57	1.59
var (e.Serine)		0.805	0.17			0.53	1.22
var (e.Total White Matter Volume)		1.342	0.22			0.98	1.85
var (e.Total Temporal Lobe WM Left Volume)		1.398	0.25			0.99	1.98
var (e.Glycine (amino acids))		0.763	0.19			0.47	1.23
var (e.Isoleucine)		1.133	0.49			0.49	2.64
var (e.Leucine)		0.797	0.19			0.50	1.28
var (e.Phenylalanine)		0.608	0.14			0.39	0.95
var (e.Valine (amino acids))		0.788	0.20			0.48	1.28
var (e.Aminoadipic acid)		0.622	0.10			0.45	0.86
var (e.Indole-3-acetic acid)		0.346	0.06			0.24	0.49
var (e.Tryptophan)		0.822	0.23			0.48	1.42
var (e.Kynurenine)		0.788	0.18			0.50	1.23
var (e.Total Right Entorhinal Area Volume)		1.053	0.17			0.77	1.44

* FA - Fractional anisotropy; CBF - Cerebral blood flow; ACG - Anterior cingulate gyrus

* Swain corrected chi-square = 2336.58; $p < 0.001$ * Yuan-Tian-Yanagihara empirically corrected chi-square = 2311.38; $p < 0.001$ * F-test value = 5.39; $p < 0.001$

There are several studies providing important insight into the potential pathophysiologic mechanisms of increased adiposity in the development of vascular cognitive aging. In the Framingham Offspring Study, studying middle-aged community participants (N=733, mean age 60 years), a significant **negative** association of BMI, WC, waist to hip ratio (WHR), subcutaneous adipose tissue (SAT) and especially the strongest visceral (abdominal fat) with total brain volume, independent of vascular risk factors were observed [233]. Pannacciulli *et al.* (2006) reported in a study of 24 obese and 36 lean, early 30s non-diabetic Caucasians, that BMI was negatively associated with GM density of the left post-central gyrus in obese but not lean subjects. The study identified structural brain differences in human obesity in several brain areas previously involved in the regulation of taste, reward, and behavioural control [234]. Raji *et al.* (2010) reported that higher BMI (25 to 30 kg/m²) was associated with lower brain volumes in 94 overweight and obese elderly (over age 70) subjects who remained cognitively normal (for at least 5 years after the scans). The study reported that obese subjects with a high BMI (> 30 kg/m²) showed atrophy in the frontal lobes, ACG, hippocampus, and thalamus compared with individuals with a normal BMI (18.5 to 25 kg/m²), using the multiple regression model. In addition, overweight subjects BMI (25 to 30 kg/m²) had atrophy in the basal ganglia and corona radiata of the WM [235]. BMI and WC were highly correlated (Spearman's $\rho=0.86$, $p < 0.001$) in the present study.

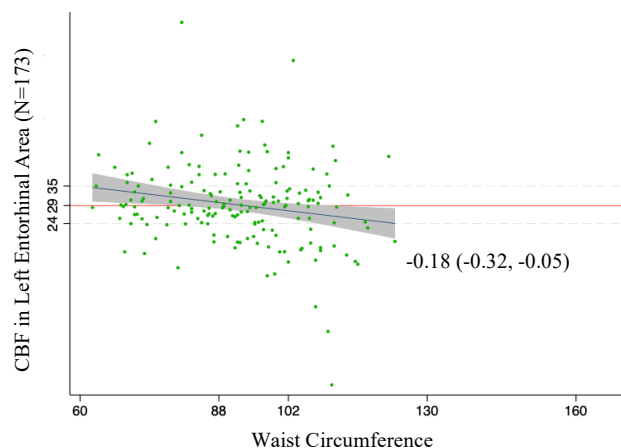


Figure 50 Scatterplot of (Mean) CBF (mm³/100gms/min) in left entorhinal area (N=173); the univariate SEM line suggests a negative association

Several studies have evaluated the effects of adiposity on brain blood flow. Willeumier *et al.* (2011) investigated the effects of high BMI on regional CBF using single photon emission CT imaging in healthy subjects. As part of a healthy brain study (HBS) conducted at the Amen Clinics, thirty-six adults recruited from the community between January 2003 and July 2009. The study showed that higher BMI was associated with lower CBF in the prefrontal cortex, which is the area of the brain responsible for executive function [236]. This finding is supported by a study conducted by Volkow *et al.* (2009) which reported diminished metabolic activity in the prefrontal cortex with high BMI [237]. The study measured the regional brain glucose metabolism to assess the relationship between BMI and brain metabolism in 21 healthy controls, with BMI from 19 to 37 during baseline and cognitive stimulation. A significant negative correlation between BMI and metabolic activity in prefrontal cortex and cingulate gyrus in healthy individuals was found. In addition, Kullmann *et al.* (2015) studied 25 healthy lean participants and 23 overweight or obese participants, and the study revealed significantly lower

CBF in both the prefrontal cortex and the hypothalamus (area of the brain that produces hormones that govern body weight homeostasis) after intranasal insulin administration compared with placebo [238]. CBF measures in the present middle-aged cohort are negatively associated with WC, respectively, and the association is in line with literature that detrimental effects of adiposity on brain blood flow, as shown as an example in Figure 50.

In this chapter, I conducted the statistical analysis using SEMs, examining physiological associations regarding cognition at Year 30. Many significant physiological markers were identified; however, the investigation needs to be expanded to include multiple time points in terms of the validity of the findings. This will be examined in the next chapter.

Chapter IV. Degeneration of cognition at Year 25 and Year 30

The cross-sectional analyses of the total path diagrams at Year 30 were performed in the previous chapter (Objectives 1 to 3). In this chapter, I investigate the associations among cardio-metabolic risk factors and the brain MRI parameters regarding cognition repeatedly measured at Year 25 and Year 30 using the Bayesian multilevel models (Objective 4). In the present chapter, I examined brain MRI parameters only; urinary metabolites were not available at Year 25. Figure 51 illustrates how this analysis fits into the overall study.

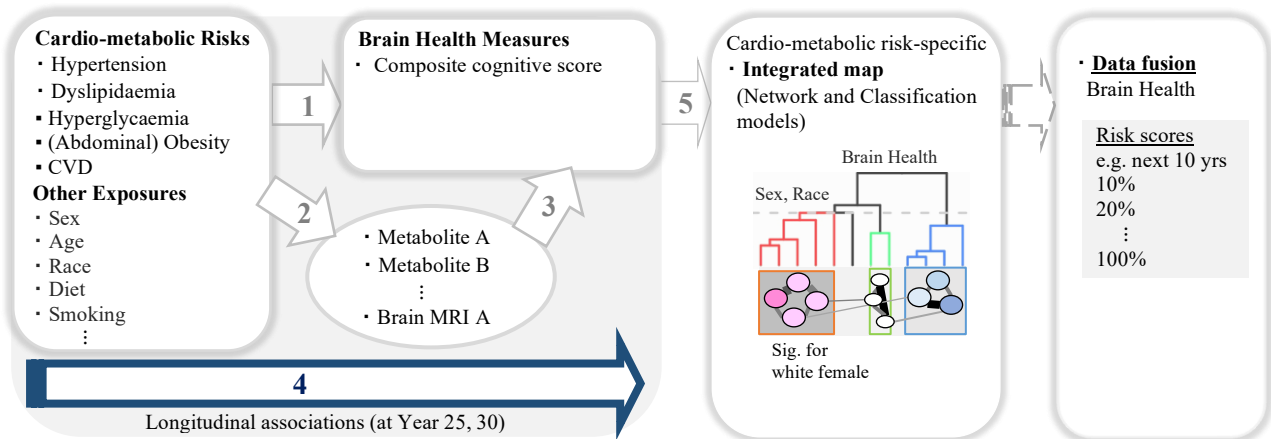


Figure 51 Schematic of the study highlighting the work focusses on Objective 4: a longitudinal analysis examining degeneration measured at Year 25 and Year 30

Bayesian multilevel models of repeatedly measured cognition were fitted with random intercepts for the participant IDs, accounting for the between-individual variability. The Bayesian approach estimates the probability that a theory is true given the observed data. It uses probability to represent all uncertainty within the model, both the uncertainty regarding the inputs and outputs of the model [239, 240]. Bayesian models are advantageous in terms of the measurement errors such as mis-measured continuous exposures under a classical measurement error model [241]. Bayesian models are especially useful in the case of multiplicity as the method avoids some of the conceptual and practical difficulties involved with the control of type I errors [242]. The hypothesis of the analyses in this chapter does not involve with multiplicity. Each Bayesian model tests only one hypothesis, i.e. whether a single cardio-metabolic factor or brain MRI measurement is significant regarding the composite cognitive score.

To investigate the degeneration of cognition over time (as illustrated in Figure 52), firstly I investigated which cardio-metabolic factor (X) is significant regarding cognition (Y) measured at Year 25 and Year 30. Secondly, I investigated which cardio-metabolic factor (X) is significant regarding a brain MRI parameter (M) measured at Year 25 and Year 30. Finally, I investigated which brain MRI parameter (M) with each cardio-metabolic risk factor (X) is significant regarding cognition (Y) measured at Year 25 and Year 30 adjusting for significant exposures.

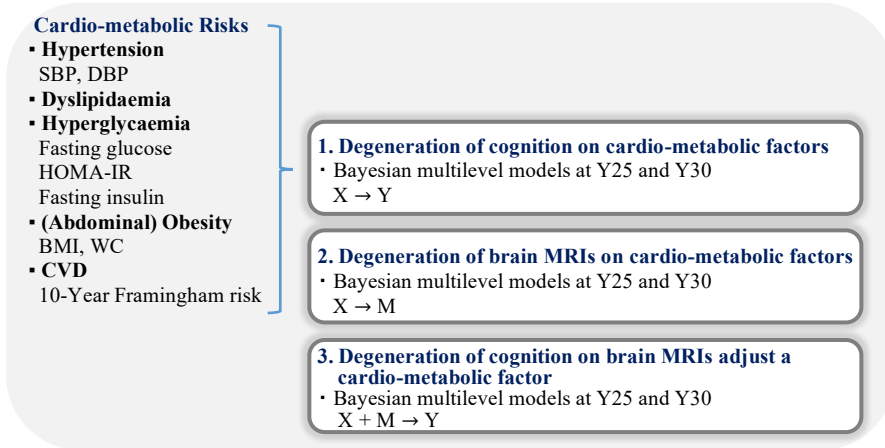


Figure 52 Workflow used in the present chapter: degeneration of cognition analysis at Year 25 and Year 30

1. Degeneration of cognition with a cardio-metabolic risk factor

In the present section, I investigate which cardio-metabolic factor is significant regarding the composite cognitive outcome measured at Years 25 and 30. The Bayesian two-level models of cognition were fitted on each cardio-vascular risk factor measured at Year 25 to Year 30, accounting for the variability between individuals. The model considering an unstructured within-subject covariance was applied after inspecting the within-subject covariance matrix. The cardio-metabolic risk factor-by-year interaction was also examined, not limited to SEM use. For the repeated cognitive measures, the composite cognitive scores which were comprised of the z-scores of RAVLT, DSST and Stroop tests measured at Year 25 and Year 30 were used (as reported in Table 7).

The Bayesian approach provided, for each parameter, a sample of 10,000 Markov chain Monte Carlo (MCMC) estimates from the simulated posterior distribution of the parameters. Normal prior distribution with mean 0 and a large variance (i.e. small precision) for the unknown beta-parameters were used as a common choice. These default priors for model parameters were used for the Bayesian multilevel regression (Table 30). The model structure and summary of the Bayesian multilevel regression for the composite condition and a cardio-metabolic risk, **SBP**, are as follows:

Participant ID
 $\{U_0\}$: *random intercepts*

Likelihood:
Composite cognition $\sim N(xb_{Composite\ cognition}, \{e.Composite\ cognition: \sigma^2\})$

Priors:
 $\{Composite\ cognition: SBP, constant\} \sim N(0, 10000)$ (1)
 $\{U_0\} \sim N(0, U_0: \sigma^2)$ (1)
 $\{e.Composite\ cognition: \sigma^2\} \sim igamma(0.01, 0.01)$

Hyperpriors:

$\{U_0: \sigma^2\}$

$\sim \text{igamma}(0.01, 0.01)$

(1) Parameters are elements of the linear form $xb_{\text{Composite cognition}}$

The parameter U_0 represents random intercepts in the model, that regression coefficients $\{\text{Composite cognition: SBP}\}$ and $\{\text{Composite cognition: constant}\}$ are assigned default normal priors with zero means and variances of 10,000. The variance component for participant ID, $\{U_0: \sigma^2\}$, is assigned the default inverse-gamma prior with 0.01 for both the shape and scale parameters. Gibbs sampling is used for regression coefficients and variance components to calculate Bayesian marginal posterior and predictive densities. After specifying a uniform on (-50, 50) prior for the regression coefficients, the same lists of the cardio-metabolic risk factors were significant regarding cognition. SBP was significant (i.e. the equal-tailed 95% CrIs do not include the null value, i.e. median -0.009; 95% CrIs: -0.036, -0.002, N=1,168 with 605 participants) regarding the composite cognitive score measured at Years 25 and 30 (Table 30).

Table 30 The Bayesian two-level model of composite cognition on **SBP** measured at Year 25 and Year 30 (N=1,168)

Bayesian multilevel regression		MCMC iterations = 12,500					
Metropolis-Hastings and Gibbs sampling		Burn-in = 2,500					
		MCMC sample size = 10,000					
Group variable: Participant ID		Number of groups = 605					
		Number of observations = 1,168					
		Acceptance rate = .9312					
		Efficiency: min = .001561					
		Avg = .001795					
Log marginal-likelihood		max = .001995					
		Mean	SD	MCSE	Median	Equal-tailed [95% Cred. Interval]	
composite Cognition	SBP	-0.011	0.008	0.002	-0.009	-0.036	-0.002
	constant	1.23	0.90	0.20	1.05	0.19	4.30
Participant ID	$U_0: \sigma^2$	4.24	1.16	0.29	4.52	0.02	5.17
e.composite Cognition	σ^2	1.22	1.04	0.26	0.94	0.83	5.20

* MCMC: Markov chain Monte Carlo

* MCSE: Monte-Carlo Standard Error

* Observation per group: min = 1, avg = 1.9, max = 2

The **DBP** was significant regarding cognition (i.e. median (95% CrIs): -0.017 (-0.055, -0.006), N=1,168). **None of dyslipidemia** (HDL-C, LDL-C and triglyceride) was significant regarding cognition (data not tabulated).

One of the significant hyperglycaemia factors, **fasting insulin**, is presented in the Table 31. **Fasting glucose** was not significant regarding the composite cognitive outcome (i.e. median (95% CrIs): -0.002 (-0.006, 0.002), N=1,162 with 604 participants).

Table 31 The Bayesian two-level model of composite cognition on **fasting insulin** measured at Year 25 and Year 30 (N=1,154)

Bayesian multilevel regression		MCMC iterations = 12,500					
Metropolis-Hastings and Gibbs sampling		Burn-in = 2,500					
Group variable: Participant ID		MCMC sample size = 10,000					
		Number of groups = 603					
		Number of observations = 1,154					
		Acceptance rate = .7997					
		Efficiency: min = .01774					
		avg = .09797					
Log marginal-likelihood		max = .2668					
		Mean	SD	MCSE	Median	Equal-tailed [95% Cred. Interval]	
composite Cognition	Fasting insulin	-0.023	0.007	0.000	-0.024	-0.036	-0.010
	constant	0.23	0.12	0.01	0.23	-0.01	0.45
Participant ID	$U_0: \sigma^2$	4.55	0.29	0.01	4.54	4.01	5.16
e.composite Cognition	σ^2	0.90	0.05	0.00	0.90	0.80	1.01

* MCMC: Markov chain Monte Carlo

* MCSE: Monte-Carlo Standard Error

* Observation per group: min = 1, avg = 1.9, max = 2

One of the (abdominal) obesity risk factors, **WC**, is presented in the Table 32. **BMI** was, moreover, significant regarding cognition (i.e. median (95% CrIs): -0.056 (-0.090, -0.031), N=1,167 with 605 participants).

Table 32 The Bayesian two-level model of composite cognition on WC measured at Year 25 and Year 30 (N=1,167)

Bayesian multilevel regression		MCMC iterations = 12,500					
Metropolis-Hastings and Gibbs sampling		Burn-in = 2,500					
Group variable: Participant ID		MCMC sample size = 10,000					
		Number of groups = 605					
		Number of observations = 1,167					
		Acceptance rate = .9283					
		Efficiency: min = .001315					
		avg = .001791					
Log marginal-likelihood		max = .002244					
		Mean	SD	MCSE	Median	Equal-tailed [95% Cred. Interval]	
composite Cognition	WC	-0.025	0.007	0.001	-0.024	-0.043	-0.015
	constant	2.36	0.64	0.14	2.25	1.37	3.98
Participant ID	$U_0: \sigma^2$	3.95	1.33	0.36	4.33	0.01	4.96
e.composite Cognition	σ^2	1.35	1.24	0.34	0.94	0.84	5.30

* MCMC: Markov chain Monte Carlo

* MCSE: Monte-Carlo Standard Error

* Observation per group: min = 1, avg = 1.9, max = 2

The reported estimates of posterior means and posterior standard deviations for the Bayesian model parameters were compared to the corresponding maximum-likelihood estimates and standard errors reported by the multilevel mixed-effects linear regressions. The significant cardio-metabolic risk factors from the Bayesian models (i.e. SBP, DBP, fasting insulin, WC and BMI) were also significant, respectively, regarding the composite cognition in the multilevel mixed-effects linear regressions. The 95% CrIs for SBP were slightly

wider (-0.036, -0.002, Table 30) than the 95% confidence intervals in the classical linear regression model (-0.015, -0.002, Table 33), corresponding to uncertainty.

Table 33 The multilevel mixed-effects linear regression of composite cognition on **SBP** measured at Year 25 and Year 30 (N=1,168)

Mixed-effects REML regression	Number of observations	=	1,168
Group variable: Participant ID	Number of groups	=	605
	Wald chi2(1)	=	6.44
Log restricted-likelihood = -2326.0733	Prob > chi2	=	0.0112

		Coef.	SD	z	P > z	[95% Conf. Interval]	
composite Cognition	SBP	-0.009	0.003	-2.540	0.011	-0.015	-0.002
	constant	0.99	0.41	2.40	0.02	0.18	1.81

Random-effects Parameters		Estimate	SE	[95% Conf. Interval]	
Participant ID: Identity	var (constant)	4.55	0.30	4.00	5.18
	var (Residual)	0.93	0.06	0.82	1.04

LR test vs linear model: chibar2(01) = 608.78, Prob >= chibar2 = 0.0000

Var (constant) and var (Residual) are the variance of the intercepts and the residual variance which correspond to the between-subject and within-subject variances, respectively.

In the Bayesian models which additionally examined the cardio-metabolic risk factor-by-year interaction (such as SBP-by-year), there was no significant year effects (i.e. Year 25 versus Year 30) regarding cognition.

2. Degeneration of the brain MRI parameters with a cardio-metabolic factor

In the present section, I investigate which cardio-metabolic factor is significant regarding each of the brain MRI parameter measured at Year 25 and Year 30. The brain MRI parameters were standardized (z-scores) for the analysis. The ranges of standardized 128 brain MRI parameters used for the analyses are -9.4 to 13 (except an outlier; abnormal right entorhinal area tissue volume) with the average range -2.6 and 5.6.

The Bayesian two-level model of each brain MRI parameter measure was fitted, respectively, on each cardiovascular risk factor measured at Year 25 and Year 30, accounting for the variability between individuals. Total ICV was adjusted in the models. The model structure and summary of the Bayesian multilevel regression for abnormal total brain tissue volume on SBP are as follows.

Participant ID
 $\{U_0\}$: *random intercepts*

Likelihood:

Abnormal Tissue Volume in Total Brain $\sim N(xb_{ATV \text{ in Total Brain}}, \{e.ATV \text{ in Total Brain}: \sigma^2\})$

Priors:

$$\begin{aligned} \{ATV \text{ in Total Brain: } SBP, \text{ Total Intracranial Volume, Constant} \} & \sim N(0, 10000) \quad (1) \\ \{U_0\} & \sim N(0, U_0: \sigma^2) \quad (1) \\ \{e. ATV \text{ in Total Brain: } \sigma^2\} & \sim \text{igamma}(0.01, 0.01) \end{aligned}$$

Hyperpriors:

$$\begin{aligned} \{U_0: \sigma^2\} & \sim \text{igamma}(0.01, 0.01) \\ (1) \text{ Parameters are elements of the linear form } xb_{ATV \text{ in Total Brain}} \end{aligned}$$

Table 34 The Bayesian two-level model of abnormal total brain tissue volume on **SBP** measured at Year 25 and Year 30 (N=469)

Bayesian multilevel regression	MCMC iterations = 12,500
Metropolis-Hastings and Gibbs sampling	Burn-in = 2,500
	MCMC sample size = 10,000
	Number of groups = 283
	Number of observations = 469
	Acceptance rate = .8026
	Efficiency: min = .05208
	avg = .0665
	max = .07389
Log marginal-likelihood	

		Mean	SD	MCSE	Median	Equal-tailed [95% Cred. Interval]	
Abnormal total brain tissue volume	SBP	0.009	0.003	0.000	0.009	0.003	0.015
	Total intracranial volume	0.071	0.054	0.002	0.072	-0.035	0.178
	Constant	-1.043	0.374	0.014	-1.037	-1.803	-0.313
Participant ID	$U_0: \sigma^2$	0.571	0.077	0.003	0.568	0.426	0.733
e. Abnormal total brain tissue volume	σ^2	0.419	0.043	0.002	0.417	0.342	0.510

* MCMC: Markov chain Monte Carlo

* MCSE: Monte-Carlo Standard Error

* Observation per group: min = 1, avg = 1.7, max = 2

SBP was significant regarding the abnormal total brain tissue volume measured at Year 25 and Year 30, adjusting for the total ICV (Table 34). Furthermore, each association of the brain MRI on SBP was examined. The median (95% CrIs) for which SBP is significant regarding the standardized brain MRI parameters in the Bayesian multilevel regressions are presented in Figure 53.

The cardio-metabolic risk factor SBP was **positively** associated with abnormal tissue volumes with the exception of the abnormal left hippocampus tissue volume (N of participants=283), and time to repeat in brain MRI parameters (TRM; time from the application of an excitation pulse to the application of the next pulse). SBP is **negatively** associated with CBF (overall median (95% CrIs): -0.009 (-0.010, -0.008)), FA (overall median (95% CrIs): -0.009 (-0.011, -0.007)), and the number of activated voxels within brain MRI parameters (overall median (95% CrIs): -0.007 (-0.010, -0.004)) (Figure 53). Overall median values by the brain MRI category were calculated for a fixed effect model using the method of Mantel and Haenszel [243]. The associations are reasonable as these are positive markers in brain health in the middle-aged cohort (as described in Chapter II, Sub-section 1-5).

In the models additionally including the SBP-by-year interaction, the brain MRI parameters which are significantly different at Years 25 and 30 are as follows: abnormal temporal (including temporal lobe WM right) and left hippocampus tissue volumes; CBF in deep WM GM, frontal, limbic, temporal, right and left MFC, right and left ACG, and temporal lobe WM right; FA in total brain and frontal.

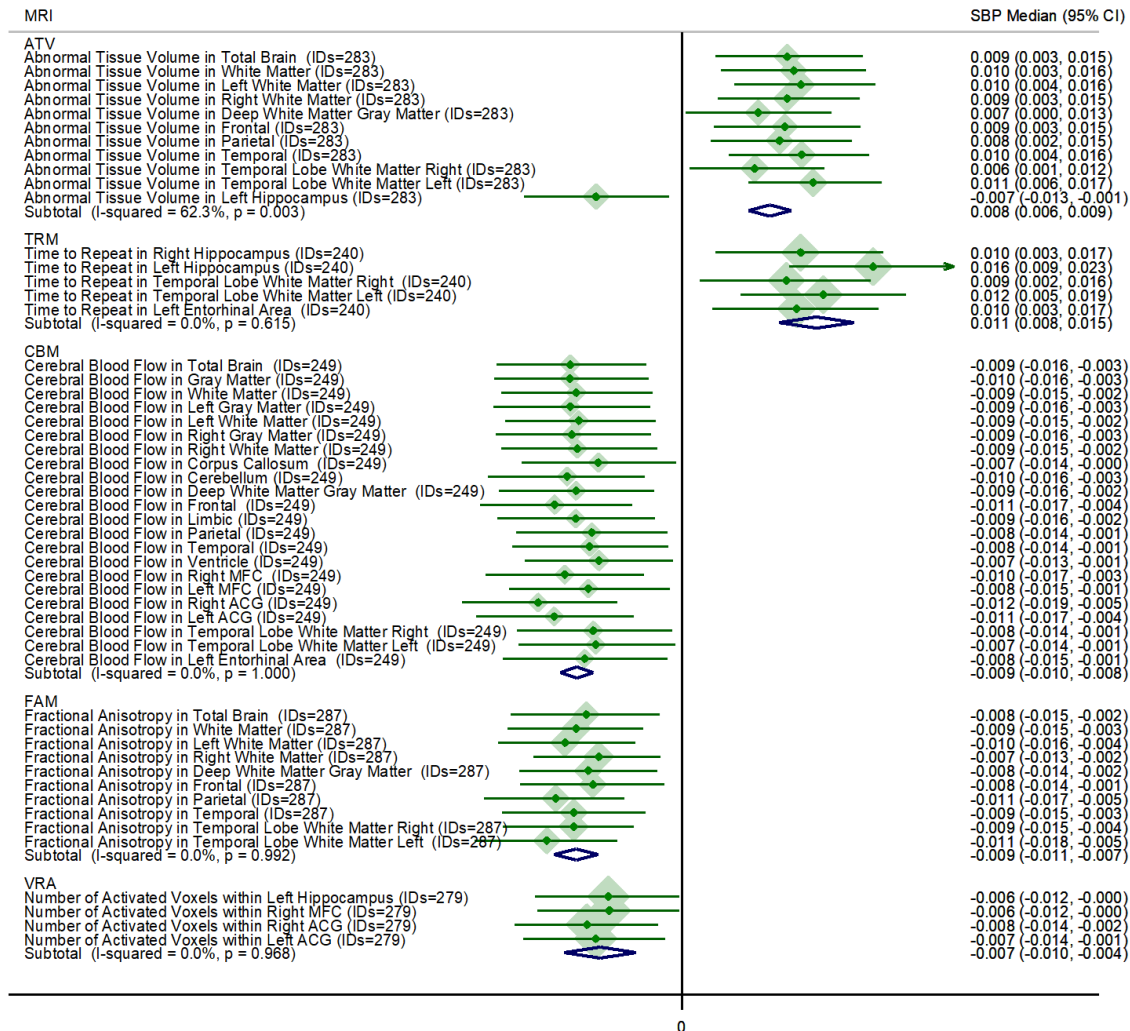


Figure 53 The standardized brain MRI measurements for which **SBP** is significant in each Bayesian multilevel regression; median (95% CrIs); total intracranial volume (ICV) was adjusted for in the models

*ATV: Abnormal tissue volume, TRM: Mean of time to repeat values in an ROI, CBM: Mean of cerebral blood flow values in an ROI, FAM: Mean of FA values in an ROI, VRA: Number of activated voxels within the ROI, IDs: Number of participants

The cardio-metabolic risk factor fasting insulin was positively associated with time to repeat in brain MRI parameters, and percent change of vascular reactivity in brain MRI parameters. Baron's (1999) statement about the association between worsening insulin resistance (which causes higher insulin and blood sugar levels) and increased vascular reactivity, supports this outcome [244]. Fasting insulin is **negatively** associated with brain volumes with an exception of total cerebrospinal fluid (N of participants=284) (overall median (95% CrIs): -0.004 (-0.005, -0.002)), CBF (overall median (95% CrIs): -0.018 (-0.021, -0.015)), FA (overall median (95% CrIs): -0.016 (-0.019, -0.013)), and the number of activated voxels within brain MRI parameters, (overall median (95% CrIs): -0.015 (-0.019, -0.011)) (Figure 54). Salamán (1991) stated that the evidence indicates that

blood-borne insulin can reach the cerebrospinal fluid and specific sites in the brain through the circumventricular organs via transport of insulin from the peripheral blood to the central nervous system (cerebrospinal fluid-brain) in species including humans [245]. This explains how fasting insulin is positively associated with cerebrospinal fluid in the present study.

In the models additionally including fasting insulin-by-year interactions, the brain MRI parameters which are significantly different at Years 25 and 30 are as follows: total volume of brain, cerebrospinal fluid, right GM, left MFC; CBF in total brain, (right) GM, right WM, frontal, limbic, parietal, temporal lobe WM right; all the measurements of FA (corrected measurement was used at Year 30); number of activated voxels within right and left MFC and ACG; time to repeat in the left hippocampus and temporal lobe WM right; percent change of vascular reactivity in temporal lobe WM right and left, left entorhinal area.

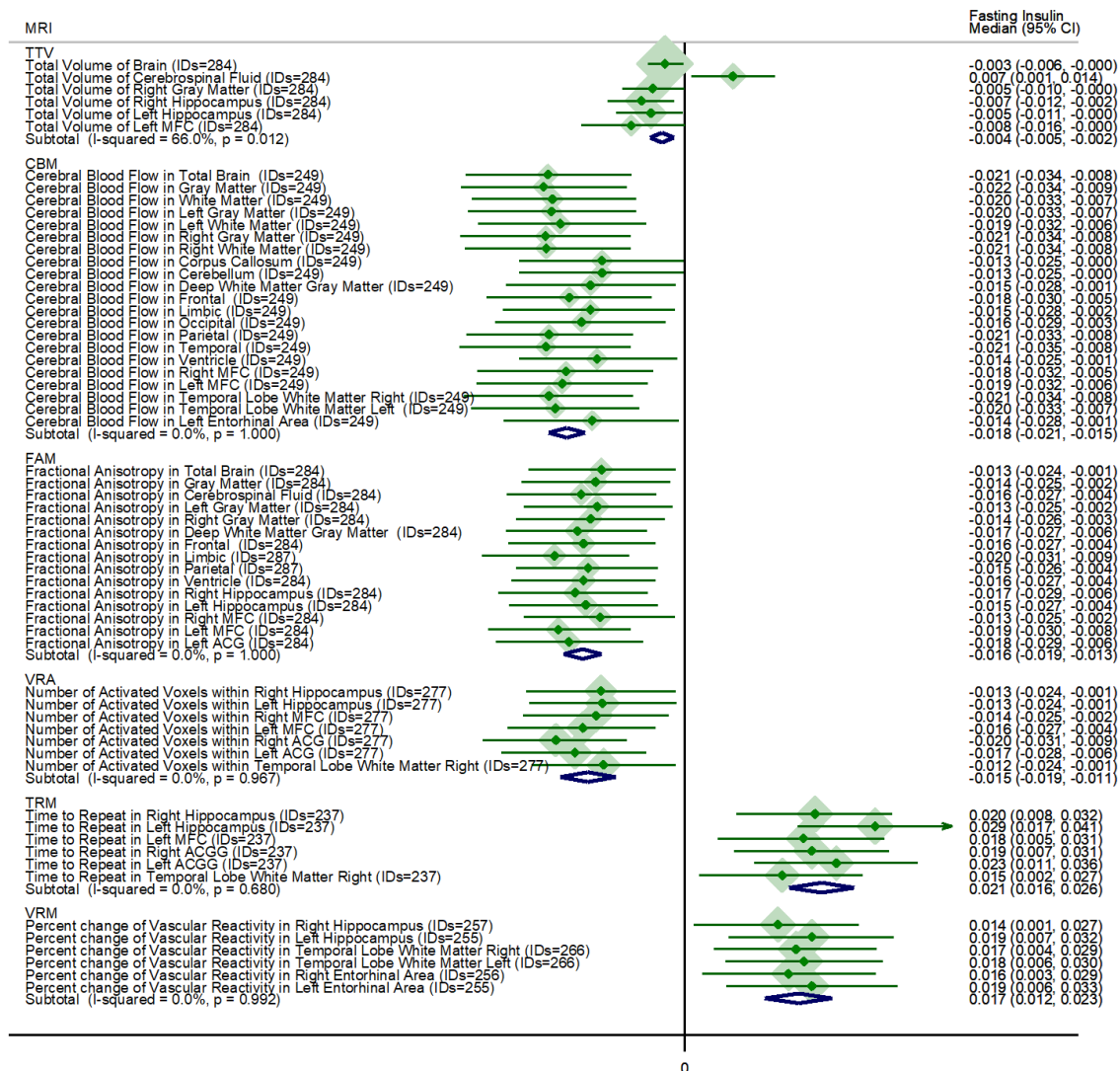


Figure 54 The standardized brain MRI measurements for which **fasting insulin** is significant in each Bayesian multilevel regression; median (95% CrIs); total intracranial volume (ICV) was adjusted in the models

*TTV: Total tissue volume, CBM: Mean of cerebral blood flow values in an ROI, FAM: Mean of FA values in an ROI, VRA: Number of activated voxels within the ROI, TRM: Mean of time to repeat values in an ROI, VRM: Mean percent change of vascular reactivity in the ROI, IDs: Number of participants

The cardio-metabolic risk factor WC is **positively** associated with abnormal tissue volume, time to repeat and percent change of vascular reactivity in brain MRI parameters. WC is **negatively** associated with CBF (overall median (95% CrIs): -0.014 (-0.015, -0.012)), FA (overall median (95% CrIs): -0.009 (-0.013, -0.006)), and the number of activated voxels within brain MRI parameters (overall median; 95% CrIs: -0.009; -0.013, -0.006) (Figure 55).

In the models additionally including the WC-by-year interactions, the brain MRI measurements which are significantly different at Years 25 and 30 are as follows: CBF in total brain, (left and right) GM and WM, deep WM GM, corpus callosum, frontal, limbic, parietal, temporal, (left and right) hippocampus, temporal lobe WM right, right entorhinal area.

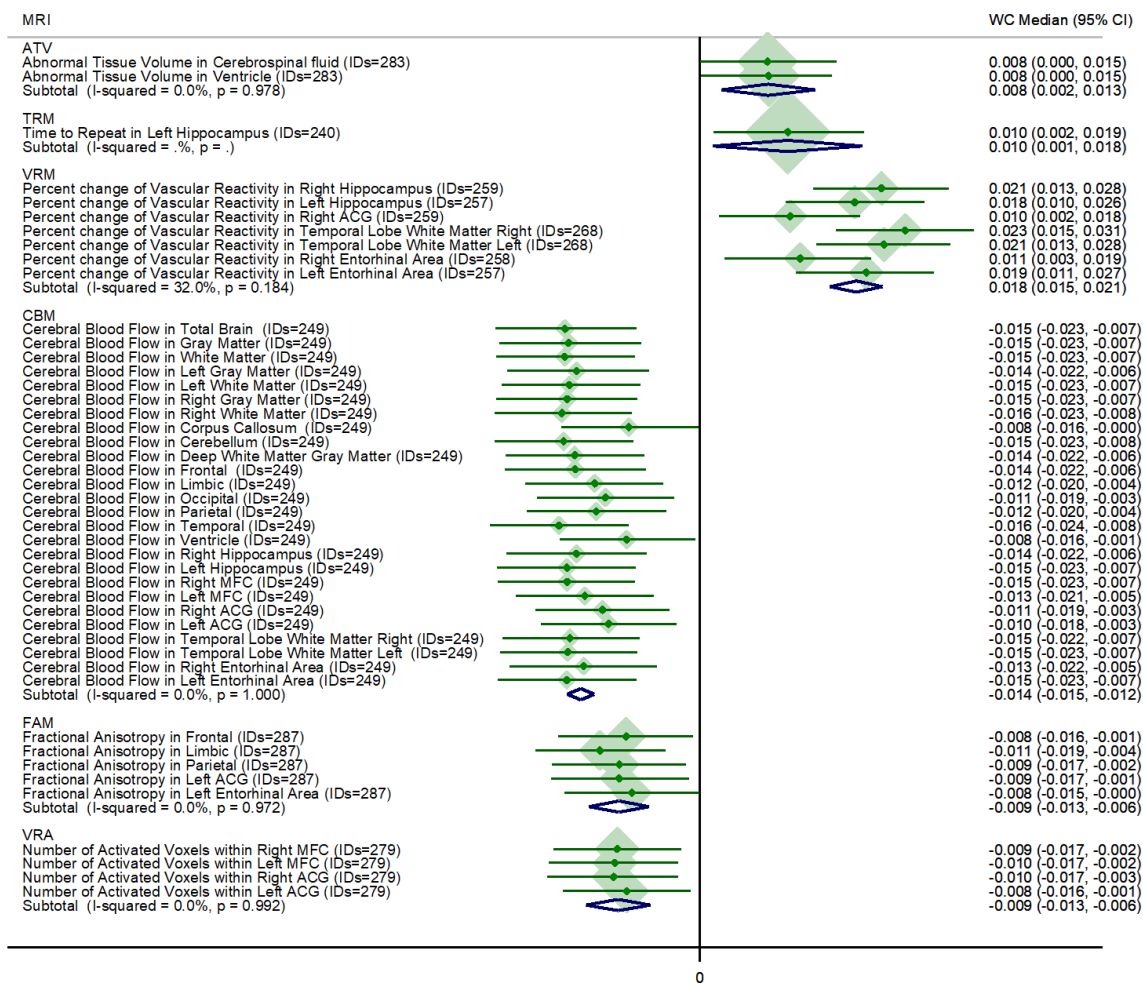


Figure 55 The standardized brain MRI measurements for which WC is significant in each Bayesian multilevel regression; median (95% CrIs); total intracranial volume (ICV) was adjusted in the models

* ATV: Abnormal tissue volume, TRM: Mean of time to repeat values in an ROI, VRM: Mean percent change of vascular reactivity in the ROI, CBM: Mean of cerebral blood flow values in an ROI, FAM: Mean of FA values in an ROI, VRA: Number of activated voxels within the ROI, IDs: Number of participants

The cardio-metabolic risk factor BMI is **positively** associated with the percent change of vascular reactivity and abnormal tissue volume in brain MRI parameters. Interestingly, BMI is **positively** associated with the brain volume parameters (except cerebrospinal fluid) in this middle-aged cohort (Figure 56).

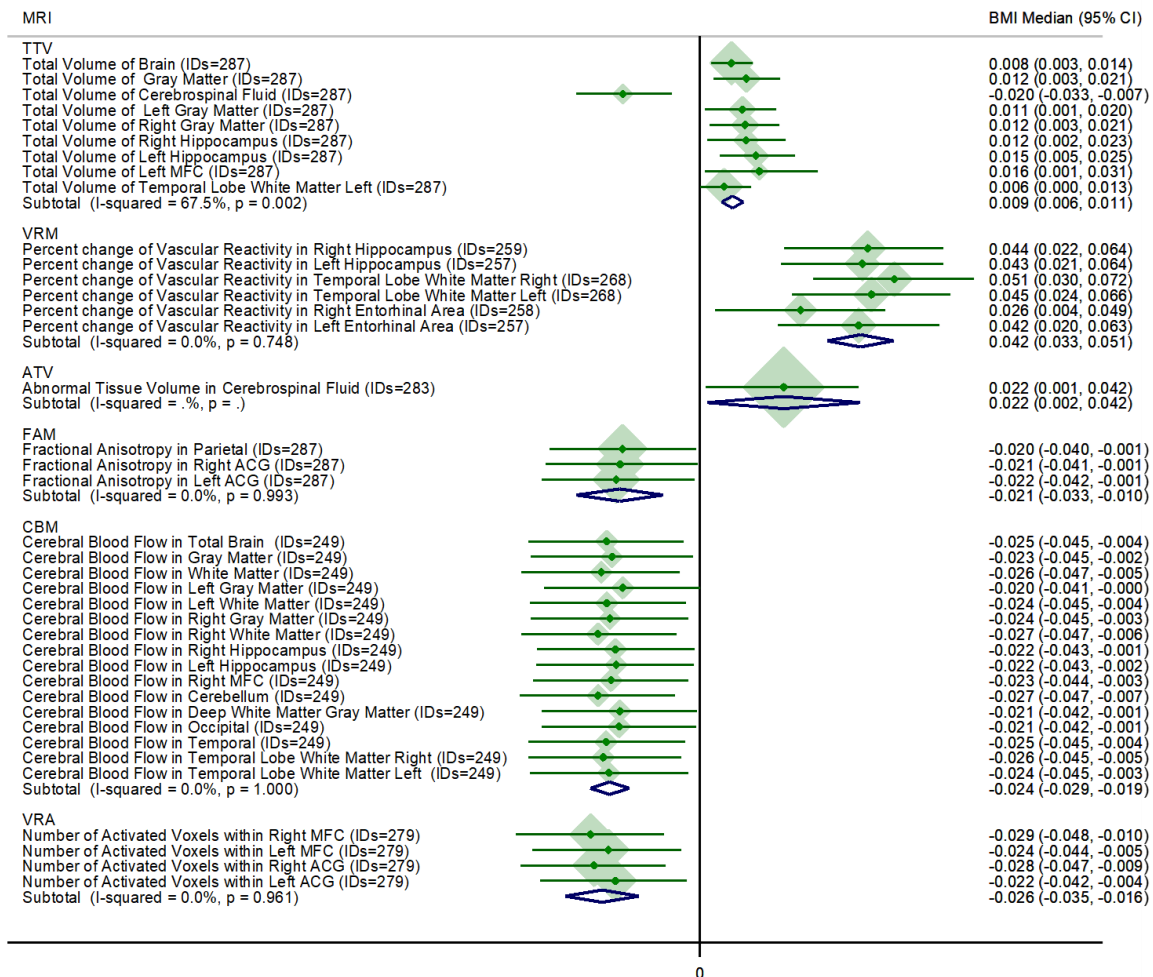


Figure 56 The standardized brain MRI measurements for which **BMI** is significant in each Bayesian multilevel regression; median (95% CrIs); total intracranial volume (ICV) was adjusted in the models

* TTV: Total tissue volume, VRM: Mean percent change of vascular reactivity in the ROI, ATV: Abnormal tissue volume, FAM: Mean of FA values in an ROI, CBM: Mean of cerebral blood flow values in an ROI, VRA: Number of activated voxels within the ROI, IDs: Number of participants

3. Degeneration of cognition with the brain MRI parameters adjusting for a cardio-metabolic factor

In the present section, after adjusting for significant exposures, I investigate which brain MRI parameters are significant regarding cognition with each cardio-metabolic risk factor (i.e. SBP, DBP, fasting insulin, WC, and BMI) measured at Year 25 and Year 30.

I examine the significant exposures adjusting for a cardio-metabolic risk factor in the models. The fully adjusted Bayesian multilevel regression of the composite cognitive outcome on the standardized brain MRI parameters with each cardio-metabolic risk factor was fitted with a random intercept for the individual ID.

Race was a confounding factor regarding cognition in the models. Total ICV was adjusted in each model and its effect was examined when it was not adjusted regarding cognition.

The Bayesian two-level model of the composite cognitive outcome on abnormal total brain tissue volume with SBP measured at Year 25 and Year 30 is presented in Table 35. Significant exposures (i.e. sex, education, physical activity, smoking, CES-D, HBP medication, family income, and urinary creatinine (mg/dL)) and the total ICV were adjusted. After additionally adjusting for the confounding factor (i.e. race), physical activity and urinary creatinine were no longer significant regarding cognition (i.e. modifiable factor by race).

The Multilevel model structure and the model summary are as follows:

Participant ID

$\{U_0\}$: random intercepts

Likelihood:

Composite cognition $\sim N(xb_{Composite\ cognition}, \{e.Composite\ cognition: \sigma^2\})$

Priors:

$\{Composite\ cognition: SBP\}$	$\sim N(0, 10000)$ (1)
$\{Composite\ cognition: Physical\ activity\}$	$\sim N(0, 10000)$ (1)
$\{Composite\ cognition: Smoking\}$	$\sim N(0, 10000)$ (1)
$\{Composite\ cognition: ICES - D\}$	$\sim N(0, 10000)$ (1)
$\{Composite\ cognition: HBP\ medication\}$	$\sim N(0, 10000)$ (1)
$\{Composite\ cognition: Family\ income\}$	$\sim N(0, 10000)$ (1)
$\{Composite\ cognition: Urinary\ Creatinine\}$	$\sim N(0, 10000)$ (1)
$\{Composite\ cognition: Sex\}$	$\sim N(0, 10000)$ (1)
$\{Composite\ cognition: Education\}$	$\sim N(0, 10000)$ (1)
$\{Composite\ cognition: Intracranial\ volume\}$	$\sim N(0, 10000)$ (1)
$\{Composite\ cognition: Abnormal\ tissue\ volume\}$	$\sim N(0, 10000)$ (1)
$\{Composite\ cognition: Constant\}$	$\sim N(0, 10000)$ (1)
$\{U_0\}$	$\sim N(0, U_0: \sigma^2)$ (1)
$\{e.Composite\ cognition: \sigma^2\}$	$\sim igamma(0.01, 0.01)$

Hyperpriors:

$\{U_0: \sigma^2\} \sim igamma(0.01, 0.01)$

(1) Parameters are elements of the linear form $xb_{Composite\ cognition}$

Table 35 The Bayesian two-level model of composite cognition on abnormal total brain tissue volume with SBP measured at Year 25 and Year 30 (N=437)

Bayesian multilevel regression
Metropolis-Hastings and Gibbs sampling

MCMC iterations = 12,500
Burn-in = 2,500
MCMC sample size = 10,000
Number of groups = 270
Number of observations = 437
Acceptance rate = .8047
Efficiency: min = .01675
Avg = .04724
max = .1356

Log marginal-likelihood

		Mean	SD	MCSE	Median	Equal-tailed [95% Cred. Interval]	
composite	SBP	-0.002	0.005	0.0003	-0.002	-0.01	0.01
Cognition	Physical activity	0.0005	0.0003	0.00001	0.0005	0.000002	0.001
	Smoking	-0.10	0.12	0.01	-0.10	-0.33	0.13
	CES-D	-0.02	0.01	0.0005	-0.02	-0.04	-0.003
	HBP medication	-0.16	0.07	0.002	-0.16	-0.30	-0.02
	Family income	0.08	0.03	0.002	0.08	0.03	0.14
	Urinary creatinine	-0.002	0.001	0.00004	-0.002	-0.004	-0.001
	Sex	2.17	0.30	0.02	2.16	1.61	2.77
	Education	0.24	0.04	0.003	0.24	0.16	0.32
	Total intracranial volume	0.70	0.14	0.01	0.70	0.43	0.97
	Abnormal total brain tissue volume	-0.22	0.08	0.004	-0.22	-0.38	-0.07
	Constant	-4.77	1.01	0.06	-4.75	-6.80	-2.83
Participant ID	$U_0: \sigma^2$	2.20	0.23	0.01	2.18	1.77	2.69
e.Abnormal total brain tissue volume	σ^2	0.54	0.06	0.00	0.53	0.43	0.68

* MCMC: Markov chain Monte Carlo

* MCSE: Monte-Carlo Standard Error

* Observation per group: min = 1, avg = 1.6, max = 2

The significant brain MRI parameters regarding the composite cognitive outcome (when SBP was fitted along this the significant exposures in each Bayesian multilevel regression) were presented in Figure 57. The brain MRI parameters measured within the same category are strongly correlated (e.g. abnormal tissue volume: ≥ 0.64 Spearman's ρ , N=469 at Y25 and Y30; number of activated voxels volume: ≥ 0.70 Spearman's ρ , N=454 at Y25 and Y30). Therefore, the overall estimated median (95% CrIs) was calculated by the significant brain MRI parameters category using a fixed effect model with the method of Mantel and Haenszel.

The brain MRI parameters measuring abnormal tissue volumes were negatively associated with cognition; not adjusted and adjusted for total ICV overall median (95% CrIs): -0.15 (-0.23, -0.07) and -0.17 (-0.24, -0.11), respectively. Mean of FA in deep WM GM, and the brain MRI parameters measuring number of activated voxels with right MFC and ACG were positively associated with cognition. These associations are reasonable as they are positive markers of brain health in the middle-aged cohort. Abnormal total brain and WM tissue volumes were additionally significant after adjusting for total intracranial volume (ICV). The number of activated voxels within left hippocampus and within left ACG were no longer significant after additionally adjusting for the confounding factor, race (i.e. modifiable factor by race).

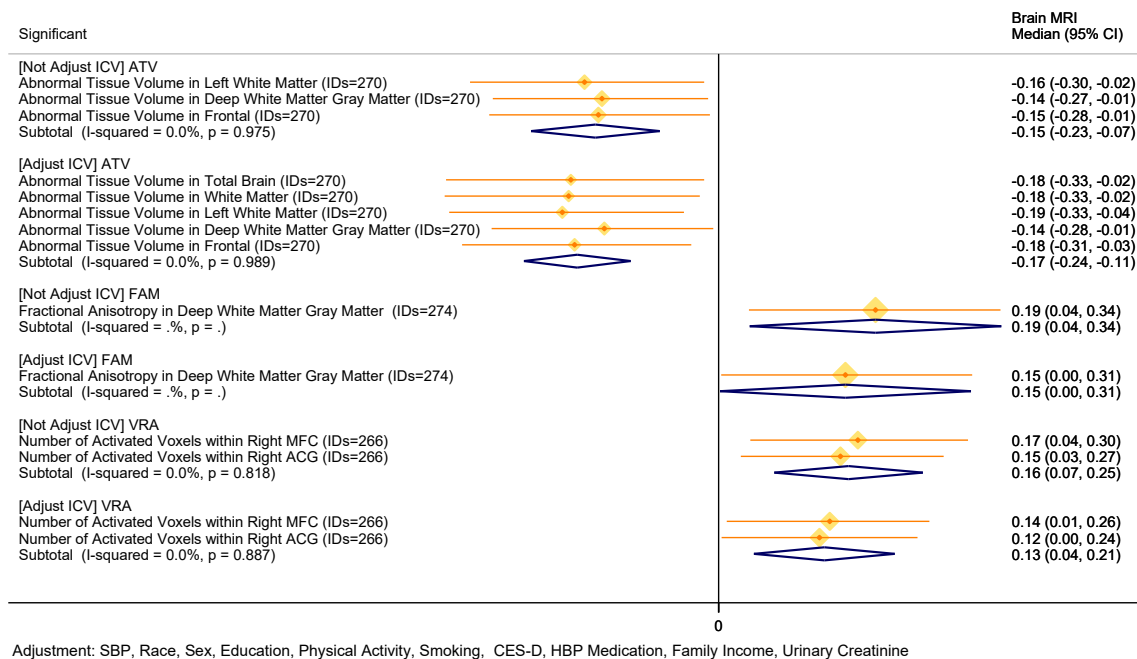


Figure 57 The significant standardized brain MRI measurements with **SBP** regarding composite cognition in each Bayesian multilevel regression; median (95% CrIs); Significant exposures adjusted for race, sex, education, physical activity, smoking, CES–D, HBP medication, family income, and urinary creatinine (mg/dL)
 * ICV: Intracranial volume, ATV: Abnormal tissue volume, FAM: Mean of FA values in an ROI, VRA: Number of activated voxels within the ROI, MFC: Medial frontal cortex, ACG: Anterior cingulate gyrus, IDs: Number of participants

The brain MRI parameters positively associated with cognition when they fitted with fasting insulin are as follows: total volume of brain, right GM, and the right hippocampus (which is no longer significant after adjusting for ICV); mean of FA in deep WM GM and right MFC (which is no longer significant after adjusting for ICV); percent change of vascular reactivity in temporal lobe WM right and left (ICV adjusted overall median (95% CrIs): 0.15 (0.06, 0.24)); number of activated voxels within right and left MFC and right ACG and temporal lobe WM right (ICV adjusted overall median (95% CrIs): 0.15 (0.10, 0.20)). Among the significant brain MRI parameters, the number of activated voxels within the right and left hippocampus were no longer significant after adjusting for ICV (i.e. modifiable factor by ICV) (Figure 58).

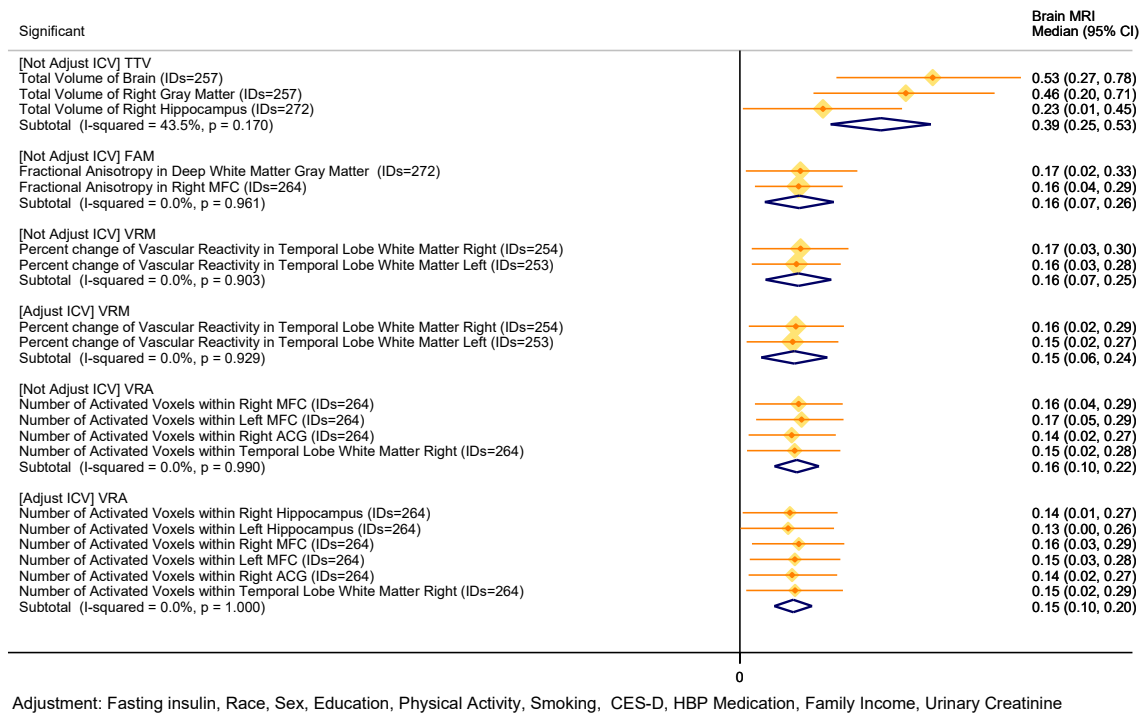


Figure 58 The significant standardized brain MRI measurements with **fasting insulin** regarding composite cognition in each Bayesian multilevel regression; median (95% CrIs); Significant exposures adjusted for are race, sex, education, physical activity, smoking, CES–D, HBP medication, family income, and urinary creatinine (mg/dL)

* ICV: Intracranial volume, TTV: Total tissue volume, FAM: Mean of FA values in an ROI, VRM: Mean percent change of vascular reactivity in the ROI, VRA: Number of activated voxels within the ROI, MFC: Medial frontal cortex, ACG: Anterior cingulate gyrus, IDs: Number of participants

The measured brain MRI parameters positively associated with cognition when they fitted with WC are as follows: percent change of temporal lobe WM right and left (ICV adjusted overall median (95% CrIs): 0.16 (0.07, 0.25)); number of activated voxels within right and left MFC and right ACG (ICV adjusted overall median (95% CrIs): 0.14 (0.06, 0.21)). Among the significant brain MRI parameters, percent change of vascular reactivity in the right hippocampus was no longer significant after additionally adjusting for the confounding factor, race (i.e. modifiable factor by race) (Figure 59).

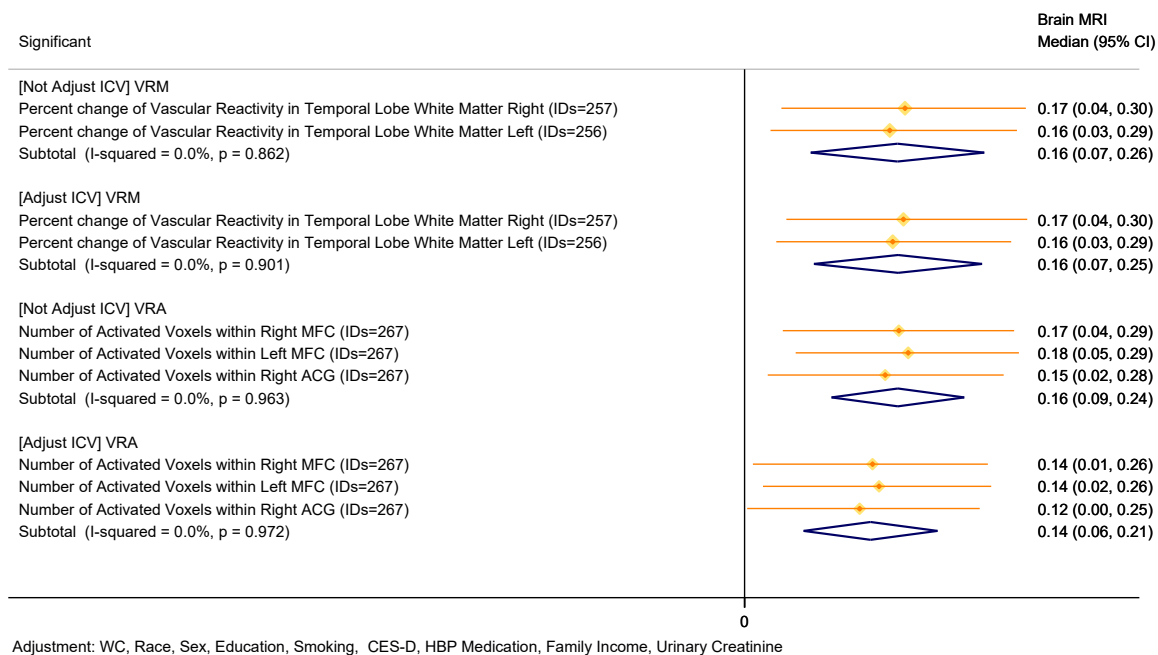


Figure 59 The significant standardized brain MRI measurements with **WC** regarding composite cognition in each Bayesian multilevel regression; median (95% CrIs); Significant exposures adjusted for are race, sex, education, smoking, CES–D, HBP medication, family income, and urinary creatinine (mg/dL)

* ICV: Intracranial volume, VRM: Mean percent change of vascular reactivity in the ROI, VRA: Number of activated voxels within the ROI, MFC: Medial frontal cortex, ACG: Anterior cingulate gyrus, IDs: Number of participants

The measured brain MRI parameters positively associated with cognition when they fitted with BMI are as follows: total volume of brain, total volume of (left) GM, cerebrospinal fluid (ICV adjusted overall median (95% CrIs): 0.10 (-0.06, 0.27)), (right) GM, (right and left) hippocampus and temporal lobe WM left; percent change of vascular reactivity in temporal lobe WM right and left (ICV adjusted overall median (95% CrIs): 0.16 (0.08, 0.25)); number of activated voxels within left MFC (ICV adjusted median (95% CrIs): 0.12 (0.01, 0.24)). In addition, the number of activated voxels within right MFC and right ACG were no longer significant after additionally adjusting for the confounding factor, race (i.e. modifiable factor by race) (Figure 60).

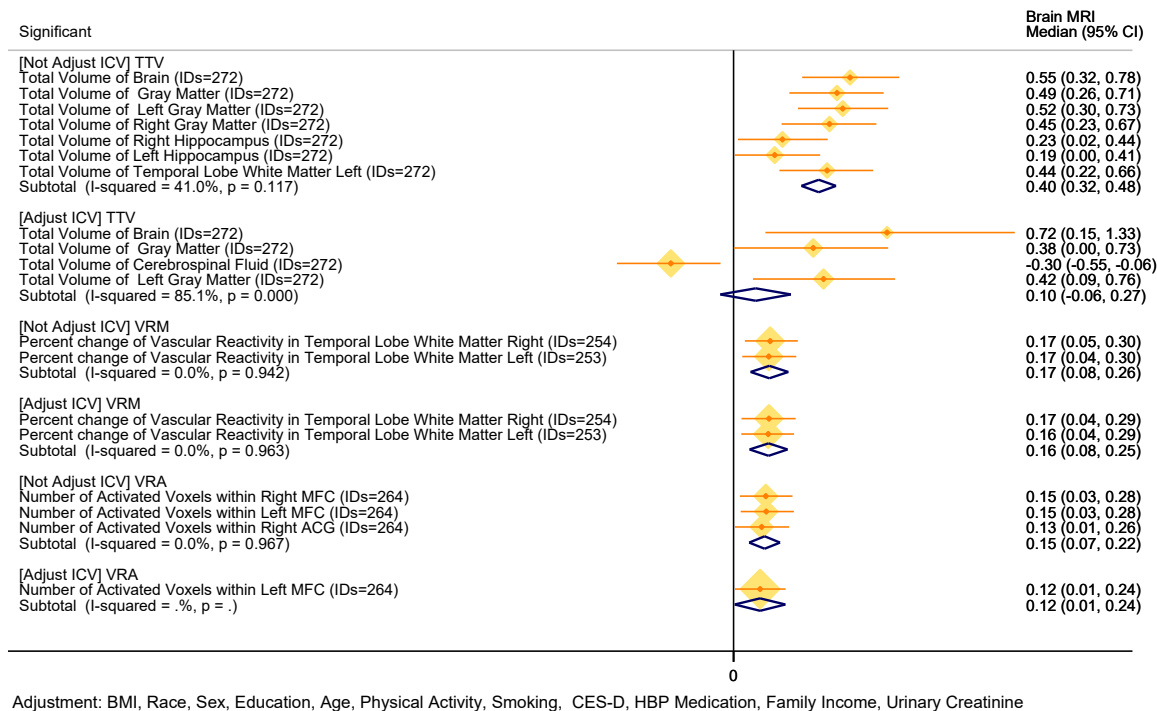


Figure 60 The significant standardized brain MRI measurements with **BMI** regarding composite cognition with BMI in each Bayesian multilevel regression; the brain MRI median (95% CrIs); Significant exposures adjusted for are race, sex, education, age, physical activity, smoking, CES–D, HBP medication, family income, and urinary creatinine (mg/dL)

* TTV: Total tissue volume, VRM: Mean percent change of vascular reactivity in the ROI, VRA: Number of activated voxels within the ROI, MFC: Medial frontal cortex, ACG: Anterior cingulate gyrus, IDs: Number of participants

The analysis of the standardized brain MRI parameters with each cardio-metabolic risk factor regarding cognition was conducted using Bayesian multilevel regression, adjusting for significant exposures. A number of brain MRI parameters were positively associated regarding cognition (with small effect sizes) at Years 25 and 30 when fasting insulin, WC and BMI were included, respectively, in the model. Higher annual rates of memory decline were associated with greater volume loss in 14 brain volumes of interest primarily within the temporal and occipital lobes in a subset of Baltimore Longitudinal Study of Aging (BLSA) participants (N=836, aged ≥ 55 years), where evaluations conducted from 1994–2018 [246]. Another longitudinal study examined insulin resistance among late middle-aged and cognitively healthy participants (N=372, aged 57.7 ± 7.5 years, and a consecutive subset of 121 participants with 4 years follow-up), confirmed that medial temporal lobe atrophy itself corresponded to cognitive deficits in the RAVLT. Temporal lobe atrophy (that was predicted by higher insulin resistance) significantly mediated worse RAVLT encoding performance [247]. Total temporal lobe WM right and left volume changes (Y30 - Y25) in the present study are also positively associated with composite cognition change (Y30 - Y25), respectively, as shown in Figure 61.

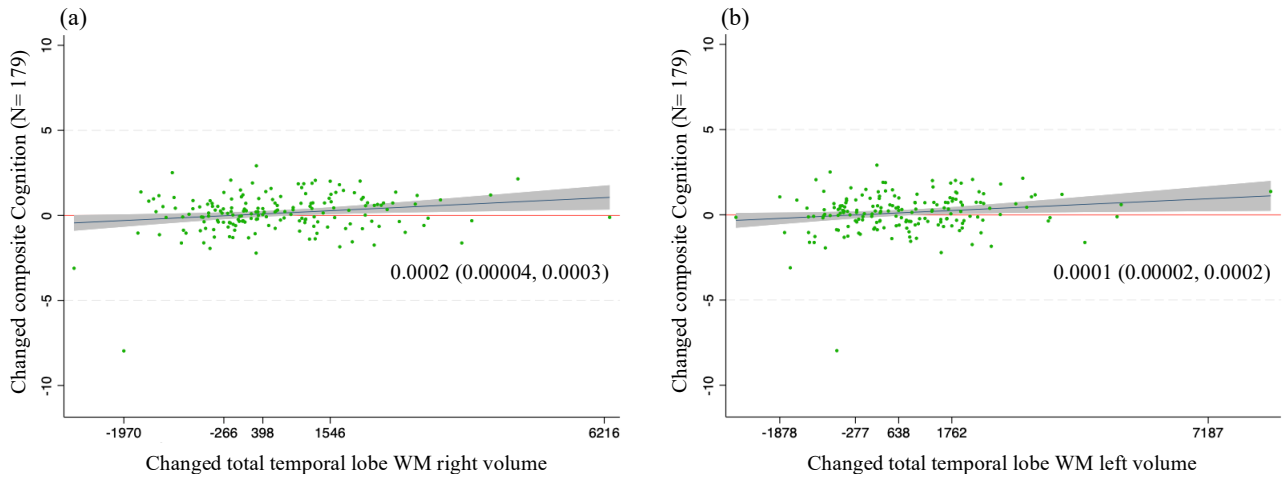


Figure 61 Scatterplot of changed composite cognition (Y30 – Y25) with changed total temporal lobe WM (a) right and (b) left volumes (cc^3) (N=179); the univariate SEM line suggests a positive association, respectively

In the next chapter, all the examined information including different types of significant other exposures (i.e. clinical, socio-demographic, lifestyle risk factors) and the physiological markers were assimilated to provide an integrated map (i.e. profile of the markers) of cardio-metabolic risk-specific brain health (i.e. cognition or dementia). The network analyses and the classification models with dimensionality-reduction will be employed to investigate cardio-metabolic risk-specific phenotypes or profiles of the significant markers. The findings will be validated with external data, or other studies from the literature in the following investigations.

Chapter V. Data fusion for cognition – Integrated map of risk-specific markers

In the previous chapters, the total path diagrams with the NMR- and LC–MS-based urinary metabolites, and the brain MRI parameters were constructed to examine mediators regarding adjusted composite cognition measured at Year 30 (Objectives 1 to 3). Next, the degeneration of composite cognitive function on a brain MRI parameter was examined per cardio-metabolic factor (e.g. SBP, fasting insulin, WC/BMI) measured at Year 25 (2010–11) and Year 30 (2015–16) (Objective 4).

In the present chapter (Objective 5), the cardio-metabolic risk-specific integrated map (i.e. profile of the markers) regarding the cognitive outcome will be visualised with significant epidemiological exposures, metabolites and brain MRI parameters identified from Objectives 1 to 4.

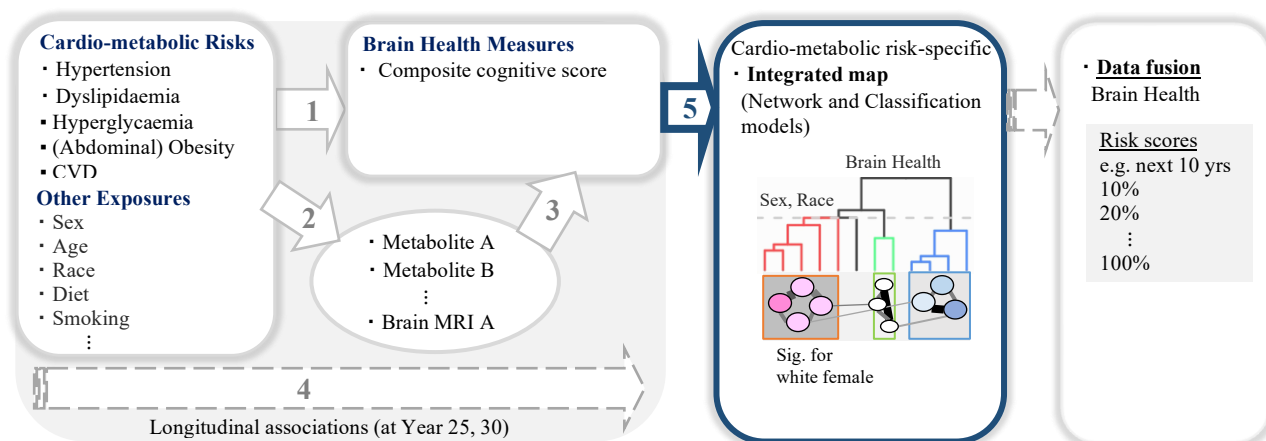


Figure 62 Schematic of the study highlighting the work focusses on Objective 5: a cardio-metabolic risk-specific integrated map regarding cognition

1. Network analysis measured at Year 30

The network analyses of metabolites and the brain MRI parameters from the SEMs at Year 30 (Table 21, Table 23, Table 25, Appendix Table 10 and Table 29, respectively) are presented in Figure 63 to Figure 66. The models regressed the composite cognitive score on cardio-metabolic risk factors (SBP, fasting glucose, WC, and 10-Year Framingham Risk). Network analysis enables complex and substantial numbers of connections to be clearly visualised.

Each circle node represents the significant urinary metabolites in the model, and the size of the circle is proportional to the number of connections. Each line corresponds to direct comparison between the metabolites.

Euclidean geometry was applied for similarity index (for the similarity association matrix upon which the clusters are based); $d_{jk} = \sqrt{\sum_i (x_{ji} - x_{ki})^2}$, where x is a variable, i.e. 95% CIs. The edge weight is proportional to the similarity which was calculated by similarity based on the coefficients (i.e. 4-digit decimal 95% CIs) in the SEM at Year 30. The colour of each node represents the type of cardio-vascular risk factor associated with the metabolites; light colours (e.g. honeydew, light-cyan, papaya-whip, misty-rose) indicate a cardio-metabolic risk factor not significantly associated with metabolites/parameters. The edge cut off is 50%; it only shows edges between nodes that are more than 50% similar [248]. The analyses were conducted using the PAleontological STatistics (PAST) version 4.03 (Copyright Øyvind Hammer 1999–2020) [248]. The Fruchterman-Reingold (1991) algorithm was used to visualize the networks, with a random starting position and the optimal distance (k) vertices calculated as follows [249]:

$$k = C \sqrt{\frac{\text{area}}{\text{number of vertices}}}$$

where the constant C is found experimentally. The vertices should be uniformly distributed in the frame, and k is the radius of the empty area around a vertex. If f_a is the attractive force (which based on Hooke's law and used to attract pairs of endpoints of the graph's edges towards each other) and f_r is nearly equivalent to the repulsive force (applying the *grid-variant algorithm*, used to separate all pairs of nodes), with d the distance between the two vertices, then

$$f_a(d) = d^2/k$$

$$f_r(d) = \frac{k^2}{d} u(2k - d), \text{ where } u(x)=1, \text{ if } x > 0; 0, \text{ otherwise}$$

The algorithm finds a configuration where the balance of attractive and repulsive forces are in equilibrium for the network plots [249].

1-1. Network analysis of the metabolites

Glycine (which stimulates insulin secretion in humans and is good for patients with diabetes [184]) and amino adipic acid (which faster global cognitive decline risk was observed in participants with lower concentrations of alpha-amino adipic acid (α -AAA) [250, 251]), appear in the white ethnicity group as significant metabolites.

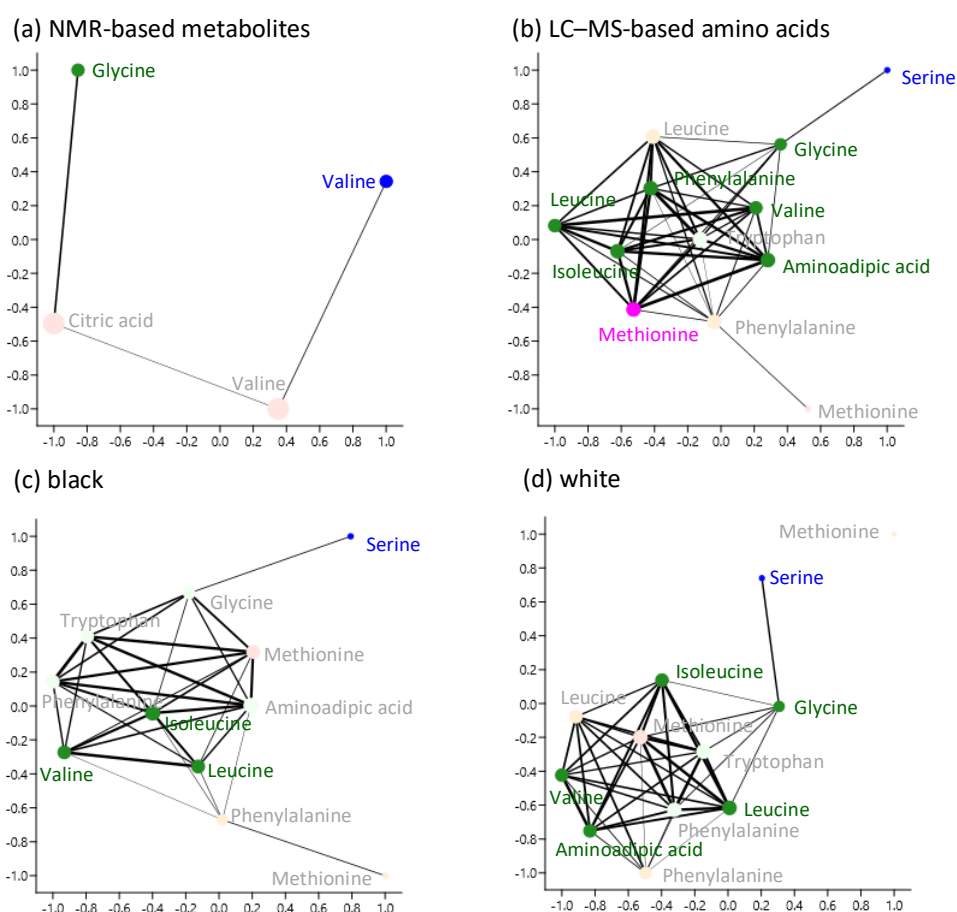
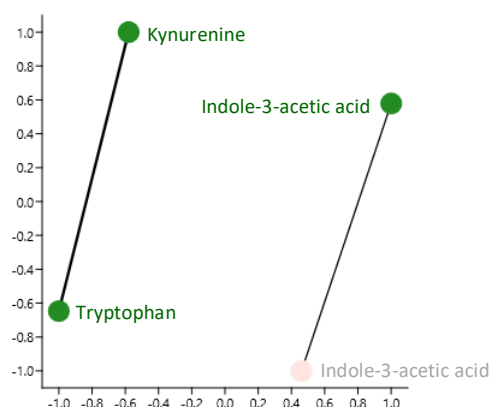
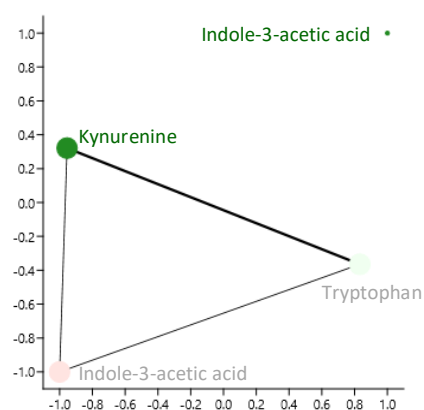


Figure 63 Network analysis of the (a) NMR-based metabolites and (b) LC-MS-based amino acids discovered in the structural equation modelling (SEM) (using the 95% CIs in Table 21; N=346 and Table 23; N=420) at Year 30; (c) black and (d) white ethnic groups in amino acids models – a colour of the node indicates each cardio-metabolic risk factor significantly associated with amino acids (pink: SBP, green: fasting glucose, blue: WC; statistically not significant amino acids were also included - light-cyan: honeydew: fasting glucose, light-cyan: WC, papaya-whip: 10-Year Framingham Risk, misty-rose: SBP)

Urinary kynurenine appears in the black ethnic group as a significant metabolite; its levels are higher in major depression [200-202].



(a) black



(b) white

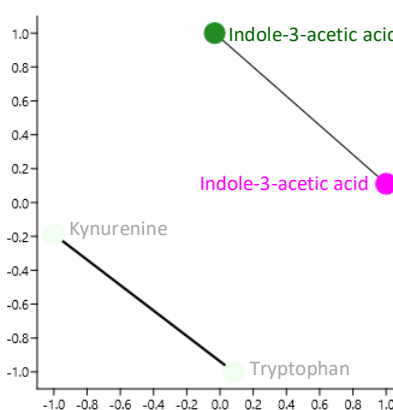


Figure 64 Network analysis of the LC–MS-based tryptophan pathway metabolites at Year 30 discovered in the structural equation modelling (SEM) (using the 95% CIs in Table 23, N=413); (a) black and (b) white ethnic groups – a colour of the node indicates each cardio-metabolic risk factor significantly associated with each metabolite (pink: SBP, green: fasting glucose, blue: WC; statistically not significant amino acids were also included - honeydew: fasting glucose, misty-rose: SBP)

1-2. Network analysis of the brain MRI parameters

Total right entorhinal area volume (right EA; a brown node on which the 10-Year Framingham Risk is significantly associated) is far from the other parameters. Total temporal lobe WM left, and WM volumes (green nodes on which fasting glucose are significantly associated) are also distinguished from other parameters of CBF and FA.

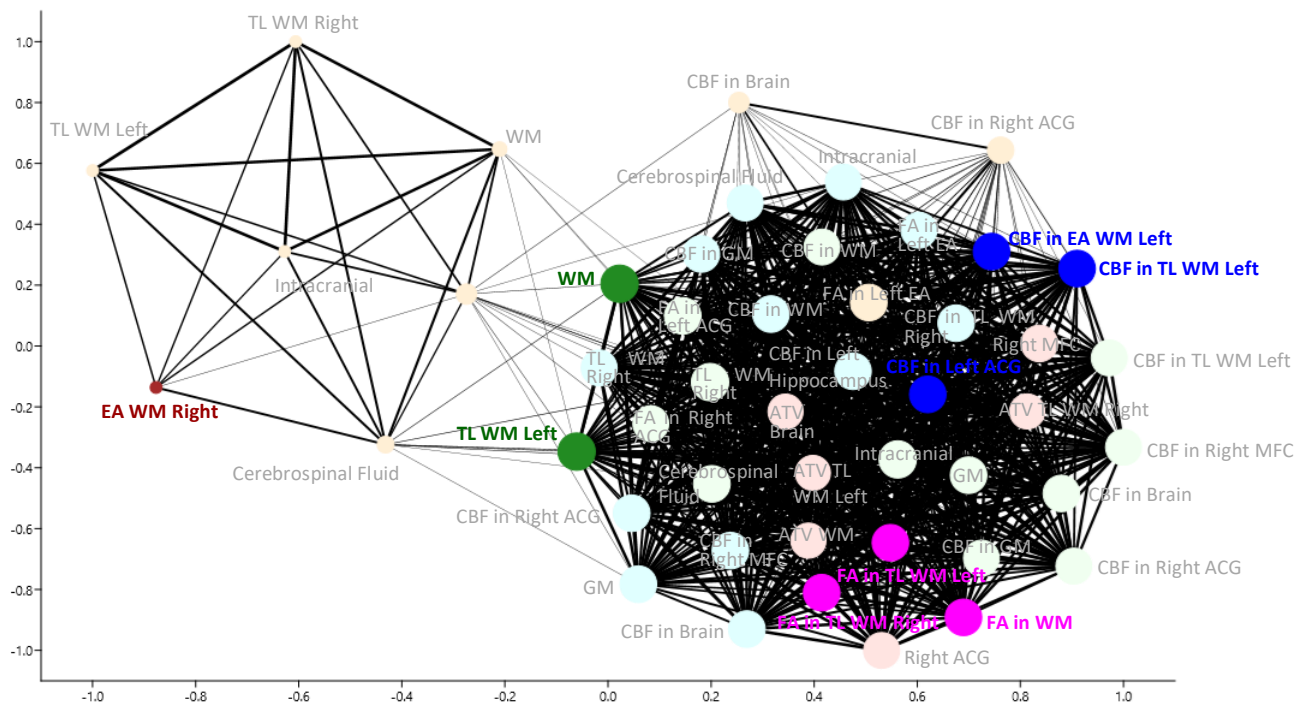


Figure 65 Network analysis of the brain MRI-derived parameters at Year 30 discovered in the structural equation modelling (SEM) (using the 95% CIs in Appendix Table 10, N=155); the colour of the node indicates each cardio-metabolic risk factor significantly associated with the parameters (pink: SBP, green: fasting glucose, blue: WC; and not statistically significant -  honeydew: fasting glucose,  light-cyan: WC,  papaya-whip: 10-Year Framingham Risk,  misty-rose: SBP)

* TL: Temporal lobe, WM: White matter, EA: Entorhinal area, CBF: Cerebral blood flow, FA: Fractional anisotropy, ATV: Abnormal tissue volume, ACG: Anterior cingulate gyrus, MFC: Medial frontal cortex, GM: Gray matter

1-3. Network analysis of the metabolites and brain MRI parameters

CBF in the left entorhinal area and CBF in temporal lobe WM left (on which WC is significantly associated) are strongly connected and in addition, indole-3-acetic acid (on which fasting glucose is significantly associated) is also connected (Figure 66). (The outcomes are presented in Table 29). Total intracranial volume and the other amino acids (aminoadipic acid, tryptophan, and phenylalanine on which fasting glucose is significantly associated) are also connected. The outcomes of the cardio-metabolic risk-specific integrated visualised information regarding the cognitive outcome are presented in this section (as a part of Objective 5). The outcomes discovered in the SEM for the metabolites and MRI parameters measured at Year 30 which are significantly connected to cardio metabolic risk factors are presented in colour nodes (pink: SBP, green: fasting glucose, and blue: WC).

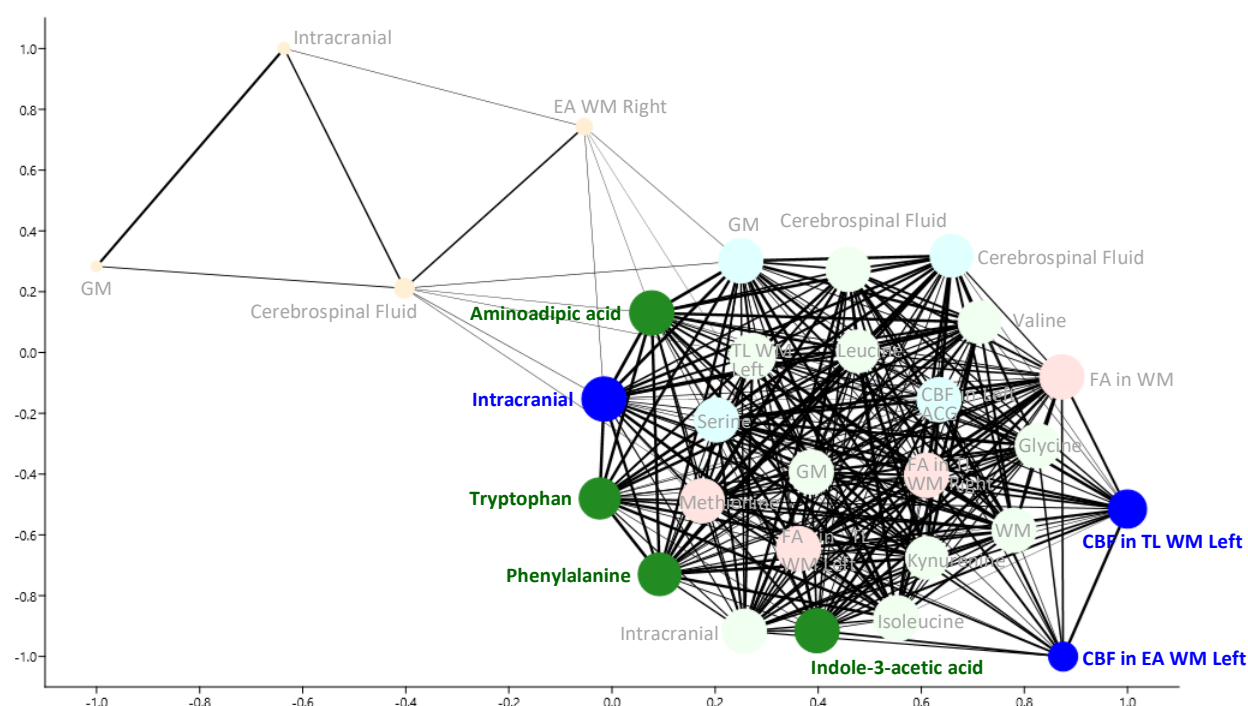


Figure 66 Network analysis of both metabolites and the brain MRI-derived parameters at Year 30 discovered in the structural equation modelling (SEM) (using the 95% CIs in Table 29, N=74); the colour of the node indicates each cardio-metabolic risk factor significantly associated with metabolites and parameters (pink: SBP, green: fasting glucose, blue: WC; and not statistically significant - light green: fasting glucose, light blue: WC, light orange: 10-Year Framingham Risk, light pink: SBP)

* TL: Temporal lobe, WM: White matter, EA: Entorhinal area, CBF: Cerebral blood flow, FA: Fractional anisotropy, ATV: Abnormal tissue volume, ACG: Anterior cingulate gyrus, GM: Gray matter

2. Network analysis of the brain MRI parameters at Year 25 and Year 30

In Chapter IV, Section 3, I investigated which brain MRI parameters measured at Year 25 and Year 30 are significant regarding composite cognition with each cardio-metabolic risk factor (i.e. SBP, DBP, fasting insulin, WC and BMI) after adjusting for other significant exposures. Total ICV was adjusted in each model. The network analysis of the brain MRI parameters (at Year 25 and Year 30) from the Bayesian two-level model of cognition (based on Figure 57 to Figure 60) are presented in Figure 67 and Figure 68.

Abnormal tissue volume parameters showed negative associations regarding cognition in the model where SBP was adjusted (based on Figure 57). Figure 67 (a) clearly shows that abnormal tissue volume parameters are distinguished from the other parameters (number of activated voxels within right ACG and MFC, respectively, and mean of FA in deep WM GM). Percent change of vascular reactivity temporal lobe WM right and left and number of activated voxels within left MFC are strongly connected in the model where BMI was adjusted.

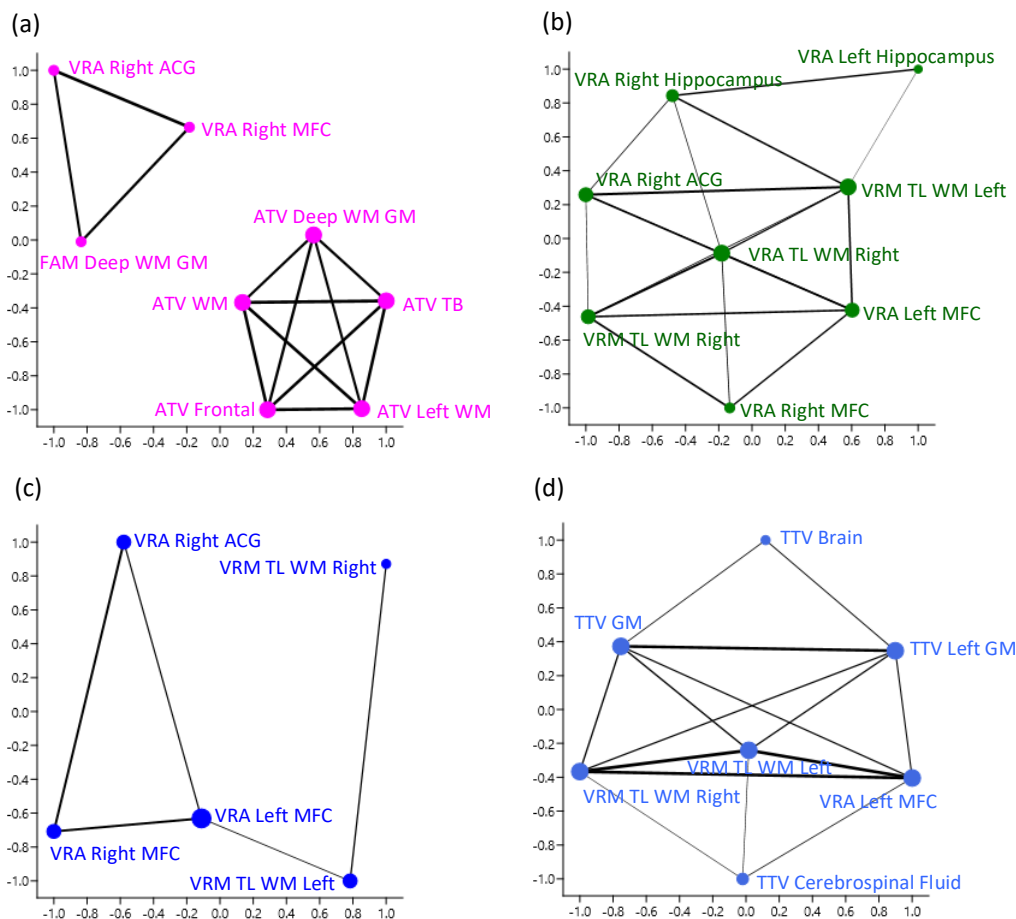


Figure 67 Network analysis of the significant brain MRI-derived parameters discovered in each Bayesian multilevel regression regarding cognition adjusted for each cardio-metabolic factor (a) SBP (pink); (b) fasting insulin (green); (c) WC (blue); (d) BMI (royal blue), measured at Year 25 and Year 30 (using the 95% CrIs in Figure 57 to Figure 60) *VRA: Number of activated voxels within the ROI, FAM: Mean of FA values in an ROI, ATV: Abnormal tissue volume, TB: Total brain, WM: White matter, GM: Gray matter, ACG: Anterior cingulate gyrus, VRM: Mean percent change of vascular reactivity in the ROI, TL: Temporal lobe, MFC: Medial frontal cortex, TTV: Total tissue volume

Figure 68 is another form of a network analysis plot which integrated estimates (95% CrIs) of the significant parameters altogether discovered in each the Bayesian two-level model of the cognitive outcome (based on Figure 57 to Figure 60). Degeneration of cognition on brain MRI parameters measured at Years 25 and 30 from the Bayesian multilevel models are presented as a network plot using the 95% CrIs. The brain MRI parameters for each cardio-metabolic factor SBP (pink); fasting insulin (green); WC (blue) are shown, and the parameters of abnormal tissue volume (pink colour nodes) show ‘negative’ associations regarding cognition where SBP was adjusted, and they were disconnected from other parameters which are ‘positively’ associated regarding cognition for fasting insulin (green colour nodes) and WC (blue colour nodes).

Figure 68 The integrated network analysis of the significant brain MRI-derived parameters discovered in each Bayesian multilevel regression regarding cognition adjusted for each cardio-metabolic factor SBP (pink); fasting insulin (green); WC (blue), measured at Year 25 and Year 30 (using the 95% CrIs in Figure 57 to Figure 60) *VRA: Number of activated voxels within the ROI, FAM: Mean of FA values in an ROI, ATV: Abnormal tissue volume, TB: Total brain, WM: White matter, GM: Gray matter, ACG: Anterior cingulate gyrus, VRM: Mean percent change of vascular reactivity in the ROI, TL: Temporal lobe, MFC: Medial frontal cortex

3. Classification and profiling using the metabolites

Classification by a decision tree induction method was employed for integration of the discovered risk factors measured at Years 25 and 30, and for classification of the risk factors at Year 30.

Using the recognised significant exposures and the standardized urinary metabolites (from Objective 1 to 4⁷), the 1st stage tree model was built regarding binary adjusted composite cognition, i.e. cognitively impaired/not impaired ($z\text{-score} \geq 0$) at Year 30. Next, the 2nd stage tree model was built, using the recognised ‘variables importance’. The variable importance, which is an indication of which predictors are most useful for predicting the response variable, based on the count (number of times in the tree that a particular variable is used in a split), the surrogate count (tallies the number of times that a variable is used in a surrogate splitting rule), residual sum of squares (RSS) and the relative importance [252].

For the tree-based statistical models, the training data was used, and these models are subsequently used to score (predict or classify) response values for new data. The most frequent response level (i.e. cognitively impaired/not impaired) of the training observations in a leaf is used to classify observations in that leaf [252]. The splitting criteria based on impurity used the decrease in the Gini index to split each variable and then to determine the split. The Gini index criterion defines $i(\tau)$ as the Gini index that corresponds to the average square error (ASE) of a class response and is given by $i(\tau) = - \sum_{j=1}^J p_j^2$ (Hastie *et al*, 2009), where p_j is the proportion of observations that have the j th response value [252, 253].

The best splitting variable (s) and the best cut-off value to produce the highest reduction (Δi) in impurity using the HPSPLIT procedure in SAS/STAT® 14.1 are as follows [252]:

$$\Delta i(s, \tau) = i(\tau) - \sum_{b=1}^B p(\tau_b | \tau) i(\tau_b)$$

where $i(\tau)$ is the impurity of a parent node τ , τ_b denotes the b^{th} child node, $p(\tau_b | \tau)$ is the proportion of observations in τ that are assigned to τ_b , and B is the number of branches after splitting τ .

3-1. Tree-based statistical models for the metabolites

A total of 293 subjects were used for the Gini split criterion for the 1st stage classification tree for the adjusted cognitive score on the significant exposures and metabolites measures at Year 30. By default, the growth process continues until the tree reaches a maximum depth of 10. The CVMODELFIT option was used to request a model assessment that is based on cross validation to assess pruning parameters [252]. The cost-complexity pruning algorithm (Breiman *et al.*, 1984) was applied, which is based on making a trade-off between the complexity

⁷Age, race, sex, education, CES-D, smoking, HBP medication, sleep quality, family income, physical activity, urinary creatinine, methionine, glycine, isoleucine, leucine, phenylalanine, valine, aminoadipic acid, serine, indole-3-acetic acid, tryptophan, kynurenine, SBP, fasting glucose, WC.

(size) of a tree and the error rate to help prevent overfitting; 20 number of leaves was used for the tree models for metabolites [254]. The 10-fold cross validation assessment (which is randomly partitioned into 10 equal sized subsamples and a single subsample is retained as the validation data for testing the model) was conducted. The 10-fold cross validation-based misclassification rate was 0.275 with the average square error 0.242.

Table 36 The 1st stage tree model for metabolites – (1) 10-fold cross validation assessment of model and (2) Variable importance

10-fold cross validation assessment of model								
N Leaves	Average square error				Misclassification rate			
	Min	Avg	SE	Max	Min	Avg	SE	Max
20	0.132	0.242	0.086	0.366	0.138	0.275	0.103	0.423

Variable importance	Training		Count-based
	Relative	RSS-based	
Race	1.0000	6.488	1
Education	0.645	4.186	2
Family income	0.496	3.218	3
Glycine	0.388	2.517	2
Methionine	0.363	2.357	2
Sex	0.322	2.087	1
Aminoadipic acid	0.322	2.087	1
Age	0.267	1.732	1
CES–D	0.265	1.719	1
WC	0.265	1.717	1
Tryptophan	0.260	1.685	1
SBP	0.256	1.663	1
Fasting glucose	0.240	1.554	1
Kynurenine	0.216	1.402	1

* CES–D: Center for epidemiologic studies-depression

The area under the receiver operating characteristic curve (AUC) of the training data was 0.96. The variable importance table shows the number of times that a variable is used as a surrogate, and the information was used to build the 2nd stage tree classification model in the following section. Glycine, methionine, aminoadipic acid, tryptophan and kynurenine were included among the urinary metabolites. Two cardio-metabolic risk factors, i.e. WC and fasting glucose, were included. The following analysis was conducted only using the variables based on the variable importance from the initial analyses to preserve statistical power (i.e. avoid missing cases in many variables). With the recognised significant other exposures and metabolites at Year 30, the cognitively impaired individuals were predicted as follows based on the 2nd stage tree model for metabolites: (1) black, education: 10–14 years, standardized glycine level less than < 1.503 (N=84); (2) black, education: 15–20 years, WC greater than ≥ 75 (N=86); (3) white, education: 12–15, 19 years, fasting glucose greater than ≥ 91 (N=50). The AUC of the training data was 0.94. A total of 390 subjects were used for the Gini split criterion. The 10-fold cross validation-based misclassification rate was $(60+58)/390 = 0.3$ with the average square error 0.25. A confusion matrix contains information about actual and predicted cognitively impaired/not impaired subjects from the 2nd stage tree model. The error rate (N of wrong predictions/ N of observations that have the response level) was reported.

Table 37 The 2nd stage tree model for metabolites – (1) 10-fold cross validation assessment of model, (2) Confusion matrices and (3) Variable importance

10-fold cross validation assessment of model								
N	Misclassification rate				Average square error			
Leaves	Min	Avg	SE	Max	Min	Avg	SE	Max
20	0.139	0.300	0.078	0.441	0.127	0.247	0.070	0.392

Confusion matrices	Actual	Predicted		Error rate
		Not impaired	Impaired	
Model based	Not impaired	186	21	0.101
	Impaired	21	162	0.115
Cross validation	Not impaired	147	60	0.290
	Impaired	58	125	0.317

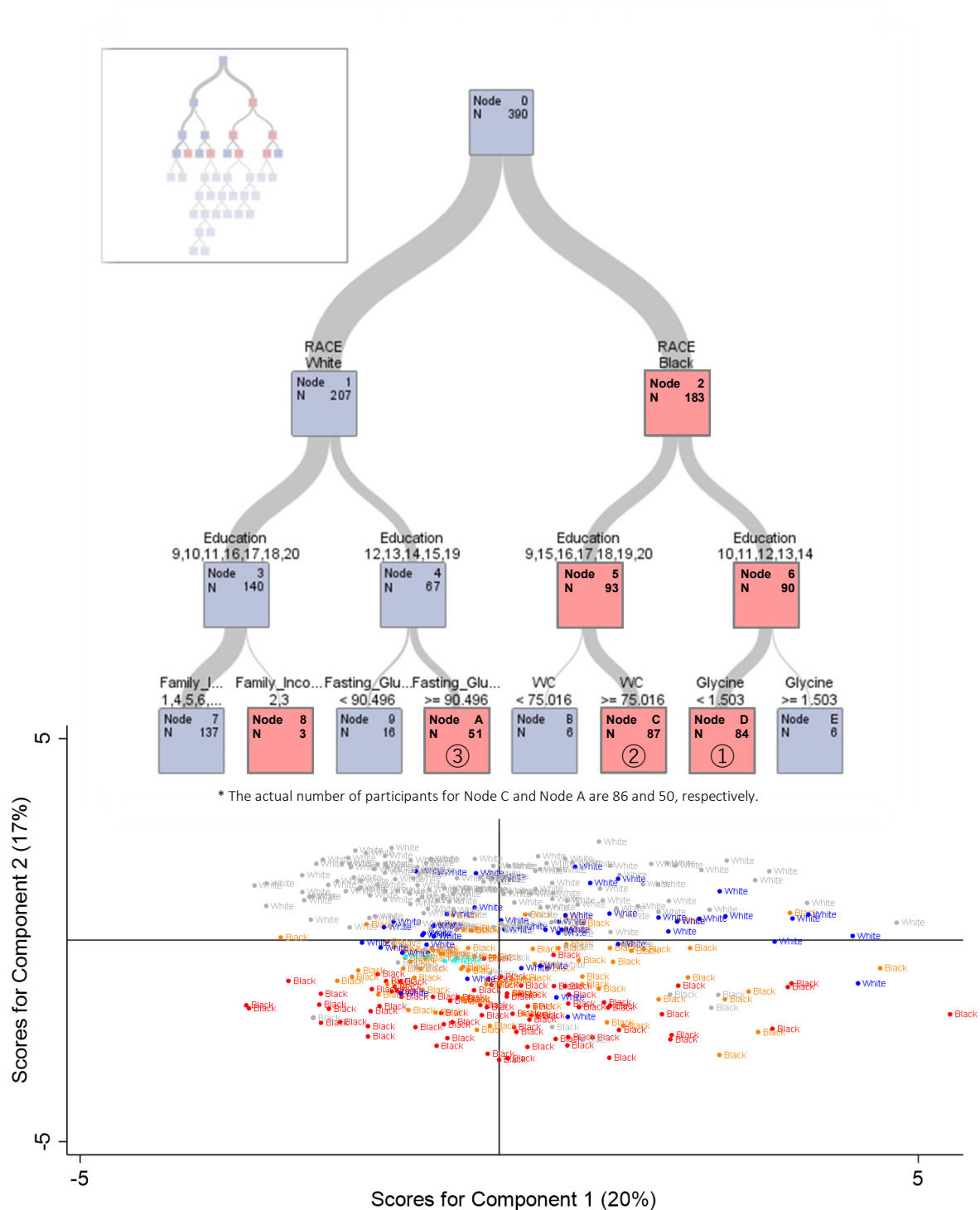
Variable importance	Training		Count-based
	Relative	RSS-based	
Race	1.000	7.194	1
Education	0.639	4.593	3
Family income	0.380	2.732	2
CES–D	0.377	2.711	2
Glycine	0.373	2.686	2
WC	0.349	2.508	2
Fasting glucose	0.347	2.498	1
Tryptophan	0.336	2.414	2
Kynurenine	0.324	2.332	2
Age	0.302	2.171	1
Methionine	0.236	1.700	1

Decision trees are a form of multiple effect analyses [255]. The model produces an output that describes the scoring model with interpretable node rules and it also produces a detailed score code output which describes the scoring algorithm [256]. A limitation of tree models is that small changes in the data can lead to very different splits, and this undermines the interpretability of the model. Another limitation of the tree model is that it often lacks the flexibility needed to capture smooth relationships between the predictor variables and the response [252, 257, 258]. I conducted other data mining methods for the model comparisons, with the variables recognised from the 1st stage tree model for metabolites (with the HPSPLIT Procedure, Table 36) using SAS® Enterprise Miner™ 14.1: High-Performance Procedures. The high-performance (HP) data mining procedures offers scalable solutions in the important common space shared between data mining and machine learning (Outcomes of other data mining methods are presented in Appendix Table 11) [259, 260]. The results showed that the decision tree (i.e. empirical tree represents a segmentation of the data that is created by applying a series of simple rules⁸) was the best model for the data, based on the misclassification rate (0.18) and average on squared error (0.14) criteria in the 30% validation data. The variable importance suggested from this model are race, education, family income, sex, CES–D, glycine, WC. Although the misclassification rate is lower than other models, the selected model using the HPSPLIT Procedure (Table 37) included many important urinary

⁸ This is different from the HPSPLIT procedure, which selected the best splitting variable and the best cut-off value based on the highest reduction in impurity, used in the tree models in Chapter V, Sub-section 3-1 and 4-1.

metabolites in prediction, and its model-based misclassification rate and average square error were much lower (0.108 and 0.087, respectively). Besides, the decision tree-based gradient boosting model showed a misclassification rate of 0.21 and average on squared error of 0.18 in the validation data. Gradient boosting is an algorithm which embraces the notion of ensemble learning, whereby multiple simple learning algorithms are used jointly to obtain better predictive performance [261].

3-2. The principal components and profiling using the metabolites



- Negative Cognition ① Black, Education: 10–14y, standardized Glycine < 1.503 (Red IDs, N=84)
 ② Black, Education: 15–20y, WC ≥ 75 (Orange IDs, N=86)
 ③ White, Education: 12–15,19y, Fasting glucose ≥ 91 (Blue IDs, N=50)

Figure 69 The 2nd stage classification tree for the adjusted cognitive score on the important exposures and metabolites at Year 30 (N=390); node colour: red–cognitively impaired/ blue–not impaired. The below figure is its PCA score plot for the component 1 (20%) and the component 2 (17%); colour–cognitively impaired/ grey–not impaired. The number of cognitively impaired participants (i.e. z-score < 0), but predicted as not impaired are 20 (i.e. false negative)

For the detailed profiling of the important metabolites recognised from the 2nd stage tree classification model, I conducted PCA, a dimensionality-reduction method transforming a number of variables based on the variable importance into a low-dimensional space that still contains most of the variance in the data (Figure 69).

I conducted PCA using the recognised variable importance in Table 37 (2) for the metabolites profiling regarding the classified participants from the 2nd tree model; it was conducted as a means of reducing the dimensionality of a multivariate data set and, also, of illuminating its interpretation by a smaller number of characteristics [262]. Among the 390 subjects, 365 subjects were used for PCA; family income (N=388), CES–D (N=385), fasting glucose (N=386), tryptophan (N=376). The five components explain 70% of the variance across the 11 variables. The orthogonal varimax rotation was used which maximizes the sum of the variances of the squared loadings leading to a model where all the coefficients will either be large or near zero, with few intermediate values. The goal is to associate each variable to at least one factor.

Table 38 The rotated PCA components (eigenvectors) for the variable importance from the 2nd tree model (N=365)

	Component 1 (20%)	Component 2 (17%)	Component 3 (13%)	Component 4 (11%)	Component 5 (10%)	Unexplained components
Age	0.03	-0.01	-0.01	-0.04	0.89	0.12
Race	0.06	0.61	-0.05	0.16	0.09	0.33
Education	-0.004	0.53	-0.04	-0.06	0.01	0.44
Family income	-0.002	0.52	0.08	-0.29	-0.19	0.30
CES–D	0.02	-0.01	-0.02	0.82	-0.05	0.20
WC	0.10	-0.01	0.54	0.25	-0.24	0.38
Fasting glucose	0.08	-0.08	0.64	-0.28	-0.02	0.32
Glycine	0.24	-0.11	-0.53	-0.14	-0.30	0.32
Tryptophan	0.62	0.09	-0.01	0.04	0.03	0.21
Kynurenine	0.54	0.09	0.09	0.13	0.09	0.37
Methionine	0.50	-0.20	-0.02	-0.18	-0.001	0.32

* CES–D: Center for epidemiologic studies-depression

* WC: Waist circumference

* Variables prominent in each component are in bold

There were no significant differences in race, education, tryptophan, kynurenine, methionine, and glycine, respectively, between false negative (i.e. cognitively negative but predicted as not cognitively impaired in the model; Table 37, N=20) versus the rest group (Kruskal-Wallis equality-of-populations rank test).

Although not observed clearly in this study, it has been reported that the kynurenine level is higher in patients with major depression, and L–methionine triggered a decline in the cognitive capacities [178]. The standardized methionine concentration was significantly higher in black (mean 0.25, N=183) vs white (mean -0.23, N=207);

$p < 0.001$, two-sample Wilcoxon rank-sum test (i.e. equality tests on unmatched data) as shown in Figure 70. Glycine is known to stimulating insulin secretion and good for patients with diabetes (as discussed in Chapter III. Predictions of cognition at Year 30) [184]. It is reported that the tryptophan uptake mechanism regulates its transport across the blood-brain barrier, leading to more focussed attention [194].

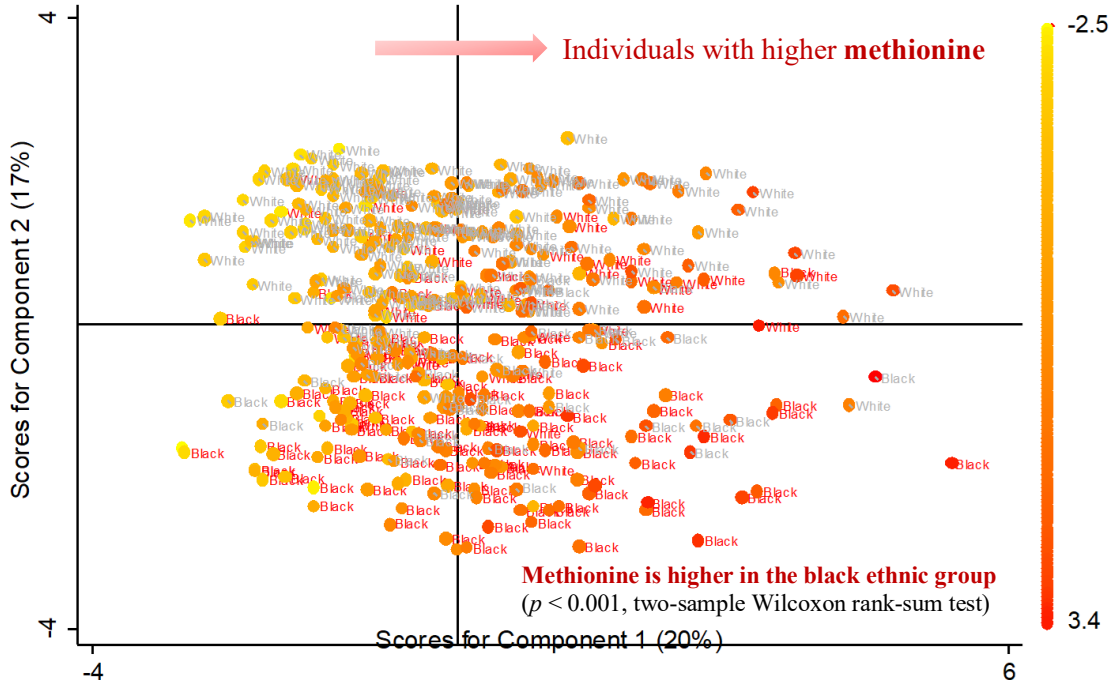


Figure 70 The PCA score plot for the component 1 (20%) and the component 2 (17%); colour—cognitively impaired/ grey—not impaired; a darker red colour scale indicates higher methionine

Turning now to the volcano plots which can help to identify potential differences in metabolite concentrations between the cognition group (N=382 from 390). Figure 71 shows statistical significance (p-value from Fisher's exact test on the $-\log_{10}$ scale) versus magnitude of change (risk difference) for the lowest and highest standardized metabolite concentrations per individual among all the metabolites. The second lowest metabolite concentration per individual in the cognition group was indole-3-acetic acid, and glycine was the second highest in the **not** impaired group; $\log_{10}$ p-value > 1 (i.e. p-value < 0.1). Besides, kynurenine (impaired 14.4% vs not impaired 10.6%) and methionine (impaired 8.0% vs not impaired 5.3%) showed relatively high frequencies of the highest metabolite concentrations per individual in the cognitively impaired group (plots are not presented here).

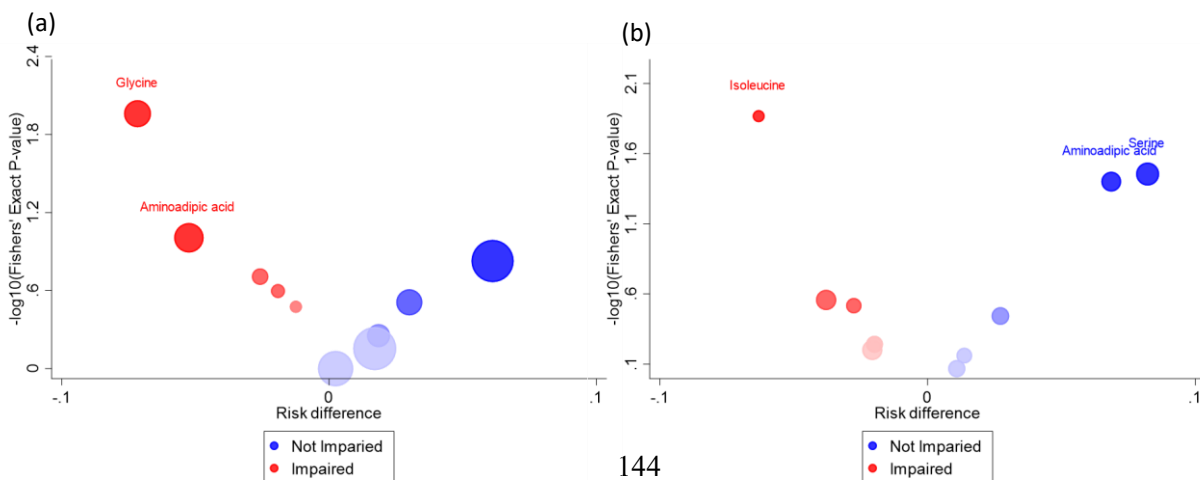


Figure 71 Volcano plots for (a) the lowest and (b) the highest standardized metabolite concentrations per individual in the cognition group (N=382); each bubble represents the event (i.e. lowest/highest standardized metabolite concentration), the size of which indicates total frequency of the event; colour is used to indicate the direction of the response level (i.e. cognitively impaired/not impaired), and colour saturation corresponds to the statistical significance ($-\log_{10}(p \text{ value})$) of the response level effect [263]. Glycine and aminoadipic acid are the lowest concentration in the cognitively impaired group; serine and aminoadipic acid are the highest concentration in the **not** impaired group; $\log_{10} p\text{-value} > 1$ (i.e. $p\text{-value} < 0.1$)

4. Classification and profiling using the brain MRI parameters

I applied the same analytic approaches for tree-based statistical models as presented in Section 3. Using the recognised significant exposures and the standardized brain MRI parameters (from Objectives 1 to 4)⁹, the 1st stage tree model was built regarding binary adjusted composite cognition, i.e. cognitively impaired/not impaired, at Year 30. Next, the 2nd stage tree model was built using the variables recognised in the 1st stage tree model.

4-1. Tree-based statistical models for the brain MRI parameters

A total of 83 subjects (who have no missing cases in the list of variables recognised from Objectives 1 to 4) were used for the Gini split criterion for the 1st stage classification tree for the adjusted cognitive score on the significant exposures and brain MRI-deprived parameters at Year 30 (Table 39). The cost-complexity pruning algorithm with 25 number of leaves was used for the tree models for brain MRI parameters.

Table 39 The 1st stage tree model for the brain MRI parameters – (1) 10-fold cross validation assessment of model and (2) Variable importance

10-Fold cross validation assessment of model								
N	Misclassification rate				Average square error			
Leaves	Min	Avg	SE	Max	Min	Avg	SE	Max
10	0	0.325	0.124	0.444	0	0.325	0.124	0.444

Variable importance	Training		Count-based
	Relative	RSS-based	
Smoking	1.000	2.513	1
Family income	0.969	2.434	1
Percent change of vascular reactivity temporal lobe WM right	0.873	2.195	1
FA right MFC	0.831	2.089	1
Number of activated voxels within temporal lobe WM right	0.829	2.083	1
Number of activated voxels within right ACG	0.713	1.792	1
WC	0.626	1.574	1
ATV frontal	0.547	1.374	1
CES-D	0.503	1.265	1

* WM: white matter, ACG: Anterior cingulate gyrus, FA right MFC: Fractional anisotropy in right medial frontal cortex, ATV: Abnormal tissue volume

⁹ Age, race, sex, education, CES-D, smoking, HBP medication, sleep quality, family income, physical activity, urinary creatinine, ATV deep WM GM, ATV frontal, ATV WM, ATV left WM, ATV total brain, ATV TL WM right, ATV TL WM left, CBF right ACG, CBF left ACG, CBF EA WM left, CBF right MFC, CBF left hippocampus, CBF GM, CBF WM, CBF brain, CBF TL WM right, CBF TL WM left, FA right ACG, FA left ACG, FA left EA, FA right MFC, FA WM, FA TL WM right, FA TL WM left, FA ventricle, Total right ACG, Total EA WM right, Total right MFC.

The variable importance list included only one cardio-metabolic risk factor, i.e. WC. I conducted the following analysis only using the variables based on variable importance from the initial analyses to preserve statistical power. The AUC of the training data was 0.97. A total of 231 subjects (who have no missing values in the list of variable importance) were used for the Gini split criterion. In addition, with the recognised significant other exposures and brain MRI parameters at year 30, cognitively impaired individuals were predicted as follows based on the 2nd stage tree model for the brain MRI parameters: (1) smoking: current, WC greater than ≥ 77.5 (N=21); (2) smoking: never/ former, family income category: 1,2,3,5,6,8 (relatively low; Table 1), WC greater than ≥ 88.6 (N=56). The 10-Fold Cross Validation-based misclassification rate was $(41+33)/231 = 0.32$ and the average square error was 0.29.

Table 40 The 2nd stage tree model for the brain MRI parameters – (1) 10-fold cross validation assessment of model, (2) Confusion matrices and (3) Variable importance

10-fold cross validation assessment of model								
N	Misclassification rate				Average square error			
Leaves	Min	Avg	SE	Max	Min	Avg	SE	Max
25	0.174	0.322	0.070	0.450	0.168	0.288	0.071	0.401

Confusion matrices	Actual	Predicted		Error rate
		Not impaired	Impaired	
Model based	Not impaired	138	4	0.028
	Impaired	9	80	0.101
Cross validation	Not impaired	101	41	0.289
	Impaired	33	56	0.371

Variable importance	Training		Count-based
	Relative	RSS-based	
Family income	1.000	4.403	4
Smoking	0.902	3.970	1
WC	0.895	3.941	4
Number of activated voxels within right ACG	0.759	3.341	4
FA right MFC	0.726	3.195	3
Percent change of vascular reactivity temporal lobe WM right	0.544	2.395	2
CES-D	0.492	2.166	3
ATV frontal	0.422	1.857	2
Number of activated voxels within temporal lobe WM right	0.321	1.414	1

* WM: white matter, ACG: Anterior cingulate gyrus, FA right MFC: Fractional anisotropy in right medial frontal cortex, ATV: Abnormal tissue volume, CES-D: Center for epidemiologic studies-depression

The variable importance recognised from the 1st stage tree model for the brain MRI parameters (Table 39) are used for building other data mining methods (with the HPSPLIT Procedure) as reported in Appendix Table 12. The HP Forest 1 model was the best model for the data, based on the misclassification rate (0.261) and average on squared error (0.178) criteria in the 30% validation data. Random forest (which was introduced by Leo

Breiman in 2001) is an ensemble approach which trains several weak learners in parallel and combines their predicting probabilities. The forest model is an ensemble of hundreds of decision trees. The bootstrap sample randomly selects number of variables in each split to reduce correlation [264]. The variables based on variable importance suggested by the random forest (HP Forest 1) model are in the order of WC, CES-D, percent change of vascular reactivity in temporal lobe WM right (3rd order; showed a positive association with WC), FA in the right MFC, FA right MFC, number of activated voxels within temporal lobe WM right, abnormal frontal tissue volume, number of activated voxels within right ACG, family income and smoking. Nevertheless, the selected model using the HPSPLIT Procedure (Table 40) was maintained; its model-based misclassification rate and average square error were lower than other models (0.056 and 0.047, respectively) and the model is simpler than the random forest one.

4-2. The principal components and profiling using the brain MRI parameters

For detailed profiling of the recognised important brain MRI parameters from the 2nd stage tree classification model, I conducted PCA, transforming a number of variables based on variable importance into a low-dimensional space (Figure 72).

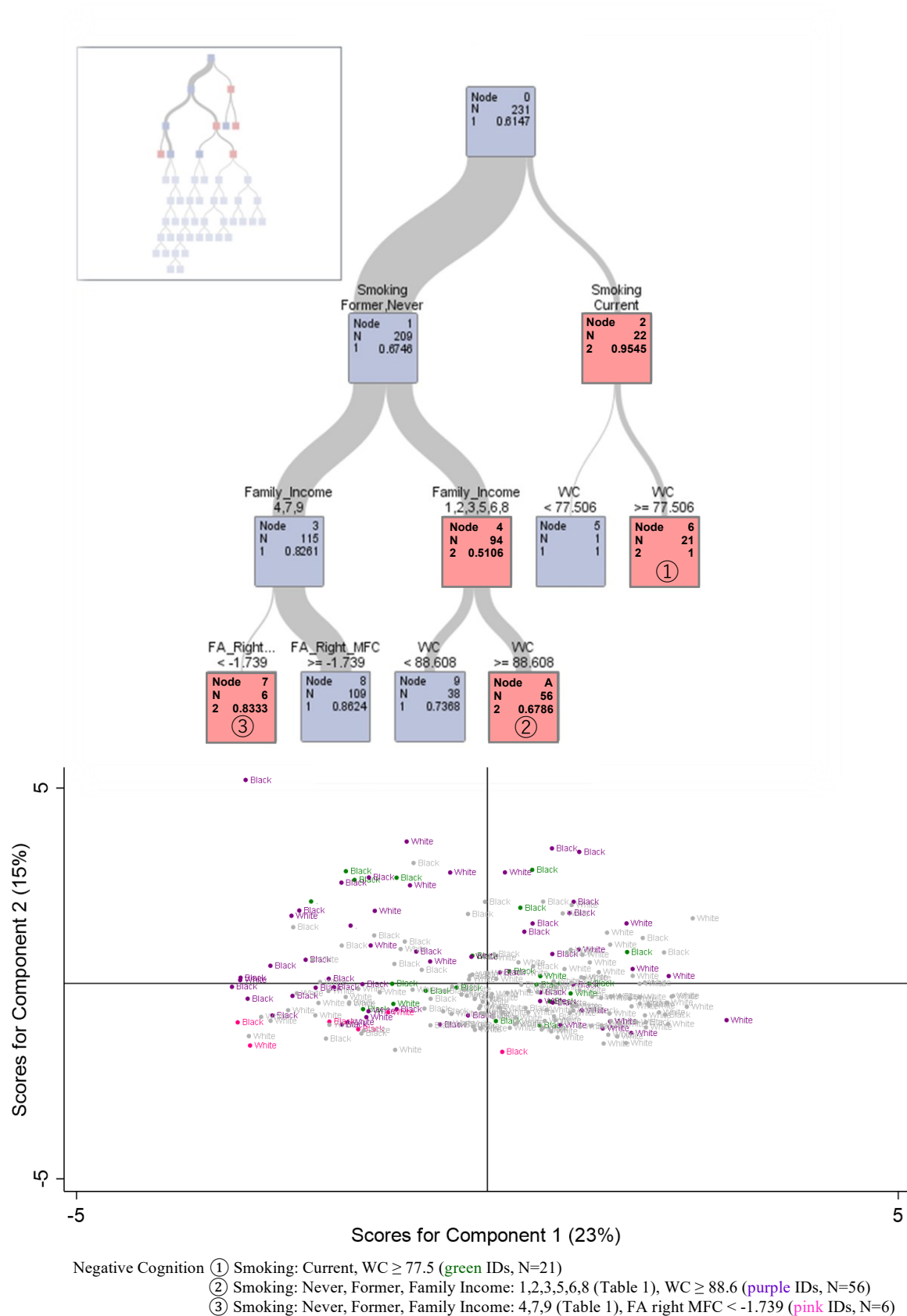


Figure 72 The 2nd stage classification tree for the adjusted cognitive score on the important exposures and brain MRI parameters at Year 30 (N=231); node colour: red–cognitively impaired/ blue–not impaired (some branches have a small Ns and this should be interpreted carefully). The below figure is its PCA score plot for the component 1 (23%) and the component 2 (15%); colour–cognitively impaired/ grey–not impaired. The number of cognitively impaired participants (i.e. z-score < 0) but predicted as not impaired are 25 (i.e. false negative)

For the brain MRI parameters profiling regarding the classified participants from the 2nd tree model, I conducted PCA using the recognised variable importance in Table 40 (2). A total of 231 subjects were used for the analysis. The five components explain 75% of the variance across the 9 variables. The orthogonal varimax rotation was used as previously (Table 41).

Table 41 The Rotated PCA components (eigenvectors) for the variable importance from the 2nd tree model (N=231)

	Component 1 (23%)	Component 2 (15%)	Component 3 (15%)	Component 4 (11%)	Component 5 (11%)	Unexplained components
Family income	0.04	-0.65	-0.01	-0.05	-0.12	0.34
CES-D	0.02	0.73	-0.002	0.01	-0.10	0.29
Smoking	-0.01	-0.01	-0.004	-0.001	0.98	0.02
WC	-0.11	0.07	0.72	-0.06	-0.06	0.29
Percent change of vascular reactivity temporal lobe WM right	0.13	-0.07	0.69	0.05	0.06	0.31
Number of activated voxels within right ACG	0.62	-0.02	-0.05	0.02	-0.07	0.19
Number of activated voxels within temporal lobe WM right	0.63	-0.04	0.06	0.11	0.04	0.17
FA right MFC	0.42	0.17	-0.02	-0.29	0.04	0.53
ATV frontal	0.03	0.02	-0.01	0.95	-0.001	0.07

* Variables prominent in each component are in bold

* WM: White matter

* ACG: Anterior cingulate gyrus

* FA right MFC: Fractional anisotropy in right medial frontal cortex

* ATV: Abnormal tissue volume

There was a significant difference in smoking ($p=0.027$, Bartlett's test for equal variances) between false negatives (which are cognitively impaired but predicted as not impaired) and the rest of the group. In addition, percent change of vascular reactivity in temporal lobe WM right differed significantly between the false negative and the rest of the group ($p=0.022$, Kruskal-Wallis equality-of-populations rank test). There were no significant differences in race or education.

A number of grey colour IDs show the darker green dots in Figure 73. The standardized number of activated voxels within temporal lobe WM right, number of activated voxels within right ACG and FA in right MFC show significant differences in the cognition group, respectively ($p < 0.01$, respectively; two-sample Wilcoxon rank-sum test). (The summation of these parameters in cognitively not impaired (mean 0.76) vs impaired (mean -0.38); $p=0.0002$, two-sample Wilcoxon rank-sum test). In addition, the standardized AV TL WM right (white: mean 0.34 vs black: -0.16) and AV Right ACG (white: mean 0.31 vs black: -0.16) and were significant in race, respectively (white: N=140 vs black: N=89; $p=0.0001$, two-sample Wilcoxon rank-sum test).

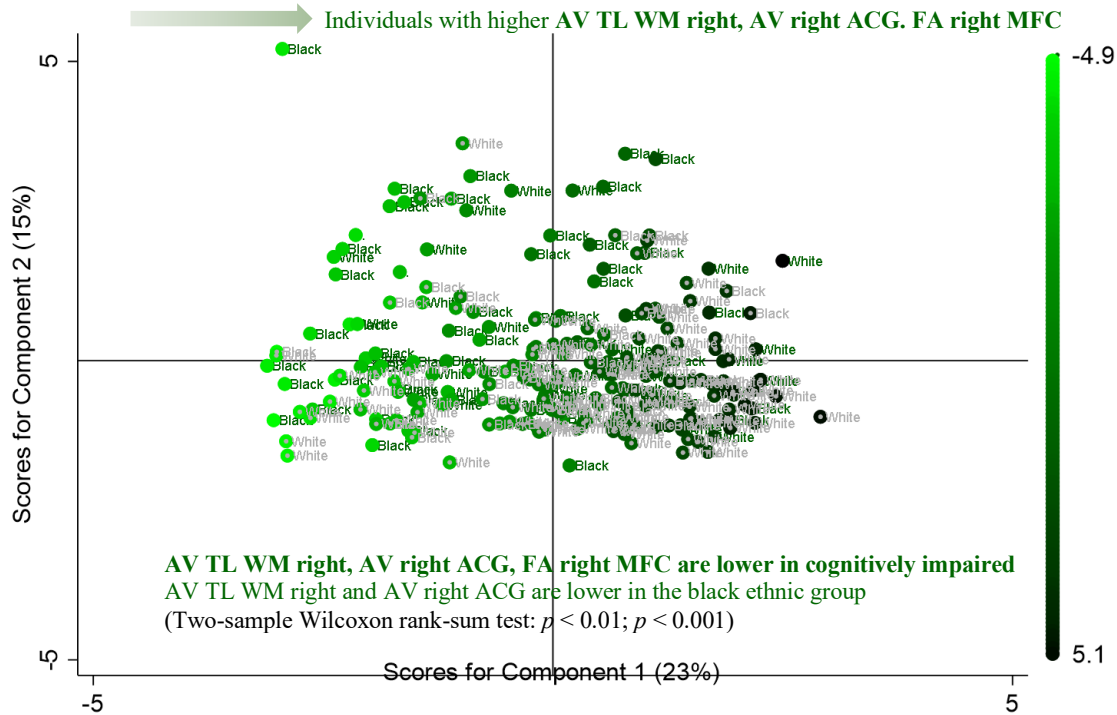


Figure 73 The PCA score plots for the component 1 (23%) and the component 2 (15%); colour—cognitively impaired/ grey—not impaired; a darker green dot indicates a higher sum of the standardized AV TL WM right, AV right ACG, FA right MFC

* AV TL WM: Number of activated voxels within temporal lobe white matter

* AV right ACG: Number of activated voxels within right anterior cingulate gyrus

* FA right MFC: Fractional anisotropy in right medial frontal cortex

Discussion on profiling using the brain MRI parameters

These parameters (AV TL WM right, AV right ACG and FA right MFC, Figure 74) showed **negative associations with a cardio-metabolic risk factor** (SBP, fasting insulin, WC, and BMI) in each Bayesian multilevel regression for Year 25 and Year 30 (Figure 53 to Figure 56). Furthermore, the number of activated voxels within temporal lobe WM right and FA in right MFC showed a **positive association regarding cognition**, respectively, in the Bayesian multilevel regression (where fasting insulin was adjusted; Figure 58). In addition, the number of activated voxels within right ACG showed a **positive association regarding cognition** in each Bayesian multilevel regression (where each SBP, fasting insulin, WC, and BMI was adjusted; Figure 57 to Figure 60). On the other hand, percent change of vascular reactivity in temporal lobe WM right and abnormal tissue volume in frontal showed **positive associations with a cardio-metabolic risk factor** (SBP or fasting insulin, WC, and BMI) in each Bayesian multilevel regression (Figure 53 to Figure 56). Moreover, abnormal frontal tissue volume showed a **negative association regarding cognition** in the Bayesian multilevel regression where SBP adjusted (Figure 57). Nevertheless, percent change of vascular reactivity in temporal lobe WM right showed a positive association regarding cognition in the Bayesian multilevel regression (where fasting insulin, WC, and BMI were adjusted, respectively; Figure 58 to Figure 60). This parameter is clearly distinguished from

the other three parameters mentioned which are negatively associated with the cardio-risk factors. The percent change of vascular reactivity in temporal lobe WM right was classified with WC in the rotated component 3. Abnormal tissue volume in frontal was also independently classified as component 4 from the other three parameters in the rotated component (Table 41). The standardized abnormal frontal tissue volume showed a distinct difference in smoking status ($p=0.013$, Analysis of Variance): The mean (SD) of smoking status ‘current’ (N=22) was 0.54 (1.42) and ‘never/former’ (N=209) was -0.09 (0.9), respectively.

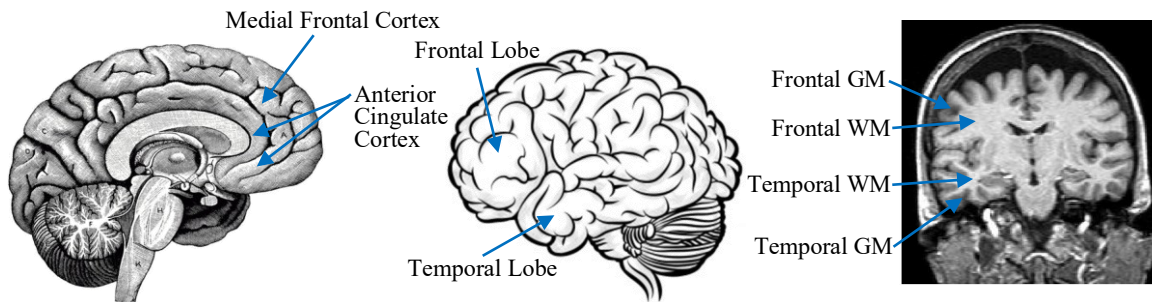


Figure 74 The schematic of brain. (Reused with permission from the corresponding author and the journal: License Number 4983770658728) [204, 265]; GM: Gray matter, WM: White matter

The volcano plots helped to identify potential differences in the lowest standardized brain MRI parameters, i.e. the total volume of temporal lobe WM right and CBF in left entorhinal area (CBF left EA), between the cognition group (N=231) (Figure 75). (The total volume of temporal lobe WM left (NE right) shows a positive association regarding cognition in the Bayesian multilevel regression where BMI was adjusted in Figure 60). The CBF in left entorhinal area (CBF left EA) shows a negative association with each cardio-metabolic risk factor (SBP, fasting insulin, and WC) in each Bayesian multilevel regression for Year 25 and Year 30 (Figure 53 to Figure 55).

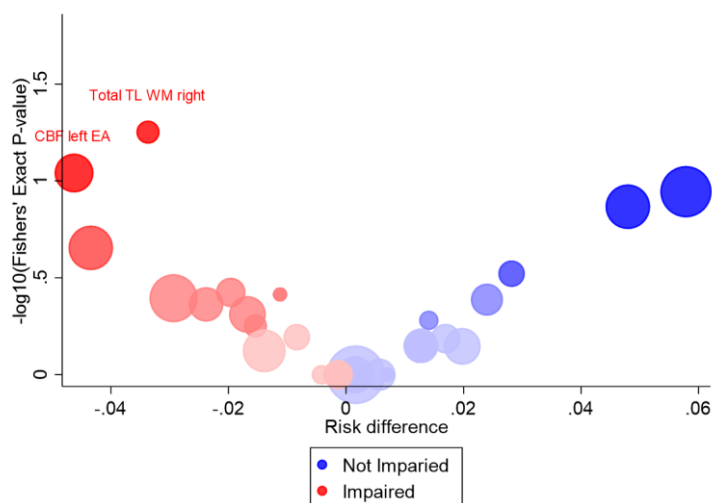


Figure 75 The volcano plots for the lowest brain MRI parameters in the cognition group (N=231); total volumes of temporal lobe white matter (Total TL WM) right and cerebral blood flow in left entorhinal area (CBF left EA) are the lowest parameters in the cognitively impaired group; $\log_{10}$ p-value >1 (i.e. p-value <0.1)

Summary of integrated map

In this chapter, I investigated the cardio-metabolic risk-specific (SBP, WC, and fasting glucose) integrated information regarding cognition. Among the significant factors identified from the previous chapters, key epidemiological exposures (race, education, family income, sex, age, CES–D, smoking), metabolites (glycine, methionine, aminoadipic acid, tryptophan and kynurenine) and the brain MRI parameters were identified as prognostic markers regarding cognition. Using the variables based on variable importance from the selected model, a dimensionality-reduction showed the **tryptophan**, **kynurenine** and **methionine** concentrations dominated component 1 (which explains 20% of the total variance). Concentrations of **kynurenine** and **methionine** were higher in the black ethnic group, respectively.

Using the variables from the selected model, a dimensionality-reduction showed the number of activated voxels within temporal lobe WM right, the number of activated voxels within right ACG and FA in right MFC dominate the component 1 (which explains 23% of the total variance). These brain MRI parameters are **positively** associated regarding cognition in each Bayesian multilevel regression where fasting insulin was adjusted at Year 25 and Year 30. The MFC and ACG are the brain structures involved in the successive stages of memory processing, and the (medial) temporal lobe volume is an early marker of AD as described in Chapter II, Sub-section 1-5. These brain MRI parameters are significantly lower in cognitively impaired individuals (i.e. $z < 0$), and the the number of activated voxels within temporal lobe WM right, and the number of activated voxels within right ACG are significantly lower in the black ethnic group. In addition, the percent change of vascular reactivity in temporal lobe WM right, and abnormal frontal tissue volume showed **positive associations with a cardio-metabolic risk factor** (SBP, fasting insulin, WC, and BMI) where abnormal frontal tissue volume showed a **negative association regarding cognition** and the percent change of vascular reactivity in temporal lobe right showed a **positive association regarding cognition** when fasting insulin, WC, and BMI were adjusted in each Bayesian multilevel regression.

In this chapter, I implemented visualisation of influential urinary metabolites and brain MRI-derived parameters, respectively, using PCA. The discovered physiological mediators were structured by individuals' characteristics of the study cohort.

Chapter VI. Design of a risk assessment tool for brain health

1. Conventional risk prediction tools and their limitations

1-1. Conventional risk tools for cardiovascular disease and the limitations

There are disadvantages in using standardized risk-scoring systems for CVD or obesity assessments. In the present section, I discuss the limitations of some of the most widely used approaches.

The cholesterol-based 10-year Framingham Risk Score (which was introduced in Chapter I) uses as a measure of CVD risk (i.e. one of the cardio-metabolic risk factors) in the study. It did not take into account ethnicity, family history or socio-economic factors. In addition, it excluded some important clinical exposures (such as fibrinogen levels, lipoprotein (a), oestrogen replacement therapy (ERT), importantly BMI, and a type of CVD in women which increases progressively with BMI [266]). Moreover, the definition of the risk, especially nonfatal end-points, is different from other cohort studies and the end-points in clinical trials [32].

The Systematic COronary Risk Evaluation (SCORE) study, representing 2.7 M person-years of follow-up with 12 European cohort studies, provided a 10-year risk of fatal CVD chart based on total serum cholesterol (TC), and TC/HDL-C (age, sex, smoking status, SBP were additionally applied for the model) [267]. It only predicts the risk of CVD mortality. It has been reported that the tool may overestimate risk in countries with a decreasing CVD mortality (and vice versa). The tool may underestimate risk in patients with diabetes, central obesity, family history of premature CVD, low HDL or elevated triglyceride, fibrinogen, lipoprotein A and B, high-sensitivity CRP, or homocysteine levels [267].

The QRESEARCH cardiovascular risk algorithm (QRISK) cardiovascular disease risk algorithm provides a 10-year risk (%) score of CVD from the England and Wales study population. QRISK2, which additionally included patients with type 2 diabetes, applied risk factors of age, ethnicity, sex, smoking status, SBP, TC/HDL, BMI, family history of CHD, Townsend deprivation score, treated hypertension, rheumatoid arthritis, chronic renal disease and AF for the model [268]. The first diagnosis of CVD (myocardial infarction, CHD, stroke and transient ischaemic attack) was recorded on the clinical computer system during the study period as the end point [269]. It may underestimate risk in European populations and the tool is not well validated yet. As the data set is incomplete for some risk factors, imputation and statistical modelling was required to reduce biased estimates. QRISK3 which applied additional clinical exposures has lack of diagnoses, information bias, and potential bias due to missing data [268]. In the next section, I discuss the limitations of some of conventional risk tools for obesity.

1-2. Conventional risk tools for obesity and their limitations

The Obesity Population Risk Tool (OPoRT), which used sex-specific generalized estimating equations (with 5,474 male and 6,402 female), was developed to predict the 10-year risk of obesity for adults (aged 18 and over) using the US longitudinal National Population Health Survey (NPHS) of Canadians (from 1994–5 to 2010–11, followed every two years) [270]. The covariates applied to the model include time, baseline BMI, age, marital status, living arrangements, smoking status, physical activity, education, equivalized income quintiles (adjusted for household size), ethnicity, immigration status, rural status, alcohol consumption, and household ownership [270]. Given the bias in self-reported BMI, the study applied a validated BMI correction equation to all analyses [271]. The logistic regression with a marginal model (i.e. generalized estimating equations (GEE) which is not the subject-specific coefficient, the population-averaged coefficients) was fitted to the longitudinal binary data using a correlation matrix to account for the clustering of observations within individuals across time [272].

A study developed a risk score to predict substantial weight gain ($\geq 10\%$ of baseline weight during follow-up) for the next 5 years with the middle aged 53,758 individuals followed for a median of 5.4 years from 5 countries (the United Kingdom, the Netherlands, Italy, Germany, and Denmark) of the European Prospective Investigation into Cancer and Nutrition (EPIC), 1992–2000 [273, 274]. The stepwise Cox regression model was fitted, and the study identified significant risk predictors for substantial weight gain and age, sex, baseline weight, the level of education, baseline smoking, sports activity, alcohol consumption, and the categorised six food intake groups (i.e. red and processed meat, poultry, fish, bread, cake and biscuits soft drinks) [273]. The risk score was calculated as a linear combination of the predictors. External validity of the score was evaluated in nine other centres of the EPIC study [273]. The pooled estimate of the background probability of avoiding substantial weight gain at 5 years estimated at average values of the predictors was 0.93, implying that under average conditions about 7% of the population will experience substantial weight gain within 5 years [273].

Douketis *et al.* (2005) underlined that weight loss studies have methodologic limitations that restrict the applicability of findings to unselected obese people assessed in everyday clinical practice, and the limitations include an inadequate study duration, large proportions of subjects lost to follow-up, a lack of an appropriate usual care group, and a lack of reporting of outcomes in high-risk subgroups [275]. In the next section, I discuss the limitations of some of conventional risk tools for dementia.

1-3. Conventional risk tools for dementia and their limitations

The CAIDE Dementia Risk Score is a validated tool to predict 20 years later life dementia risk based on midlife (40–65 years), which was re-examined 21 years later, and with additional 7 years in total up to 30 years follow-up, vascular risk factors using a large population-based cohort from Finland [276]. This risk score is based on multifactorial risk estimation using age, hypertension, hypercholesterolemia, physical inactivity, obesity, and educational level as the model parameters. Later, the score was modified to place greater weight on age and education and to include carrier status for the APOE $\epsilon 4$. The study assigned risk scores for each factor with the

respective β coefficients from the final logistic regression models for dementia risk in 20 years. The risk score for an individual was obtained by summing the scores for the appropriate level of each of the risk factors. The risk score was categorised into five groups (based on quintiles; lower than the normal range 0–7; normal range 8–9; higher than average 10–15). The study also calculated the risk of dementia using the dementia risk score (modelled as a continuous variable) with the formula $P(\text{dementia}) = \frac{e^{(\beta_0 + \beta_1 + \beta_2[\text{score}])}}{1 + e^{(\beta_0 + \beta_1 + \beta_2[\text{score}])}}$, where β_0 is the intercept, β_1 is the coefficient for the follow-up time of 20 years, and β_2 represents the regression coefficients of the risk score [277].

The CAIDE Dementia Risk Score was externally validated in a multi-ethnic population in the United States [278] using the retrospective cohort study; a total of 9,480 Kaiser Permanente members (aged 40–55 years) from 1964 to 1973 with average follow-up 36.1 years. The dementia risk per sum score was calculated with Kaplan-Meier estimates. A total of 2,767 participants (25%) were diagnosed with any type of dementia. The study (Exalto *et al.*, 2014) reported that the CAIDE score also predicted well within different race strata, and other midlife risk factors (central obesity, depressed mood, diabetes mellitus, head trauma, lung function, and smoking) did not improve predictability. The risk score allowed stratification of participants into those with a 40-year low (9%) and high (29%) dementia risk [278]. Another study tested the predictive capacity of the CAIDE dementia risk score among a cohort of 3,582 Japanese American men in the Honolulu Heart Program (1965–1968, average age = 53.1 years; follow-up exam in 1991 (average age = 77.8 years), the CAIDE dementia risk score demonstrates significant association with later life severe cognitive impairment (OR = 1.48, 95% CIs: 1.39–1.58). However, its predictive capacity in this population is weak (i.e. the sensitivity is low, 0.50) [279].

Kivipelto *et al.* (2006) stressed that the CAIDE score is applicable to the prediction of dementia risk in 20 years only among middle-aged people who survive for the next 20 years. The survival bias could have affected the score, which would underestimate the effect of those risk factors (from who died with dementia before the late-life examination) in the total population [277].

There has been a need to evaluate conventional and novel risk factors for prediction of dementia. In the next section, I discuss an approach that might fill the gaps and improve the prediction of conventional risk tools.

2. Development of a novel risk assessment tool of brain health

The cognition risk scores (%) on the cardio-metabolic risk can be estimated with the discovered information (e.g. physiological markers) (Figure 76) using longitudinal, population-based data. In the present section, the importance of applying novel risk factors (which were not applied in the conventional risk assessment tools) will be discussed.

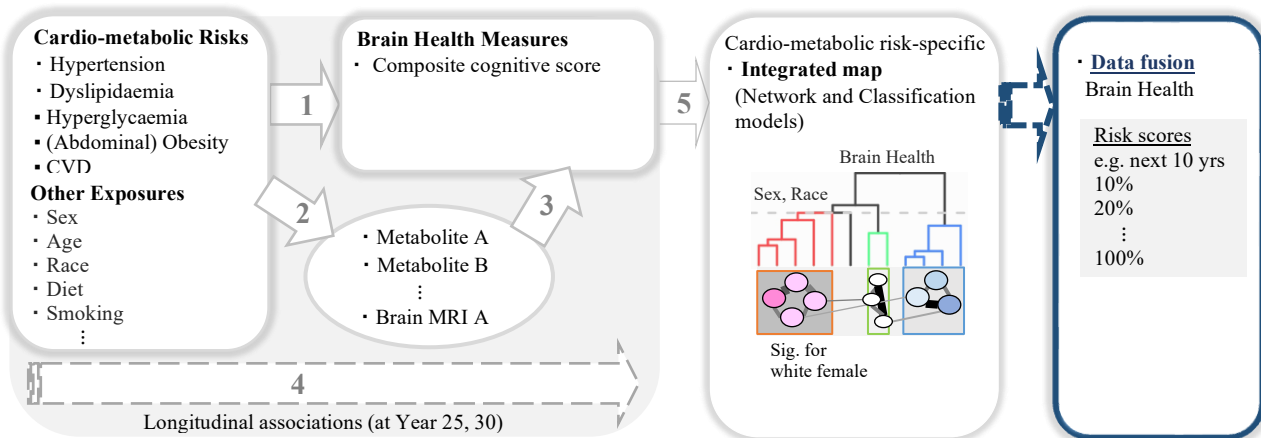


Figure 76 Schematic of the study highlighting the work focusses on the development of a risk assessment tool of brain health (i.e. cognition or dementia)

2-1. Identifying people who are liable to develop dementia in later life

MCI is defined as a clinical stage of the symptomatic and pre-dementia phase of the trajectory of cognitive decline (by the National Institute on Aging and Alzheimer's Association (NIA-AA), 2011) [280].

Kang (2018) reported a sharp increase in examinations of *Dementia symptoms* (which include mini-mental state examinations, the six item cognitive impairment test, referrals to memory clinics and behaviour assessments) since 2013 within UK primary care databases, The Health Improvement Network (THIN) (1995–2015) [87]. Usage of the Six-Item Cognitive Impairment Test (6CIT) and the General Practitioner Assessment of Cognition (GPCOG which was introduced in 2002) increased sharply from 2013. Memory assessments increased in 2013, and the elderly at risk rating scale, the 6CIT, and referrals to AF clinics showed the highest frequencies over time [87]. The results highlight the importance of early detection and risk management of cognitive impairment. Patients were very likely to have more than one diagnosis over time even within a short period time (e.g. a year). It was found that usage of the *Other diagnoses* records (which include dementia annual review, dementia monitoring, medication review, etc.) in addition to the Quality and Outcomes Framework (QOF) defined codes, and *Prescribed dementia medications* would at least ensure that a number of people with dementia who are not officially diagnosed would not be missed. In addition, the codes related to *Dementia symptoms* seem to cover a broad definition of dementia or pre-existing dementia population in the UK primary care records [87]. The

information on whether an individual had the related records in primary care before later-life seems to be significant as a risk factor in terms of the development of a risk assessment tool for dementia in later-life.

In addition to the limitations of risk scoring itself, several barriers to implementing the existing risk assessments in clinical practice have been reported by health professionals, i.e. primary care physicians. Clear guidelines for the prediction rules (which does not lead to medical over/under-treatment) and using risk information derived from multiple risk factors (not just a single risk value) would be helpful [281].

2-2. Metabolic phenotyping and its application to stratified medicine

The metabolic phenotype represents an integrated chemical profile at the molecular level derived from environmental, lifestyle, genotype and endogenous processes [147]. There have been many reports of direct applications of metabolic phenotyping in a clinical setting which support personalized medicine by providing a comprehensive measure of bio-sample chemistry for in-depth characterization of the metabolic phenotype [147, 151]. Holmes *et al.* (2015) highlighted that research into population level phenotyping to inform health care, use of clinical-image-based diagnostics and journey modelling were particularly important to enrich stratified medicine research [148, 151].

Current practice assumes that specific epidemiological information, metabolic and molecular information are minor pieces of information which are simply neglected in the management of adult brain health; there are no specific guidelines. Evaluating novel exposures which have not been used in the conventional risk tools will be as important as previously identified risk factors. The risk factors incorporated in the models could be different in those specific models. The important exposures from the novel approaches (e.g. scores using new knowledge of metabolomics) could enhance a brain health-related risk (i.e. cognition or dementia) prediction tool for stratified or targeted cohorts in primary care and public health.

Omics data fusion and decision level fusion

The combined analysis of multiple data sets is referred to as ‘data fusion or data integration’. Information from the unsupervised network approaches (e.g. a Bayesian method and SNF) can be integrated for metabolic data. Decision level fusion is one of the main strategies of data fusion; this requires constructing separate prognostic models for each data set (e.g. men and woman on cognitive impairment) and integrating the results [282, 283].

2-3. Development of a novel risk assessment tool of cardio-metabolic risk-specific brain health

The cognition risk scores (%) based on cardio-metabolic risk can be estimated with the discovered information using longitudinal, population-based data. Applying the issue of whether an individual had the related records (including cardio-metabolic risk factors and mediators) before later-life seems to contribute as a risk factor concerning a risk assessment tool for dementia in later-life. The information from the network-based approaches

can be integrated. A decision level data fusion constructs separate prognostic models for each data set (e.g. a cardio-metabolic dysfunction category, black and white, or men and women) and integrates results. The CARDIA Study, the prospective cohort study followed since 1985–86, measured cognitive and brain MRI measurements since 2010–11 (Year 25; participants aged 43–55 years). The present study accessed the data at Year 25 and Year 30, further follow-up of the cohort will be a great resource to develop a risk assessment tool for dementia with each 5 years of follow-up. The information from the knowledge of metabolomics could enhance a brain health-related risk prediction tool for stratified or targeted cohorts.

Table 42 adopted an international model to predict recurrent CVD, and it demonstrates a score sheet to estimate the risk of the dementia over years of follow-up [224]. The patient information for demographic, cardio-metabolic, history related to cognition, and newly discovered metabolites and brain MRI parameters can be matched with the status for each variable, and the assigned number of points are then summed. The total number of points is then matched with the estimated risk. The prediction model can be validated in other external cohorts to be established.

Table 42 An example of the score sheet to estimate the risk of dementia over years of follow-up using multiple regression-estimating equations, based on the recognised prognostic factors in the study

	Factor	Event: factors and points			Dementia event points
Demographic	Age	< 60	60–64	≥ 65	
	Race	Black	White		
	Education	10–14 years	15–20 years		
	Family income				
	Sex	Male	Female		
	Depression (CES-D)	< 16	≥ 16		
	Smoking	Never	Former	Current	
Cardio-metabolic	Fasting glucose	< 91	≥ 91		
	Waist circumference	< 88.6	≥ 88.6		
	:				
History related to cognition	Cognitive tests	No	Yes		
	Referral to memory clinic	No	Yes		
	Prescribed dementia medications	No	Yes		
Metabolites	<i>Standardized</i> Glycine	< 1.503	≥ 1.503		
	<i>Standardized</i> Methionine				
	<i>Standardized</i> Kynurenine				
	<i>Standardized</i> Tryptophan				
	<i>Standardized</i> Amino adipic acid				
	:				
Brain MRI parameters	<i>Standardized</i> FA right MFC	< -1.739	≥ -1.739		
	<i>Standardized</i> VR TL WM right				
	<i>Standardized</i> AV TL WM right				
	<i>Standardized</i> AV right ACG				
	<i>Standardized</i> ATV frontal				

* The sheet can be used by assigning points for each factor, summing the points, and determining the percentage risk of either the dementia event [284].

In terms of developing a tool, the following investigations are recommended: whether a cardio-metabolic risk-specific risk assessment tool will improve the prevention and management of brain health in comparison to a global risk assessment tool; whether the cognition risk scores (%) on a cardio-metabolic risk can be improved with **novel markers** (e.g. medical history, physiological markers); and whether the detailed investigation of **confounders** and **mediators** of the association will improve the prevention and management of brain health. This can be examined using longitudinal, population-based large cohort data, with external data for validation.

Advanced and flexible risk prediction algorithms (such as discrete/continuous-time hazard models, Bayesian methods, or multivariable fractional polynomial models), and advanced data mining techniques (such as Bayesian classification and unsupervised networks) can be employed for integration and classification of the population-based records and identified risk factors.

Previously identified risk factors should be evaluated to ascertain if they are predictive indicators. In addition, novel risk factors not included in previous models, but potentially associated with the risks, can be examined. Internal and external validity can be performed on the models, describing performance in terms of discrimination (e.g. concordance index (C-index) and Cohen's d statistic (D-statistic)), calibration (e.g. Hosmer-Lemeshow statistic), and measures of fit combining calibration and discrimination (e.g. R^2 and the Brier score).

Chapter VII. Discussion

1. Main findings

The primary aim of this study was to investigate whether targeted urinary metabolites and brain MRI parameters are prognostic markers (i.e. mediators) regarding the cognitive outcome with each cardio-metabolic risk factor measured at Year 30 in the CARDIA Study.

The evidence of CVD-induced cognitive impairment or dementia were reviewed in Chapter I, Section 3. In addition, other cardio-metabolic risk factors (hypertension, hyperglycaemia, and obesity) and other exposures influencing cognitive impairment or dementia have been reported. Other exposures in the study include lifestyle exposures (sleep quality, family income, alcohol consumption, physical activity, diet quality, depression), socio-demographic exposures (age, race, sex, education, smoking) and clinical exposures (medication for hypertension and depression, and family history, e.g. immediate family member diagnosed with dementia, serum and urinary creatinine).

In addition to these exogenous variables, the values of which are determined outside the system, it is important to investigate endogenous variables (factors in a causal system whose value are determined by the states of other variables in the system) with respect to risk prediction of cognition or dementia. In the present study, I investigated the metabolic phenotype and the brain MRI parameters underlying cognitive impairment/reserve.

The total path diagram of metabolites

I investigated the associations regarding the adjusted composite cognitive score at Year 30 using NMR-and LC–MS-based PQN scaling urinary metabolites, adjusting for significant exposures. In the total multivariate linear SEM to investigate the urinary metabolites by NMR regarding the cognitive outcome, **valine** showed a positive association on WC, and **glycine** showed a negative association on fasting glucose ($p < 0.05$, Table 21). Furthermore, **alanine** showed positive associations with fasting insulin in the model which used it instead of fasting glucose.

In this total path diagram for amino acids, age, race, sex, education (up to Year 30), and family income were significant ($p < 0.05$) regarding the adjusted composite cognitive score. Sex, or race (fitted with education) were confounding factors, respectively, in the association between fasting glucose and cognition. In the model with the amino acids assayed by LC–MS, methionine showed a positive association with the mean SBP; **glycine** showed a negative association with fasting glucose; **isoleucine**, **leucine**, **phenylalanine**, **valine**, and **aminoadipic acid** showed positive associations with fasting glucose; serine showed a negative association with WC (Table 23).

In the model of tryptophan metabolites measured by LC–MS, **indole-3-acetic acid** showed a negative association with fasting glucose; **tryptophan**, and **kynurenine** which is higher in major depression showed a negative association with fasting glucose, respectively (Table 25).

The total path diagram of the brain MRI parameters

I used the adjusted composite cognitive score measured at Year 30 for the examinations of midlife vascular risk factors and the brain health associations in the study.

Race, sex and education up to Year 30 were significant regarding the adjusted composite cognitive score. In the SEM to investigate the brain MRI-derived parameters regarding the cognitive outcome, there was a negative association with each standardized **FA in WM**, **FA in temporal lobe WM right and left** on SBP; the standardized **total WM volume** showed a positive association with fasting glucose; each standardized **CBF in left ACG** and **CBF in left entorhinal area** showed negative associations with WC. The standardized **total right entorhinal area volume** was positively associated on the 10-Year Framingham risk score (Appendix Table 10).

The total path diagram of the metabolites and the brain MRI parameters

I included the statistically significant associations in the urinary metabolites and the brain MRI parameters from each mediation model simultaneously in the total path diagram (Figure 49).

Phenylalanine, **aminoadipic acid**, and **tryptophan** showed significant positive associations, respectively, with fasting glucose in terms of hyperglycaemia; **indole-3-acetic acid** showed a significant negative association with fasting glucose. This information is consistent with the outcomes before applying the brain MRI parameters in the model. **WM CBF in temporal lobe WM left**, and **CBF in left entorhinal area** showed a significant negative association with WC, respectively, in terms of (abdominal) obesity (N=74); the association is in line with literature that detrimental effects of adiposity on brain blood flow. **Temporal lobes** and **entorhinal area** were investigated as early markers of AD in the study [111, 144]. This outcome is supported by the review that obesity triggering VaD increases fat cells that damage the brain WM leading to loss of cognitive and intellectual capacity [46].

The degeneration of cognition with the brain MRI parameters adjusting for a cardio-metabolic factor

I employed the Bayesian multilevel models to examine the degeneration of cognitive outcome on each of the brain MRI parameter fitted with a cardio-metabolic factor measured at Year 25 (2010–11) and Year 30 (2015–16) (Figure 57 to Figure 60).

In the analysis of the standardized brain MRI measurements with **SBP** regarding cognition in each Bayesian multilevel regression); the brain MRI parameters measured **abnormal tissue volumes** were negatively associated regarding cognition (with small effect sizes); **FA in deep WM GM** and the **number of activated voxels with right MFC and ACG** were positively associated with cognition. Significant exposures (i.e. race, sex, education, physical activity, smoking, CES–D, HBP medication, family income, urinary creatinine and ICV) were adjusted for in the model.

In the analysis of the standardized brain MRI measurements with **fasting insulin** regarding composite cognition in each Bayesian multilevel regression, the brain MRI parameters positively associated regarding cognition (with small effect sizes) when they fitted with fasting insulin are as follows: **percent change of vascular reactivity in temporal lobe WM right and left; number of activated voxels within right and left hippocampus, MFC, right ACG and temporal lobe WM right**. Significant exposures (i.e. race, sex, education, physical activity, smoking, CES–D, HBP medication, family income, urinary creatinine and ICV) were adjusted for in the model.

In the analysis of the standardized brain MRI measurements with WC regarding cognition in each Bayesian multilevel regression, the brain MRI parameters positively associated regarding cognition (with small effect sizes) when they fitted with WC are as follows: **percent change of temporal lobe WM right and left; number of activated voxels within right and left MFC and right ACG**. Significant exposures (i.e. race, sex, education, smoking, CES–D, HBP medication, family income, urinary creatinine and ICV) were adjusted for in the model.

In the analysis of the standardized brain MRI measurements with BMI regarding cognition in each Bayesian multilevel regression the brain MRI parameters positively associated regarding cognition (with small effect sizes) when they fitted with BMI are as follows: **total volume of brain, total volume of (left) GM and cerebrospinal fluid; and percent change of vascular reactivity in temporal lobe WM right and left; number of activated voxels within right and left MFC**. Significant exposures (i.e. race, sex, education, age, physical activity, smoking, CES–D, HBP medication, family income, urinary creatinine and ICV) were adjusted for in the model.

To provide a cardio-metabolic risk-specific integrated map

To provide a cardio-metabolic risk-specific integrated overview (i.e. profile of the markers) regarding cognition with significant epidemiological exposures, metabolites, and brain MRI parameters (identified from Objectives 1 to 4), I employed **network** and **classification models**. I used the methods to further investigate brain health in middle life on metabolic phenotypes and brain MRI parameters with significant exposures for cardio-metabolic risk-specific information (i.e. integrated map, [5] Figure 7).

I integrated all the examined information including different types of significant clinical, socio-demographic, lifestyle risk factors, metabolite and brain MRI parameter markers to provide an integrated overview relating to

cardio-metabolic risk-specific brain health. I employed tree-based classification models and principal component models to investigate cardio-metabolic risk-specific metabolic phenotypes and brain MRI parameters.

The cardio-metabolic risk-specific (SBP, WC, and fasting glucose) integrated information regarding the cognitive outcome was investigated. Among the significant factors identified from the previous chapters, key epidemiological exposures (race, education, family income, sex, age, CES–D, smoking), metabolites (glycine, methionine, aminoadipic acid, tryptophan, and kynurenine) and brain MRI parameters were identified as prognostic markers regarding cognition; in particular **methionine** concentrations were higher in the black ethnic group. The number of activated voxels within temporal lobe WM right, the number of activated voxels within right ACG and FA in right MFC showed **negative** associations with a cardio-metabolic risk factor (SBP, fasting insulin, WC, and BMI) and **positive associations regarding cognition**, respectively. These brain MRI-parameters are significantly lower in cognitively impaired individuals, and the AV TL WM right and AV right ACG are significantly lower in the black ethnic group. In addition, percent change of vascular reactivity in temporal lobe WM right, and abnormal frontal tissue volume showed **positive** associations with a cardio-metabolic risk factor (SBP, fasting insulin, WC, and BMI) where abnormal tissue volume in frontal showed a **negative association regarding cognition** and VR TL WM right showed a **positive association regarding cognition** when fasting insulin, WC and BMI were adjusted in each Bayesian multilevel regression.

1-1. Discussion and suggestions for future research

Metabolic pathways for the gut-brain axis or changes in metabolites associated with brain inflammation

Further detailed studies of metabolites of the tryptophan and/or kynurenine pathways related to the gut-brain axis or changes in metabolites associated with brain inflammation need to be undertaken. There are a number of potential mechanisms that underlie the associations between obesity and the risk of cognitive impairment include the gut-brain axis, and central inflammation processes [285]. Solas M *et al.* (2017) emphasized that gut microbiota could be part of a mechanism linking the consumption of high fat and other unbalanced diets to impaired cognition (known as the 'gut-brain axis') [285]. Clinical studies have reported positive associations in obese individuals between peripheral inflammatory status (a kind of protective response status of the body) and cognitive decline or mood symptoms [286]. Solas M *et al.* (2017) underlined that the inflammation becomes activated in metabolic disease which can cause it to spread from the peripheral tissue to the brain (disrupting synaptic plasticity leading to neurodegeneration and eventually brain atrophy). This inflammatory activation can be commonly triggered by diet-related changes (such as consumption of high fats and sugars) via alterations in the gut microbiota composition [285].

In the review by Tim L *et al.* (2014), a decade (2003–2013) of epidemiologic observational studies investigating the associations between mid-life BMI and AD were highlighted. Studies investigating the association between

midlife BMI and the risk of dementia generally demonstrated an increased risk among overweight and obese adults. The author state that there are potential underlying mechanisms accompanying a high BMI at midlife that may influence dementia/AD due to increased inflammation or higher cytokine and hormone levels produced by adipose tissue [287-289]. Serum proinflammatory cytokines are known to be risk factors for AD [290].

In terms of the analytic methods, network-based methods can be employed to discover modules and symbolic representations of disease-associated mechanisms. Moreover, unsupervised network-based methods detect significant factors within pathways, discover disease sub-clusters, or find co-expression network modules [291]. Significant associations were found in the present study; however further research is needed. Metabolic pathways regarding cardio-metabolic risk-specific associations with cognition should be further examined. It will enable an examination of cardio-metabolic risk differences in metabolites (as phenotypic traits) by a specific risk factor which is eventually highly associated with/or from the result of metabolism functions, or physiological reactions.

Metabolic group separation and network-based modelling

Metabolic fingerprints (i.e. finding the compounds most responsible for discriminating samples) associated with group separation can be employed using several unsupervised clustering methodologies such as Bayesian hierarchical cluster analysis for targeted metabolomics data in favour of flexibility. For high dimensional data with variable selection, Bayesian agglomerative clustering, which includes a preliminary response filtering step in a hierarchical structure, or the non-negative matrix factorization (NMF) can be employed. NMF generates non-orthogonal and non-negative factors (Zhang *et al.*, 2011, 2012) which is an appropriate method when high-dimensional data contains a large number of irrelevant or redundant information (Li *et al.*, 2013). Unsupervised network approaches can be employed such as a flexible and robust Bayesian method which is based on assumptions of different types of data sets with various distributions and correlations among the data sets. The multiple dataset integration (MDI) method uses the Dirichlet-multinomial allocation (DMA) mixture model and cluster metabolites across multiple data sets under the assumption that they are measured in all different levels [291, 292]. In addition, each cardio-metabolic risk-specific difference in metabolic data can be analysed using a metabolic association network approach, examining the risk factor associations with the clustered network [293]. The SNF is a method of constructing networks of individuals for each available data type and then efficiently fusing these into one network that represents the full spectrum of underlying data; it is also an effective approach to perform risk prediction [294].

1-2. Strengths and limitations

Strengths

A key strength of this study is the use of SEM in which the indirect and total effects can be estimated from a single model, and the simple mediation model can be inserted in a larger model (even using latent variables) to measure any piece of the mediation model ([1-3] Figure 7) [219, 220]. The study adjusted other exposures (lifestyle, socio-demographic and clinical exposures) that might influence the exposures and outcomes. The study employed **robust analytic methods** such as Bayesian multilevel regression which is advantageous in terms of the measurement errors such as mis-measured continuous exposures, especially for the brain MRI parameters. The Bayesian approach is known to be useful especially in the case of multiplicity situations as the method avoids some of the conceptual and practical difficulties involved with the control of type I error [242]. The multiplicity problem occurs when one infers a subset of parameters selected based on the observed values (M Shields, 2019). I used systematic multiple-analytic processes (i.e. a combination of the univariate regressions to the multivariate regressions). The test hypothesis is only one for each association in the univariate regression. In the total multiple regression model, each different parameter could support each hypothesis, but those were from the univariate regression initially, and I used the bootstrapped standard errors and confidence intervals for conservative estimates of significance. In addition, Bayesian methods are advantageous compared to classical path analysis in that the former add the concept of latent variables, allowing the capture of communal explanations across highly interrelated variables and in this way control for some of the measured biases as well as interrelations. In this thesis, I investigated the brain MRI parameters using the Bayesian two-level models measured over time considering cardio-metabolic risk factor-by-year interactions.

A further strength of the present study was that it employed multi-stage examinations to examine the study objectives using various statistical approaches (including network and classification models) to verify the outcomes, and in addition to using of multiple study designs, i.e. both cross-sectional and longitudinal designs (Year 25 and Year 30), to integrate the findings.

Limitations

There is a number of limitations in the study design and assay data. The CARDIA Study is a longitudinal prospective design, and the present study examined the urinary metabolites focusing on the sub-cohort data measured at Year 30. The internal validity of metabolite findings, and development of a population-based risk assessment tool were not available.

The path diagram using the SEM examined one direction (i.e. variables do not affect each other). This casualty issue in path diagrams can be improved using randomization and removing endogenous variables with the clear correlation (i.e. $X \rightarrow Y$). There are almost certainly unmeasured confounding factors present since this study is not an RCT. For instance, the CVD risk factor high-sensitivity C-reactive protein (CRP) and genetic variables

(such as plasma clusterin and APOE ϵ 4) which would be potential confounding factors for the associations were not available in the sub-cohort data. Nevertheless, the study adjusted many other significant exposures. In addition, the data is not free from measurement error issues, and the therefore robust analytic methods were applied. Finally, the finding has not yet validated with the external data (or other studies from the literature).

2. Conclusions

Prior to this study, a small effect size of CVD-induced cognitive impairment or dementia was reported in the literature where other exposures (lifestyle, socio-demographic and clinical exposures) were adjusted. In addition, midlife obesity has been associated with an increased risk of dementia. Nevertheless, the underlying mechanisms are poorly understood. There is an acute need to investigate clinical prognostic markers (i.e. mediators) of these associations. Therefore, the present study examined whether urinary metabolites and brain MRI parameters are mediators regarding cognition on each cardio-metabolic risk factor using the CARDIA Study measured at Year 25 and Year 30. In the study, the metabolic phenotype and the brain MRI parameters underlying cognitive impairment/reserve were investigated.

A fusion of bio-statistical analytic methods such as a path analysis, Bayesian multilevel models, network and classification analyses were employed in the study. A number of significant metabolites and brain MRI parameters were discovered including phenylalanine, aminoadipic acid, and tryptophan which showed positive associations with fasting glucose in terms of hyperglycaemia, while indole-3-acetic acid showed a negative association. The measures of brain physiology, i.e. CBF in temporal lobe WM left, and CBF in left entorhinal area showed negative associations with WC, respectively, in terms of (abdominal) obesity. For cognitively impaired individuals, metabolites (glycine, methionine, aminoadipic acid, tryptophan and kynurenine) and brain MRI-derived parameters (the number of activated voxels within temporal lobe, right ACG, and FA in right MFC) were identified as physiological markers.

The study findings need to be interpreted carefully as the study findings are based on the CARDIA cohort at two-time points, and ideally the findings need to be verified by using additional follow-up data. Future investigations may be extended to research metabolic pathways for the gut-brain axis and/or changes in metabolite associated with brain inflammation. Newly discovered prognostic factors could be used in the development of risk assessment tools for brain health. Current usual practice assumes specific clinical, metabolic, or genetic factors to be minor information, and they are simply neglected in the management of adult-CVD-related risk, or brain health. Nevertheless, this study has shown that these factors are significant mediators in the prognostic link between known cardio-metabolic risk factors and cognition. The study also showed how discovered prognostic makers can be applied in the management of adult brain health. These advances may lead, in the long term, to improved prediction of cognition, and therefore significant advances in brain health of the general population.

References

1. Tromp, D., *DTI Scalars (FA, MD, AD, RD) - How do they relate to brain structure?* The Winnower, 2016.
2. Sotak, C.H., *The role of diffusion tensor imaging in the evaluation of ischemic brain injury - a review.* NMR Biomed, 2002. **15**(7-8): p. 561-9.
3. Moseley, M., R. Bammer, and J. Illes, *Diffusion-tensor imaging of cognitive performance.* Brain and Cognition, 2002. **50**(3): p. 396-413.
4. National Heart, L., and Blood Institute, *Third Report of the National Cholesterol Education Program (NCEP) Expert Panel on Detection, Evaluation, and Treatment of High Blood Cholesterol in Adults (Adult Treatment Panel III) final report.* Circulation., 2002. **106**(25): p. 3143-421.
5. Carnethon, M.R., et al., *Risk factors for the metabolic syndrome: the Coronary Artery Risk Development in Young Adults (CARDIA) study, 1985-2001.* Diabetes Care, 2004. **27**(11): p. 2707-15.
6. Shin, D., K. Kongpakpaisarn, and C. Bohra, *Trends in the prevalence of metabolic syndrome and its components in the United States 2007–2014.* International Journal of Cardiology, 2018. **259**: p. 216-219.
7. DiBello, J.R., et al., *Prevalence of metabolic syndrome and its components among men with and without clinical benign prostatic hyperplasia: a large, cross-sectional, UK epidemiological study.* BJU International, 2016. **117**(5): p. 801-808.
8. Marangos, P.J., L.J. Okamoto, and J.J. Caro, *Economic Burden of the Components of the Metabolic Syndrome*, in *Handbook of Disease Burdens and Quality of Life Measures*, V.R. Preedy and R.R. Watson, Editors. 2010, Springer New York: New York, NY. p. 1135-1149.
9. Excellence, T.N.I.f.H.a.C., *Hypertension in adults: diagnosis and management*, in *NICE guideline*. 2019: The National Institute for Health and Care Excellence.
10. Manolis, A.J., et al., *Hypertension and atrial fibrillation: diagnostic approach, prevention and treatment. Position paper of the Working Group 'Hypertension Arrhythmias and Thrombosis' of the European Society of Hypertension.* J Hypertens, 2012. **30**(2): p. 239-52.
11. Hahn, C. and M.A. Schwartz, *Mechanotransduction in vascular physiology and atherogenesis.* Nat Rev Mol Cell Biol, 2009. **10**(1): p. 53-62.
12. Pires, P.W., et al., *The effects of hypertension on the cerebral circulation.* American Journal of Physiology-Heart and Circulatory Physiology, 2013. **304**(12): p. H1598-H1614.
13. Guyton, A.C., *Blood pressure control--special role of the kidneys and body fluids.* Science, 1991. **252**(5014): p. 1813-1816.
14. Ram, C.V.S., *Hypertensive emergencies: recognition and management.* Current problems in cardiology, 1982. **7**(1): p. 4-70.
15. Moretti, R., et al., *Risk factors for vascular dementia: hypotension as a key point.* Vasc Health Risk Manag, 2008. **4**(2): p. 395-402.
16. Poblador-Plou, B., et al., *Comorbidity of dementia: a cross-sectional study of primary care older patients.* BMC psychiatry, 2014. **14**(1): p. 84.
17. Florez, H., et al., *Prevalence and risk factors associated with the metabolic syndrome and dyslipidemia in White, Black, Amerindian and Mixed Hispanics in Zulia State, Venezuela.* Diabetes Res Clin Pract, 2005. **69**(1): p. 63-77.
18. Vallejo-Vaz Antonio, J., et al., *Triglyceride-Rich Lipoprotein Cholesterol and Risk of Cardiovascular Events Among Patients Receiving Statin Therapy in the TNT Trial.* Circulation, 2018. **138**(8): p. 770-781.
19. Yang, R.-L., et al., *Increasing Oxidative Stress with Progressive Hyperlipidemia in Human: Relation between Malondialdehyde and Atherogenic Index.* Journal of Clinical Biochemistry and Nutrition, 2008. **43**(3): p. 154-158.
20. Ceriello, A., et al., *Evidence for an independent and cumulative effect of postprandial hypertriglyceridemia and hyperglycemia on endothelial dysfunction and oxidative stress generation: effects of short- and long-term simvastatin treatment.* Circulation, 2002. **106**(10): p. 1211-8.
21. Organization, W.H., *Waist Circumference and Waist–Hip Ratio: Report of a WHO Expert Consultation.* 2008: the World Health Organization.

22. Goto, A., et al., *Severe hypoglycaemia and cardiovascular disease: systematic review and meta-analysis with bias analysis*. BMJ : British Medical Journal, 2013. **347**.
23. Association, A.D., *Standards of Medical Care in Diabetes — 2020*. Journal of clinical and applied research and education, 2020. **43**.
24. Tomlinson, S., et al., *Mechanisms of Disease : An Introduction to Clinical Science*. 2008, Cambridge, UNITED KINGDOM: Cambridge University Press.
25. Organization, W.H. *Controlling the global obesity epidemic*. 2018; Available from: <http://www.who.int/nutrition/topics/obesity/en/>.
26. Organization, W.H. *Obesity and overweight*. 2018; Available from: <http://www.who.int/mediacentre/factsheets/fs311/en/>.
27. England, P.H., *Health matters: obesity and the food environment*, T.D.o.H.a.S. Care, Editor. 2017.
28. Hales, C.M., et al., *Prevalence of Obesity Among Adults and Youth: United States, 2015-2016*. NCHS Data Brief, 2017(288): p. 1-8.
29. Ulijaszek, S., *Obesity: Preventing and Managing the Global Epidemic. Report of a WHO Consultation. WHO Technical Report Series 894. Pp. 252. (World Health Organization, Geneva, 2000.) SFr 56.00, ISBN 92-4-120894-5, paperback*. Vol. 35. 2003. 624-625.
30. Salamat, M.S., *Robbins and Cotran: Pathologic Basis of Disease, 8th Edition*. Journal of Neuropathology & Experimental Neurology, 2010. **69**(2): p. 214-214.
31. Ritchie, S.A. and J.M. Connell, *The link between abdominal obesity, metabolic syndrome and cardiovascular disease*. Nutr Metab Cardiovasc Dis., 2007. **17**(4): p. 319-26. doi: 10.1016/j.numecd.2006.07.005. Epub 2006 Nov 15.
32. Wilson, P.W.F., et al., *Prediction of Coronary Heart Disease Using Risk Factor Categories*. Circulation, 1998. **97**(18): p. 1837.
33. Armstrong, A.C., et al., *Framingham score and LV mass predict events in young adults: CARDIA study*. Int J Cardiol, 2014. **172**(2): p. 350-5.
34. Debette, S. and H.S. Markus, *The clinical importance of white matter hyperintensities on brain magnetic resonance imaging: systematic review and meta-analysis*. BMJ, 2010. **341**: p. c3666.
35. Olsson, E., et al., *White Matter Lesion Assessment in Patients with Cognitive Impairment and Healthy Controls: Reliability Comparisons between Visual Rating, a Manual, and an Automatic Volumetrical MRI Method-The Gothenburg MCI Study*. J Aging Res, 2013. **2013**: p. 198471.
36. Yano, Y., et al., *Long-term blood pressure variability throughout young adulthood and cognitive function in midlife: the Coronary Artery Risk Development in Young Adults (CARDIA) study*. Hypertension, 2014. **64**(5): p. 983-8.
37. Abell, J.G., et al., *Association between systolic blood pressure and dementia in the Whitehall II cohort study: role of age, duration, and threshold used to define hypertension*. Eur Heart J, 2018. **39**(33): p. 3119-3125.
38. Notkola, I.L., et al., *Serum total cholesterol, apolipoprotein E epsilon 4 allele, and Alzheimer's disease*. Neuroepidemiology, 1998. **17**(1): p. 14-20.
39. Kivipelto, M., et al., *Midlife vascular risk factors and Alzheimer's disease in later life: longitudinal, population based study*. Bmj, 2001. **322**(7300): p. 1447-51.
40. Mielke, M.M., et al., *The 32-year relationship between cholesterol and dementia from midlife to late life*. Neurology, 2010. **75**(21): p. 1888-1895.
41. Anstey, K.J., K. Ashby-Mitchell, and R. Peters, *Updating the Evidence on the Association between Serum Cholesterol and Risk of Late-Life Dementia: Review and Meta-Analysis*. J Alzheimers Dis, 2017. **56**(1): p. 215-228.
42. Cox, D.J., et al., *Relationships between hyperglycemia and cognitive performance among adults with type 1 and type 2 diabetes*. Diabetes Care., 2005. **28**(1): p. 71-7.
43. Sommerfield, A.J., I.J. Deary, and B.M. Frier, *Acute hyperglycemia alters mood state and impairs cognitive performance in people with type 2 diabetes*. Diabetes Care, 2004. **27**(10): p. 2335-40.
44. Whitmer, R.A., et al., *Obesity in middle age and future risk of dementia: a 27 year longitudinal population based study*. BMJ, 2005. **330**(7504): p. 1360.
45. Dye, L., et al., *The relationship between obesity and cognitive health and decline*. Proceedings of the Nutrition Society, 2017. **76**(4): p. 443-454.
46. Anjum, I., et al., *Does Obesity Increase the Risk of Dementia: A Literature Review*. Cureus, 2018. **10**(5): p. e2660-e2660.

47. Clare, L., et al., *Potentially modifiable lifestyle factors, cognitive reserve, and cognitive function in later life: A cross-sectional study*. PLoS Med., 2017. **14**(3): p. e1002259. doi: 10.1371/journal.pmed.1002259. eCollection 2017 Mar.
48. NHS. *Atherosclerosis* 2016 2016; Available from: <https://www.nhs.uk/conditions/atherosclerosis/>.
49. NHS. *Coronary heart disease*. 2017 2017; Available from: <https://www.nhs.uk/conditions/coronary-heart-disease/>.
50. NHS. *Cardiovascular disease*. 2016; Available from: <https://www.nhs.uk/conditions/cardiovascular-disease/>.
51. Ikram, M.A., et al., *Unrecognized myocardial infarction in relation to risk of dementia and cerebral small vessel disease*. Stroke, 2008. **39**(5): p. 1421-6.
52. Vidal, J.S., et al., *Coronary artery calcium, brain function and structure: the AGES-Reykjavik Study*. Stroke, 2010. **41**(5): p. 891-7.
53. Vuorinen, M., et al., *Coronary Heart Disease and Cortical Thickness, Gray Matter and White Matter Lesion Volumes on MRI*. PLOS ONE, 2014. **9**(10): p. e109250.
54. Sujin Kang, A.M., Kosh Ray, *Kaplan-Meier failure estimates for non-acute cardiovascular disease*. 2018: Unpublished: Primary Care and Public Health, School of Public Health, Imperial College London.
55. Kivipelto, M., et al., *Apolipoprotein E epsilon4 allele, elevated midlife total cholesterol level, and high midlife systolic blood pressure are independent risk factors for late-life Alzheimer disease*. Ann Intern Med, 2002. **137**(3): p. 149-55.
56. Ross, G.W., et al., *Characterization of risk factors for vascular dementia: the Honolulu-Asia Aging Study*. Neurology, 1999. **53**(2): p. 337-43.
57. Kalmijn, S., et al., *Cerebrovascular disease, the apolipoprotein e4 allele, and cognitive decline in a community-based study of elderly men*. Stroke, 1996. **27**(12): p. 2230-5.
58. Kurella Tamura, M., et al., *Vascular risk factors and cognitive impairment in chronic kidney disease: the Chronic Renal Insufficiency Cohort (CRIC) study*. Clinical journal of the American Society of Nephrology : CJASN, 2011. **6**(2): p. 248-256.
59. Mast, B.T., et al., *Vascular disease and future risk of depressive symptomatology in older adults: findings from the Health, Aging, and Body Composition study*. Biol Psychiatry, 2008. **64**(4): p. 320-6.
60. Jacob, L., J. Bohlken, and K. Kostev, *Risk Factors for Mild Cognitive Impairment in German Primary Care Practices*. J Alzheimers Dis, 2017. **56**(1): p. 379-384.
61. Verhaegen, P., M. Borchelt, and J. Smith, *Relation between cardiovascular and metabolic disease and cognition in very old age: cross-sectional and longitudinal findings from the berlin aging study*. Health Psychol, 2003. **22**(6): p. 559-69.
62. Rostamian, S., et al., *Executive function, but not memory, associates with incident coronary heart disease and stroke*. Neurology, 2015. **85**(9): p. 783-9.
63. Vos, S.J.B., et al., *Modifiable Risk Factors for Prevention of Dementia in Midlife, Late Life and the Oldest-Old: Validation of the LIBRA Index*. J Alzheimers Dis, 2017. **58**(2): p. 537-547.
64. Haring, B., et al., *Cardiovascular disease and cognitive decline in postmenopausal women: results from the Women's Health Initiative Memory Study*. Journal of the American Heart Association, 2013. **2**(6): p. e000369-e000369.
65. Chinn, S., *A simple method for converting an odds ratio to effect size for use in meta-analysis*. Stat Med, 2000. **19**(22): p. 3127-31.
66. Hayden, K.M., et al., *Vascular Risk Factors for Incident Alzheimer Disease and Vascular Dementia: The Cache County Study*. Alzheimer Disease and Associated Disorders, 2006. **20**(2): p. 93-100.
67. Newman, A.B., et al., *Dementia and Alzheimer's disease incidence in relationship to cardiovascular disease in the Cardiovascular Health Study cohort*. J Am Geriatr Soc, 2005. **53**(7): p. 1101-7.
68. Rovio, S., et al., *Leisure-time physical activity at midlife and the risk of dementia and Alzheimer's disease*. Lancet Neurol, 2005. **4**(11): p. 705-11.
69. Liu, C.C., et al., *Apolipoprotein E and Alzheimer disease: risk, mechanisms and therapy*. Nat Rev Neurol, 2013. **9**(2): p. 106-18.
70. Sundbøll, J., et al., *Higher Risk of Vascular Dementia in Myocardial Infarction Survivors*. Circulation, 2018. **137**(6): p. 567-577.
71. Luchsinger, J.A., et al., *Aggregation of vascular risk factors and risk of incident Alzheimer disease*. Neurology, 2005. **65**(4): p. 545-51.

72. Eriksson, U.K., et al., *Nonstroke cardiovascular disease and risk of Alzheimer disease and dementia*. Alzheimer Dis Assoc Disord, 2010. **24**(3): p. 213-9.
73. Li, J., et al., *Vascular risk factors promote conversion from mild cognitive impairment to Alzheimer disease*. Neurology, 2011. **76**(17): p. 1485-91.
74. Petrovitch, H., et al., *Influence of myocardial infarction, coronary artery bypass surgery, and stroke on cognitive impairment in late life*. Am J Cardiol, 1998. **81**(8): p. 1017-21.
75. Malek, N., et al., *Vascular disease and vascular risk factors in relation to motor features and cognition in early Parkinson's disease*. Mov Disord, 2016. **31**(10): p. 1518-1526.
76. Louis, E.D. and J.A. Luchsinger, *History of Vascular Disease and Mild Parkinsonian Signs in Community-Dwelling Elderly Individuals*. Archives of Neurology, 2006. **63**(5): p. 717-722.
77. Zhong, G., et al., *Smoking Is Associated with an Increased Risk of Dementia: A Meta-Analysis of Prospective Cohort Studies with Investigation of Potential Effect Modifiers*. PLOS ONE, 2015. **10**(3): p. e0118333.
78. Rovio, S., et al., *Leisure-time physical activity at midlife and the risk of dementia and Alzheimer's disease*. The Lancet Neurology, 2005. **4**(11): p. 705-711.
79. Malm, T., et al., *Activation of the nuclear receptor PPAR δ is neuroprotective in a transgenic mouse model of Alzheimer's disease through inhibition of inflammation*. Journal of Neuroinflammation, 2015. **12**(1): p. 7.
80. Webers, A., M.T. Heneka, and P.A. Gleeson, *The role of innate immune responses and neuroinflammation in amyloid accumulation and progression of Alzheimer's disease*. Immunol Cell Biol, 2020. **98**(1): p. 28-41.
81. Cervellati, C., et al., *Oxidative balance, homocysteine, and uric acid levels in older patients with Late Onset Alzheimer's Disease or Vascular Dementia*. J Neurol Sci, 2014. **337**(1-2): p. 156-61.
82. Luca, M., A. Luca, and C. Calandra, *The Role of Oxidative Damage in the Pathogenesis and Progression of Alzheimer's Disease and Vascular Dementia*. Oxidative medicine and cellular longevity, 2015. **2015**: p. 504678-504678.
83. Gackowski, D., et al., *Oxidative stress and Oxidative DNA damage is characteristic for mixed Alzheimer disease/vascular dementia*. Journal of the neurological sciences, 2008. **266**: p. 57-62.
84. Raphael Wittenberg, B.H., Luis Barraza-Araiza, Amritpal Rehill *Projections of older people living with dementia and costs of dementia care in the United Kingdom, 2019-2040*, in CPEC Working Paper 5. 2019: London School of Economics and Political Science, Care Policy and Evaluation Centre.
85. Association, A.s., *2018 Alzheimer's disease facts and figures*. Alzheimer's & Dementia, 2018. **14**(3): p. 367-429.
86. Stern, Y., *An approach to studying the neural correlates of reserve*. Brain Imaging Behav, 2017. **11**(2): p. 410-416.
87. Kang, S., et al., *Identifying People with Dementia in Electronic Primary Care Records in the UK*. European Journal of Public Health, 2019. **29**(Supplement_4).
88. Center, C.C. *Coronary Artery risk Development in Young Adults Study (CARDIA)*. 2014; Available from: <https://www.cardia.dopm.uab.edu/>.
89. Baron, R.M. and D.A. Kenny, *The moderator-mediator variable distinction in social psychological research: conceptual, strategic, and statistical considerations*. Journal of personality and social psychology, 1986. **51** 6: p. 1173-82.
90. Fritz, M.S. and D.P. Mackinnon, *Required sample size to detect the mediated effect*. Psychological science, 2007. **18**(3): p. 233-239.
91. Nasreddine, Z.S., et al., *The Montreal Cognitive Assessment, MoCA: a brief screening tool for mild cognitive impairment*. J Am Geriatr Soc, 2005. **53**(4): p. 695-9.
92. Lewis, M.R., et al., *Development and Application of Ultra-Performance Liquid Chromatography-TOF MS for Precision Large Scale Urinary Metabolic Phenotyping*. Anal Chem, 2016. **88**(18): p. 9004-13.
93. Narayan, R., *Encyclopedia of biomedical engineering*. 2019.
94. Koshiyama, D., et al., *White matter microstructural alterations across four major psychiatric disorders: mega-analysis study in 2937 individuals*. Molecular Psychiatry, 2020. **25**(4): p. 883-895.
95. Parker, E.D., et al., *Physical activity in young adults and incident hypertension over 15 years of follow-up: the CARDIA study*. American journal of public health, 2007. **97**(4): p. 703-709.
96. Pereira, M.A., et al., *A collection of Physical Activity Questionnaires for health-related research*. Med Sci Sports Exerc, 1997. **29**(6 Suppl): p. S1-205.

97. Gooding, H.C., et al., *Application of a Lifestyle-Based Tool to Estimate Premature Cardiovascular Disease Events in Young Adults: The Coronary Artery Risk Development in Young Adults (CARDIA) Study*. JAMA Internal Medicine, 2017. **177**(9): p. 1354-1360.
98. Martha Sajatovic, L.F.R., *Rating Scales in Mental Health*. 3rd ed. ed. 2012: Baltimore, Md. : Johns Hopkins University Press, 2012.
99. Ware, J.E., S.D. Keller, and M. Kosinski, *SF-12 : how to score the SF-12 physical and mental health summary scales*. 1995.
100. D'Agostino, R.B., Sr., et al., *General cardiovascular risk profile for use in primary care: the Framingham Heart Study*. Circulation, 2008. **117**(6): p. 743-53.
101. Kautzky-Willer, A., J. Harreiter, and G. Pacini, *Sex and Gender Differences in Risk, Pathophysiology and Complications of Type 2 Diabetes Mellitus*. Endocr Rev, 2016. **37**(3): p. 278-316.
102. Jahangiry, L., M.A. Farhangi, and F. Rezaei, *Framingham risk score for estimation of 10-years of cardiovascular diseases risk in patients with metabolic syndrome*. J Health Popul Nutr, 2017. **36**(1): p. 36.
103. Muntner, P., et al., *Validation of the Atherosclerotic Cardiovascular Disease Pooled Cohort Risk Equations*. JAMA, 2014. **311**(14): p. 1406-1415.
104. Schmidt, M., *Rey Auditory Verbal Learning Test : RAVLT : a handbook*. 1996, Los Angeles, CA: Western Psychological Services.
105. Wechsler, D., *WAIS-R : manual : Wechsler adult intelligence scale-revised*. 1981: New York, NY : Harcourt Brace Jovanovich [for] Psychological Corp.
106. Lawton, G., *THE MEASUREMENT OF ADULT INTELLIGENCE**. American Journal of Orthopsychiatry, 1939. **9**(4): p. 808-812.
107. Stroop, J.R., *Studies of interference in serial verbal reactions*. 1935. **18**(6): p. 643.
108. Hoops, S., et al., *Validity of the MoCA and MMSE in the detection of MCI and dementia in Parkinson disease*. Neurology, 2009. **73**(21): p. 1738-1745.
109. Rosenzweig, A. *Montreal Cognitive Assessment (MoCA) Test for Dementia*. 2019; Available from: <https://www.verywellhealth.com/alzheimers-and-montreal-cognitive-assessment-moca-98617>.
110. Beier, M., et al., *Brief and cost-effective tool for assessing verbal learning in multiple sclerosis: Comparison of the Rey Auditory Verbal Learning Test (RAVLT) to the California Verbal Learning Test - II (CVLT-II)*. Journal of the neurological sciences, 2019. **400**: p. 104-109.
111. Haight, T., et al., *Associations of plasma clusterin and Alzheimer's disease-related MRI markers in adults at mid-life: The CARDIA Brain MRI sub-study*. PLoS One, 2018. **13**(1): p. e0190478.
112. Armstrong, A.C., et al., *Association of early left ventricular dysfunction with advanced magnetic resonance white matter and gray matter brain measures: The CARDIA study*. Echocardiography, 2017. **34**(11): p. 1617-1622.
113. Zhu, N., et al., *Cardiorespiratory fitness and brain volume and white matter integrity: The CARDIA Study*. Neurology, 2015. **84**(23): p. 2347-53.
114. Susan Standing, R.T., *Gray's Anatomy: The Anatomical Basis of Clinical Practice*. 2015: Elsevier; 41 edition. 1584.
115. Andreasen, N., et al., *Cerebrospinal fluid tau protein as a biochemical marker for Alzheimer's disease: a community based follow up study*. J Neurol Neurosurg Psychiatry, 1998. **64**(3): p. 298-305.
116. Blennow, K., et al., *Tau protein in cerebrospinal fluid: a biochemical marker for axonal degeneration in Alzheimer disease?* Mol Chem Neuropathol, 1995. **26**(3): p. 231-45.
117. Tato, R.E., A. Frank, and A. Hernanz, *Tau protein concentrations in cerebrospinal fluid of patients with dementia of the Alzheimer type*. J Neurol Neurosurg Psychiatry, 1995. **59**(3): p. 280-3.
118. Blennow, K., *Cerebrospinal fluid protein biomarkers for Alzheimer's disease*. NeuroRx, 2004. **1**(2): p. 213-25.
119. Kabani, N.J., et al., *3D Anatomical Atlas of the Human Brain*. NeuroImage, 1998. **7**(4).
120. Zacharaki, E.I., et al. *Measuring Brain Lesion Progression with a Supervised Tissue Classification System*. in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2008*. 2008. Berlin, Heidelberg: Springer Berlin Heidelberg.
121. Sheejakumari, V. and B. Sankara Gomathi, *MRI brain images healthy and pathological tissues classification with the aid of improved particle swarm optimization and neural network*. Comput Math Methods Med, 2015. **2015**: p. 807826.

122. N. Kabani, D.C., and A. Evans, *A 3d neuroanatomical atlas*, in *4th International Conference on Functional Mapping of the Human Brain*. 1998, Organization for Human Brain Mapping.
123. Walther, S., et al., *Alterations of white matter integrity related to motor activity in schizophrenia*. Neurobiol Dis, 2011. **42**(3): p. 276-83.
124. Douaud, G., et al., *Anatomically related grey and white matter abnormalities in adolescent-onset schizophrenia*. Brain, 2007. **130**(Pt 9): p. 2375-86.
125. Kumra, S., et al., *White matter abnormalities in early-onset schizophrenia: a voxel-based diffusion tensor imaging study*. J Am Acad Child Adolesc Psychiatry, 2005. **44**(9): p. 934-41.
126. Wolters, F.J., et al., *Cerebral Perfusion and the Risk of Dementia*. 2017. **136**(8): p. 719-728.
127. Jayashri Srinivasan, C.C., Brian Scott, Juan E. Small, *Netter's Neurology E-Book*. 3 edition ed. 2019: Elsevier
128. Melton, A.W., *Implications of short-term memory for a general theory of memory*. Journal of Verbal Learning and Verbal Behavior, 1963. **2**(1): p. 1-21.
129. Bubb, E.J., L. Kinnavane, and J.P. Aggleton, *Hippocampal - diencephalic - cingulate networks for memory and emotion: An anatomical guide*. Brain and neuroscience advances, 2017. **1**(1): p. 2398212817723443.
130. Cerf-Ducastel, B. and C. Murphy, *Neural substrates of cross-modal olfactory recognition memory: An fMRI study*. NeuroImage, 2006. **31**(1): p. 386-396.
131. Duvernoy, H.M., *The human hippocampus : functional anatomy, vascularization, and serial sections with MRI*. 3rd ed. ed. 2005, Berlin :: Springer.
132. Watchman, K., *Intellectual Disabilities and Dementia: A Guide for Families*. 2017: Jessica Kingsley Publishers; 1 edition (21 Mar. 2017). 120.
133. Todd Vanderah, D.J.G., *Nolte's The Human Brain: An Introduction to its Functional Anatomy*. 2015: Elsevier. 720.
134. Khan, U.A., et al., *Molecular drivers and cortical spread of lateral entorhinal cortex dysfunction in preclinical Alzheimer's disease*. Nature Neuroscience, 2014. **17**(2): p. 304-11.
135. Braak, H., et al., *Staging of Alzheimer disease-associated neurofibrillary pathology using paraffin sections and immunocytochemistry*. Acta neuropathologica, 2006. **112**(4): p. 389-404.
136. Society, A.s., *Dementia and the brain*. 2015: alzheimers.org.uk.
137. Bailey, H.R., et al., *Medial temporal lobe volume predicts elders' everyday memory*. Psychological science, 2013. **24**(7): p. 1113-1122.
138. Van Hoesen, G.W. and D.N. Pandya, *Some connections of the entorhinal (area 28) and perirhinal (area 35) cortices of the rhesus monkey. III. Efferent connections*. Brain Research, 1975. **95**(1): p. 39-59.
139. Dickerson, B.C. and R.A. Sperling, *Functional abnormalities of the medial temporal lobe memory system in mild cognitive impairment and Alzheimer's disease: insights from functional MRI studies*. Neuropsychologia, 2008. **46**(6): p. 1624-35.
140. Visser, P.J., et al., *Medial temporal lobe atrophy predicts Alzheimer's disease in patients with minor cognitive impairment*. J Neurol Neurosurg Psychiatry, 2002. **72**(4): p. 491-7.
141. Takehara-Nishiuchi, K., *Entorhinal cortex and consolidated memory*. Neurosci Res, 2014. **84**: p. 27-33.
142. Gómez-Isla, T., et al., *Profound loss of layer II entorhinal cortex neurons occurs in very mild Alzheimer's disease*. J Neurosci, 1996. **16**(14): p. 4491-500.
143. Du, A.-T., et al., *Age effects on atrophy rates of entorhinal cortex and hippocampus*. Neurobiology of aging, 2006. **27**(5): p. 733-740.
144. Van Hoesen, G.W., B.T. Hyman, and A.R. Damasio, *Entorhinal cortex pathology in Alzheimer's disease*. Hippocampus, 1991. **1**(1): p. 1-8.
145. Mengxi, Z., et al., *Entorhinal cortex: a good biomarker of mild cognitive impairment and mild Alzheimer's disease*. Reviews in the Neurosciences, 2016. **27**(2): p. 185-195.
146. Dona, A.C., et al., *Precision high-throughput proton NMR spectroscopy of human urine, serum, and plasma for large-scale metabolic phenotyping*. Anal Chem., 2014. **86**(19): p. 9887-94. doi: 10.1021/ac5025039. Epub 2014 Sep 16.
147. Nicholson, J., et al., *Metabolic Phenotyping in Personalized and Public Healthcare*. 2016, San Diego, UNITED STATES: Elsevier Science.
148. Holmes, E., et al., *The promise of metabolic phenotyping in gastroenterology and hepatology*. Nat Rev Gastroenterol Hepatol., 2015. **12**(8): p. 458-71. doi: 10.1038/nrgastro.2015.114. Epub 2015 Jul 21.

149. Jacobsen, N.E., *NMR Data Interpretation Explained: Understanding 1D and 2D NMR Spectra of Organic Compounds and Natural Products*. 2016: Wiley.
150. Jellum, E., et al., *Application of glass capillary-column gas chromatography-mass spectrometry to the studies of human diseases*. J Chromatogr, 1976. **126**: p. 487-93.
151. Nicholson, J.K., et al., *Metabolic phenotyping in clinical and surgical environments*. Nature, 2012. **491**: p. 384.
152. Plumb, R., et al., *Ultra-performance liquid chromatography coupled to quadrupole-orthogonal time-of-flight mass spectrometry*. Rapid Commun Mass Spectrom, 2004. **18**(19): p. 2331-7.
153. Dunn, W.B., et al., *Procedures for large-scale metabolic profiling of serum and plasma using gas chromatography and liquid chromatography coupled to mass spectrometry*. Nature Protocols, 2011. **6**(7): p. 1060-1083.
154. van der Kloet, F.M., et al., *Analytical Error Reduction Using Single Point Calibration for Accurate and Precise Metabolomic Phenotyping*. Journal of Proteome Research, 2009. **8**(11): p. 5132-5141.
155. Dunn, W.B., et al., *Systems level studies of mammalian metabolomes: the roles of mass spectrometry and nuclear magnetic resonance spectroscopy*. Chem Soc Rev, 2011. **40**(1): p. 387-426.
156. Fiehn, O., *Metabolomics – the link between genotypes and phenotypes*. Plant Molecular Biology, 2002. **48**(1): p. 155-171.
157. Daviss, B., *Growing Pains for Metabolomics*. Scientist, 2005. **19**: p. 25-28.
158. Di Guida, R., et al., *Non-targeted UHPLC-MS metabolomic data processing methods: a comparative investigation of normalisation, missing value imputation, transformation and scaling*. Metabolomics, 2016. **12**: p. 93.
159. Cook, T., Y. Ma, and S. Gamagedara, *Evaluation of statistical techniques to normalize mass spectrometry-based urinary metabolomics data*. Journal of Pharmaceutical and Biomedical Analysis, 2020. **177**: p. 112854.
160. Rosen Vollmar, A.K., et al., *Normalizing Untargeted Periconceptional Urinary Metabolomics Data: A Comparison of Approaches*. Metabolites, 2019. **9**(10).
161. David Balding, I.M., John Marioni, *Handbook of Statistical Genomics: Two Volume Set, Fourth Edition*. 2019: John Wiley & Sons Ltd.
162. Dieterle, F., et al., *Probabilistic quotient normalization as robust method to account for dilution of complex biological mixtures. Application in 1H NMR metabonomics*. Anal Chem, 2006. **78**(13): p. 4281-90.
163. Teo, B.W., et al., *Spot urine estimations are equivalent to 24-hour urine assessments of urine protein excretion for predicting clinical outcomes*. Int J Nephrol, 2015. **2015**: p. 156484.
164. Abubacker, N.R.T., et al., *A comparative study of spot urine versus 24 hour urine in assessment of proteinuria in varying degree of renal dysfunction*. 2017, 2017. **3**(1): p. 4.
165. Ulrich, K., *BIPLLOT: Stata module to generate biplots*. 1998, Boston College Department of Economics.
166. Nelson, D.L., *Lehninger principles of biochemistry*. 2005: Fourth edition. New York : W.H. Freeman, 2005.
167. Tuokko, H.A. and D.F. Hulstsch, *Mild Cognitive Impairment: International Perspectives*. 2019: Taylor & Francis.
168. Hox, J. and T. Bechger, *An Introduction to Structural Equation Modeling*. Family Science Review, 1999. **11**.
169. Strong, R. and G. Alvarez, *Using simulation and resampling to improve the statistical power and reproducibility of psychological research*. 2019.
170. Whitwell, J.L., et al., *Normalization of Cerebral Volumes by Use of Intracranial Volume: Implications for Longitudinal Quantitative MR Imaging*. American Journal of Neuroradiology, 2001. **22**(8): p. 1483.
171. VanderWeele, T.J. and S. Vansteelandt, *Mediation Analysis with Multiple Mediators*. Epidemiologic methods, 2014. **2**(1): p. 95-115.
172. Smith, K.E., et al., *Network analysis: An innovative framework for understanding eating disorder psychopathology*. The International journal of eating disorders, 2018. **51**(3): p. 214-222.
173. Jia, Z., *Controlling the Overfitting of Heritability in Genomic Selection through Cross Validation*. Scientific reports, 2017. **7**(1): p. 13678-9.
174. Jodie B. Ullman, P.M.B., *Structural Equation Modeling*, in *Handbook of Psychology*. 2005, John Wiley & Sons, Ltd. p. 607-634.
175. Forest, J. *When do you need to standardize the variables in a regression model?* 2017.

176. Nikitina, L., R. Paidi, and F. Furuoka, *Using bootstrapped quantile regression analysis for small sample research in applied linguistics: Some methodological considerations*. PLOS ONE, 2019. **14**(1): p. e0210668.
177. Friedman, M., *Analysis, Nutrition, and Health Benefits of Tryptophan*. International journal of tryptophan research : IJTR, 2018. **11**: p. 1178646918802282-1178646918802282.
178. Tapia-Rojas, C., et al., *Is L-methionine a trigger factor for Alzheimer's-like neurodegeneration?: Changes in A β oligomers, tau phosphorylation, synaptic proteins, Wnt signaling and behavioral impairment in wild-type mice*. Molecular Neurodegeneration, 2015. **10**(1): p. 62.
179. Deborah, M., et al., *Spectroscopic studies on sidewall carboxylic acid functionalization of multi-walled carbon nanotubes with valine*. Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy, 2015. **139**: p. 138-144.
180. Glenn, J.M., E.N. Madero, and N.T. Bott, *Dietary Protein and Amino Acid Intake: Links to the Maintenance of Cognitive Health*. Nutrients, 2019. **11**(6).
181. Halson, S.L., *Sleep in elite athletes and nutritional interventions to enhance sleep*. Sports medicine (Auckland, N.Z.), 2014. **44 Suppl 1**(Suppl 1): p. S13-S23.
182. Inagawa, K., et al., *Subjective effects of glycine ingestion before bedtime on sleep quality*. Sleep and Biological Rhythms, 2006. **4**(1): p. 75-77.
183. Gannon, M.C., J.A. Nuttall, and F.Q. Nuttall, *The metabolic response to ingested glycine*. Am J Clin Nutr, 2002. **76**(6): p. 1302-7.
184. Yan-Do, R., et al., *A Glycine-Insulin Autocrine Feedback Loop Enhances Insulin Secretion From Human β -Cells and Is Impaired in Type 2 Diabetes*. Diabetes, 2016. **65**(8): p. 2311.
185. Ott, A., et al., *Association of diabetes mellitus and dementia: the Rotterdam Study*. Diabetologia, 1996. **39**(11): p. 1392-7.
186. Leibson, C.L., et al., *Risk of dementia among persons with diabetes mellitus: a population-based cohort study*. Am J Epidemiol, 1997. **145**(4): p. 301-8.
187. Arvanitakis, Z., et al., *Diabetes mellitus and risk of Alzheimer disease and decline in cognitive function*. Arch Neurol, 2004. **61**(5): p. 661-6.
188. Janson, J., et al., *Increased risk of type 2 diabetes in Alzheimer disease*. Diabetes, 2004. **53**(2): p. 474-81.
189. Norton, L.E. and D.K. Layman, *Leucine regulates translation initiation of protein synthesis in skeletal muscle after exercise*. J Nutr, 2006. **136**(2): p. 533s-537s.
190. Breen, L. and T.A. Churchward-Venne, *Leucine: a nutrient 'trigger' for muscle anabolism, but what more?* The Journal of physiology, 2012. **590**(9): p. 2065-2066.
191. Layman, D.K., *The role of leucine in weight loss diets and glucose homeostasis*. J Nutr, 2003. **133**(1): p. 261s-267s.
192. Kapalka, G.M., *Chapter 4 - Substances Involved in Neurotransmission*, in *Nutritional and Herbal Therapies for Children and Adolescents*, G.M. Kapalka, Editor. 2010, Academic Press: San Diego. p. 71-99.
193. William K. Ovalle, P.C.N., *Netter's Essential Histology: With Correlated Histopathology*. 3rd ed. 2020: Elsevier. 568.
194. Palego, L., et al., *Tryptophan Biochemistry: Structural, Nutritional, Metabolic, and Medical Aspects in Humans*. Journal of amino acids, 2016. **2016**: p. 8952520-8952520.
195. Litwack, G., *Chapter 18 - Membrane Transport*, in *Human Biochemistry*, G. Litwack, Editor. 2018, Academic Press: Boston. p. 553-589.
196. Pardridge, W.M., *CNS drug design based on principles of blood-brain barrier transport*. J Neurochem, 1998. **70**(5): p. 1781-92.
197. Zinger, A., et al., *The involvement of neuroinflammation and kynurenine pathway in Parkinson's disease*. Parkinsons Dis, 2011. **2011**: p. 716859.
198. Lin, Y.-T., et al., *Indole-3 acetic acid increased risk of impaired cognitive function in patients receiving hemodialysis*. NeuroToxicology, 2019. **73**: p. 85-91.
199. Karu, N., et al., *Tryptophan metabolism, its relation to inflammation and stress markers and association with psychological and cognitive functioning: Tasmanian Chronic Kidney Disease pilot study*. BMC Nephrology, 2016. **17**(1): p. 171.
200. Sublette, M.E., et al., *Plasma kynurenine levels are elevated in suicide attempters with major depressive disorder*. Brain, Behavior, and Immunity, 2011. **25**(6): p. 1272-1278.

201. Krause, D., et al., *High Kynurenine (a Tryptophan Metabolite) Predicts Remission in Patients with Major Depression to Add-on Treatment with Celecoxib*. Front Psychiatry, 2017. **8**: p. 16.
202. Myint, A., M. Schwarz, and N. Müller, *The role of the kynurenine metabolism in major depression*. Journal of neural transmission (Vienna, Austria : 1996), 2011. **119**: p. 245-51.
203. Myint, A.-M., et al., *Kynurenine pathway in major depression: Evidence of impaired neuroprotection*. Journal of Affective Disorders, 2007. **98**(1): p. 143-151.
204. Shan, Z., et al., *Selective Atrophy of Left Hemisphere and Frontal Lobe of the Brain in Old Men*. The journals of gerontology. Series A, Biological sciences and medical sciences, 2005. **60**: p. 165-74.
205. Heck, R., *An Introduction to Multilevel Modeling Techniques: MLM and SEM Approaches Using Mplus*. 3 edition ed. Quantitative Methodology Series. 2015: Routledge.
206. Emmanuel Lesaffre, A.B.L., *Bayesian Biostatistics*. 2012: Wiley-Blackwell; 1 edition (27 July 2012). 534.
207. Bunch, T.J., et al., *Atrial Fibrillation and Dementia: Exploring the Association, Defining Risks and Improving Outcomes*. Arrhythmia & electrophysiology review, 2019. **8**(1): p. 8-12.
208. Peters, A., et al., *Feature article: are neurons lost from the primate cerebral cortex during normal aging?* Cereb Cortex, 1998. **8**(4): p. 295-300.
209. Terry, R.D., R. DeTeresa, and L.A. Hansen, *Neocortical cell counts in normal human adult aging*. Ann Neurol, 1987. **21**(6): p. 530-9.
210. Farkas, E. and P.G. Luiten, *Cerebral microvascular pathology in aging and Alzheimer's disease*. Prog Neurobiol, 2001. **64**(6): p. 575-611.
211. Riddle, D.R., W.E. Sonntag, and R.J. Lichtenwalner, *Microvascular plasticity in aging*. Ageing Res Rev, 2003. **2**(2): p. 149-68.
212. Giorgio, A., et al., *Age-related changes in grey and white matter structure throughout adulthood*. NeuroImage, 2010. **51**(3): p. 943-951.
213. Sowell, E.R., et al., *Mapping cortical change across the human life span*. Nature Neuroscience, 2003. **6**(3): p. 309-315.
214. Ge, Y., et al., *Age-related total gray matter and white matter changes in normal adult brain. Part II: quantitative magnetization transfer ratio histogram analysis*. AJNR Am J Neuroradiol, 2002. **23**(8): p. 1334-41.
215. Enzinger, C., et al., *Risk factors for progression of brain atrophy in aging: six-year follow-up of normal subjects*. Neurology, 2005. **64**(10): p. 1704-11.
216. Pagani, E., et al., *Voxel-based analysis derived from fractional anisotropy images of white matter volume changes with aging*. Neuroimage, 2008. **41**(3): p. 657-67.
217. Rovaris, M., et al., *Age-related changes in conventional, magnetization transfer, and diffusion-tensor MR imaging findings: study with whole-brain tissue histogram analysis*. Radiology, 2003. **227**(3): p. 731-8.
218. Selzer, P.M.a.a.h.i.l.g.v.r.a., R.J.a.a.h.i.l.g.v.r.a. Marhöfer, and O.a.a.h.i.l.g.v.r.a. Koch, *Applied Bioinformatics An Introduction*. 2nd ed. 2018. ed. 2018.
219. Preacher, K.J., M.J. Zyphur, and Z. Zhang, *A general multilevel SEM framework for assessing multilevel mediation*. Psychol Methods, 2010. **15**(3): p. 209-33.
220. Krull, J.L. and D.P. MacKinnon, *Multilevel Modeling of Individual and Group Level Mediated Effects*. Multivariate Behav Res, 2001. **36**(2): p. 249-77.
221. Sturm, W. and K. Willmes, *Efficacy of reaction training on various attentional and cognitive functions in stroke patients*. Neuropsychological Rehabilitation - NEUROPSYCHOL REHABIL, 1991. **1**: p. 259-280.
222. Langer, W., *How to use Stata's sem with small samples? New corrections for the Likelihood-Ratio chi2 statistics and fit indices*. 2018.
223. Herzog, W. and A. Boomsma, *Small-Sample Robust Estimators of Noncentrality-Based and Incremental Model Fit*. Structural Equation Modeling A Multidisciplinary Journal, 2009. **16**: p. 1-27.
224. Herzog, W., A. Boomsma, and S. Reinecke, *The Model-Size Effect on Traditional and Modified Tests of Covariance Structures*. Structural Equation Modeling: A Multidisciplinary Journal, 2007. **14**(3): p. 361-390.
225. Yuan, K.H., Y. Tian, and H. Yanagihara, *Empirical Correction to the Likelihood Ratio Statistic for Structural Equation Modeling with Many Variables*. Psychometrika, 2015. **80**(2): p. 379-405.

226. McNeish, D., *Should We Use F-Tests For Model Fit instead of Chi-Square in Over-Identified Structural Equation Models?* Organizational Research Methods, 2018. **23**.
227. Antonakis, J. and N. Bastardo, *SWAIN: Stata module to correct the SEM chi-square overidentification test in small sample sizes or complex models*. 2020.
228. Gelman, A., J. Hill, and M. Yajima, *Why We (Usually) Don't Have to Worry About Multiple Comparisons*. Journal of Research on Educational Effectiveness, 2009. **5**.
229. Kang, Y.E., et al., *The Roles of Adipokines, Proinflammatory Cytokines, and Adipose Tissue Macrophages in Obesity-Associated Insulin Resistance in Modest Obesity and Early Metabolic Dysfunction*. PLOS ONE, 2016. **11**(4): p. e0154003.
230. Kahn, B.B. and J.S. Flier, *Obesity and insulin resistance*. The Journal of clinical investigation, 2000. **106**(4): p. 473-481.
231. Whitmer, R.A., *The epidemiology of adiposity and dementia*. Curr Alzheimer Res, 2007. **4**(2): p. 117-22.
232. Kiliaan, A.J., I.A.C. Arnoldussen, and D.R. Gustafson, *Adipokines: a link between obesity and dementia?* The Lancet Neurology, 2014. **13**(9): p. 913-923.
233. DeBette, S., et al., *Visceral fat is associated with lower brain volume in healthy middle-aged adults*. Ann Neurol, 2010. **68**(2): p. 136-44.
234. Pannacciulli, N., et al., *Brain abnormalities in human obesity: a voxel-based morphometric study*. Neuroimage, 2006. **31**(4): p. 1419-25.
235. Raji, C.A., et al., *Brain structure and obesity*. Hum Brain Mapp, 2010. **31**(3): p. 353-64.
236. Willeumier, K.C., D.V. Taylor, and D.G. Amen, *Elevated BMI is associated with decreased blood flow in the prefrontal cortex using SPECT imaging in healthy adults*. Obesity (Silver Spring), 2011. **19**(5): p. 1095-7.
237. Volkow, N.D., et al., *Inverse association between BMI and prefrontal metabolic activity in healthy adults*. Obesity (Silver Spring), 2009. **17**(1): p. 60-5.
238. Kullmann, S., et al., *Selective insulin resistance in homeostatic and cognitive control brain areas in overweight and obese adults*. Diabetes Care, 2015. **38**(6): p. 1044-50.
239. Eddy, S.R., *What is Bayesian statistics?* Nature Biotechnology, 2004. **22**(9): p. 1177-1178.
240. van Oijen, M., *Bayesian Methods for Quantifying and Reducing Uncertainty and Error in Forest Models*. Current Forestry Reports, 2017. **3**(4): p. 269-280.
241. Espino-Hernandez, G., P. Gustafson, and I. Burstyn, *Bayesian adjustment for measurement error in continuous exposures in an individually matched case-control study*. BMC Medical Research Methodology, 2011. **11**(1): p. 67.
242. Bender, R. and S. Lange, *Adjusting for multiple testing—when and how?* Journal of Clinical Epidemiology, 2001. **54**(4): p. 343-349.
243. Mantel, N. and W. Haenszel, *Statistical Aspects of the Analysis of Data From Retrospective Studies of Disease*. JNCI: Journal of the National Cancer Institute, 1959. **22**(4): p. 719-748.
244. Baron, A.D., *Vascular reactivity*. The American Journal of Cardiology, 1999. **84**(1, Supplement 1): p. 25-27.
245. Plata-Salamán, C.R., *Insulin in the cerebrospinal fluid*. Neuroscience & Biobehavioral Reviews, 1991. **15**(2): p. 243-258.
246. Armstrong, N., et al., *Associations between Cognitive and Brain Volume Changes in Cognitively Normal Older Adults*. NeuroImage, 2020. **223**: p. 117289.
247. Willette, A.A., et al., *Insulin resistance, brain atrophy, and cognitive performance in late middle-aged adults*. Diabetes Care, 2013. **36**(2): p. 443-9.
248. Hammer, O., D. Harper, and P. Ryan, *PAST: Paleontological Statistics Software Package for Education and Data Analysis*. Palaeontologia Electronica, 2001. **4**: p. 1-9.
249. Fruchterman, T.M.J. and E.M. Reingold, *Graph drawing by force-directed placement*. Software: Practice and Experience, 1991. **21**(11): p. 1129-1164.
250. Li, D., et al., *Prospective associations of plasma phospholipids and mild cognitive impairment/dementia among African Americans in the ARIC Neurocognitive Study*. Alzheimer's & dementia (Amsterdam, Netherlands), 2016. **6**: p. 1-10.
251. Toledo, J.B., et al., *Metabolic network failures in Alzheimer's disease: A biochemical road map*. Alzheimers Dement, 2017. **13**(9): p. 965-984.
252. Inc., S.I., *SAS/STAT® 14.1 User's Guide*. 2015, SAS Institute Inc., Cary, NC, USA.

253. Hastie, T., R. Tibshirani, and J.H. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. 2009: Springer.
254. Breiman, L., et al., *Classification and Regression Trees*. 1984: Taylor & Francis.
255. Barry de Ville, P.N., *Decision Trees for Analytics Using SAS® Enterprise Miner™*, S.I. Inc., Editor. 2013: SAS Institute Inc., Cary, NC, USA.
256. Inc., S.I., *SAS® Factory Miner 15.1: User's Guide*, C. SAS Institute Inc., NC, USA, Editor. 2018.
257. Kuhn, M. and K. Johnson, *Applied Predictive Modeling*. 2013.
258. Inc., S.I., *SAS® Visual Data Mining and Machine Learning 8.1: Statistical Procedures*. 2016: SAS Institute Inc., Cary, NC, USA.
259. Inc., S.I., *SAS® Enterprise Miner™ 14.1: High-Performance Procedures*. 2015: SAS Institute Inc., Cary, NC, USA.
260. Patrick Hall, J.D., Ilknur Kaynar Kabul, Jorge Silva, *An Overview of Machine Learning with SAS® Enterprise Miner™*, S.I. Inc., Editor. 2014: SAS Enterprise Miner, 100 SAS Campus Dr., Cary, NC 27513.
261. Friedman, J.H., *Greedy Function Approximation: A Gradient Boosting Machine*. The Annals of Statistics, 2001. **29**(5): p. 1189-1232.
262. Bartholomew, D.J., *Principal Components Analysis*, in *International Encyclopedia of Education (Third Edition)*, P. Peterson, E. Baker, and B. McGaw, Editors. 2010, Elsevier: Oxford. p. 374-377.
263. Phillips, R. and S. Cro, *AEVOLCANO: Stata module to produce volcano plot for adverse event data*. 2020.
264. Panneerselvam, N.D., *The More Trees, the Better! Scaling Up Performance Using Random Forest in SAS® Enterprise Miner™*, O.S.U. Spears School of Business, Stillwater, OK 74078, Editor. 2013: SAS Institute Inc.
265. Bartzokis, G., et al., *Age-Related Changes in Frontal and Temporal Lobe Volumes in Men: A Magnetic Resonance Imaging Study*. Archives of General Psychiatry, 2001. **58**(5): p. 461-465.
266. Canoy, D., et al., *Body mass index and incident coronary heart disease in women: a population-based prospective study*. BMC Medicine, 2013. **11**: p. 87-87.
267. Conroy, R.M., et al., *Estimation of ten-year risk of fatal cardiovascular disease in Europe: the SCORE project*. European Heart Journal, 2003. **24**(11): p. 987-1003.
268. Hippisley-Cox, J., C. Coupland, and P. Brindle, *Development and validation of QRISK3 risk prediction algorithms to estimate future risk of cardiovascular disease: prospective cohort study*. BMJ, 2017. **357**: p. j2099.
269. Hippisley-Cox, J., et al., *Performance of the QRISK cardiovascular risk prediction algorithm in an independent UK sample of patients from general practice: a validation study*. Heart, 2008. **94**(1): p. 34.
270. Lebenbaum, M., et al., *Development and validation of a population based risk algorithm for obesity: The Obesity Population Risk Tool (OPoRT)*. PLOS ONE, 2018. **13**(1): p. e0191169.
271. Shields, M., et al., *Bias in self-reported estimates of obesity in Canadian health surveys: an update on correction equations for adults*. Health Rep, 2011. **22**(3): p. 35-45.
272. Hanley, J.A., et al., *Statistical analysis of correlated data using generalized estimating equations: an orientation*. Am J Epidemiol, 2003. **157**(4): p. 364-75.
273. Steffen, A., et al., *Development and validation of a risk score predicting substantial weight gain over 5 years in middle-aged European men and women*. PLoS One, 2013. **8**(7): p. e67429.
274. Bachlechner, U., et al., *Predicting risk of substantial weight gain in German adults—a multi-center cohort approach*. European Journal of Public Health, 2017. **27**(4): p. 768-774.
275. Douketis, J.D., et al., *Systematic review of long-term weight loss studies in obese adults: clinical significance and applicability to clinical practice*. International Journal of Obesity, 2005. **29**(10): p. 1153-1167.
276. Sindi, S., et al., *The CAIDE Dementia Risk Score App: The development of an evidence-based mobile application to predict the risk of dementia*. Alzheimer's & dementia (Amsterdam, Netherlands), 2015. **1**(3): p. 328-333.
277. Kivipelto, M., et al., *Risk score for the prediction of dementia risk in 20 years among middle aged people: a longitudinal, population-based study*. Lancet Neurol, 2006. **5**(9): p. 735-41.
278. Exalto, L.G., et al., *Midlife risk score for the prediction of dementia four decades later*. Alzheimer's & Dementia, 2014. **10**(5): p. 562-570.

279. Chosy, E.J., et al., *The CAIDE Dementia Risk Score and the Honolulu-Asia Aging Study*. Dement Geriatr Cogn Disord, 2019. **48**(3-4): p. 164-171.
280. Jack, C.R., Jr., et al., *NIA-AA Research Framework: Toward a biological definition of Alzheimer's disease*. Alzheimers Dement, 2018. **14**(4): p. 535-562.
281. Hobbs, F.D.R., et al., *Barriers to cardiovascular disease risk scoring and primary prevention in Europe*. QJM: An International Journal of Medicine, 2010. **103**(10): p. 727-739.
282. Fang, X., N.Z. Gebraeel, and K. Paynabar, *Scalable prognostic models for large-scale condition monitoring applications*. IISE Transactions, 2017. **49**(7): p. 698-710.
283. Petrakos, M., J.A. Benediktsson, and I. Kanellopoulos, *The effect of classifier agreement on the accuracy of the combined classifier in decision level fusion*. IEEE Transactions on Geoscience and Remote Sensing, 2001. **39**(11): p. 2539-2546.
284. Wilson, P.W., et al., *An international model to predict recurrent cardiovascular disease*. Am J Med., 2012. **125**(7): p. 695-703.e1. doi: 10.1016/j.amjmed.2012.01.014.
285. Solas, M., et al., *Inflammation and gut-brain axis link obesity to cognitive dysfunction: plausible pharmacological interventions*. Current Opinion in Pharmacology, 2017. **37**: p. 87-92.
286. Sellbom, K.S. and J. Gunstad, *Cognitive function and decline in obesity*. J Alzheimers Dis, 2012. **30 Suppl 2**: p. S89-95.
287. Emmerzaal, T.L., A.J. Kiliaan, and D.R. Gustafson, *2003-2013: A Decade of Body Mass Index, Alzheimer's Disease, and Dementia*. Journal of Alzheimer's Disease, 2015. **43**: p. 739-755.
288. Arnoldussen, I., A. Kiliaan, and D. Gustafson, *Obesity and dementia: Adipokines interact with the brain*. European Neuropsychopharmacology, 2014. **24**.
289. Yaffe, K., et al., *The metabolic syndrome, inflammation, and risk of cognitive decline*. Jama, 2004. **292**(18): p. 2237-42.
290. Hata, J., et al., *Association Between Serum β -Alanine and Risk of Dementia: The Hisayama Study*. American Journal of Epidemiology, 2019. **188**(9): p. 1637-1645.
291. Huang, S., K. Chaudhary, and L.X. Garmire, *More Is Better: Recent Progress in Multi-Omics Data Integration Methods*. Front Genet, 2017. **8**: p. 84.
292. Gershman, S.J. and D.M. Blei, *A tutorial on Bayesian nonparametric models*. Journal of Mathematical Psychology, 2012. **56**(1): p. 1-12.
293. Krumsiek, J., et al., *Gender-specific pathway differences in the human serum metabolome*. Metabolomics, 2015. **11**(6): p. 1815-1833. doi: 10.1007/s11306-015-0829-0. Epub 2015 Aug 4.
294. Wang, B., et al., *Similarity network fusion for aggregating data types on a genomic scale*. Nat Methods., 2014. **11**(3): p. 333-7. doi: 10.1038/nmeth.2810. Epub 2014 Jan 26.

Appendix

Appendix Table 1 Lifestyle exposures at Year 30 (by race and sex)

Total		Mean (SD)	Black, Mean (SD)		White, Mean (SD)		Male, Mean (SD)		Female, Mean (SD)	
Alcohol consumption per day (mean ml) †††	600	11.4 (17.7)	271	9.3 (16.7)	329	13.2 (18.3)	269	15.5 (20.3)	331	8.1 (14.4)
Physical activity exercise units †††	601	341.6 (287.2)	271	258.8 (259.4)	330	409.7 (291.4)	269	405.7 (307.2)	332	289.7 (259)
Diet quality score (v2) †	598	2.6 (5.8)	270	1.4 (5.6)	328	3.7 (5.8)	266	1.9 (5.5)	332	3.2 (6)
CES-D score*	598	8.2 (7.5)	269	9.4 (8.4)	329	7.2 (6.5)	269	8.1 (6.9)	329	8.3 (7.9)
SF-12 Physical Components †	603	50.2 (8.6)	274	48.4 (9.3)	329	51.7 (7.6)	271	50.8 (7.8)	332	49.6 (9.1)
SF-12 Mental Components	603	52 (8.8)	274	52 (9.5)	329	52.1 (8.2)	271	52.5 (7.9)	332	51.7 (9.4)
		Total, N(%)	Black, N(%)		White, N(%)		Male, N(%)		Female, N(%)	
My sleep was restless	Rarely/none	238 (39.4)	109 (39.8)		129 (39.1)		115 (42.4)		123 (36.9)	
	Some	268 (44.4)	112 (40.9)		156 (47.3)		122 (45)		146 (43.8)	
	Much	65 (10.8)	32 (11.7)		33 (10)		25 (9.2)		40 (12)	
	Most/all	33 (5.5)	21 (7.7)		12 (3.6)		9 (3.3)		24 (7.2)	
	Total	604 (100)	274 (100)		330 (100)		271 (100)		333 (100)	
Participate in the PRF and Glucose Tolerance in Women Study	Not restrict carbs intake due to medical conditions ¹⁾	175 (78.8)	77 (82.8)		98 (76.0)				175 (78.8)	
	Oral glucose tolerance test not done	47 (21.2)	16 (17.2)		31 (24.0)				47 (21.2)	
	Total	222 (100)	93 (100)		129 (100)				222 (100)	
On weight reducing diet because of disease ²⁾	No	117 (79.6)	30 (69.8)		87 (83.7)		33 (73.3)		84 (82.4)	
	Yes	30 (20.4)	13 (30.2)		17 (16.4)		12 (26.7)		18 (17.7)	
	Total	147 (100)	43 (100)		104 (100)		45 (100)		102 (100)	
Family income	[1] < \$5,000	22 (3.6)	16 (5.8)		6 (1.8)		9 (3.3)		13 (3.9)	
	[2] \$5,000 - \$11,999	26 (4.3)	19 (6.9)		7 (2.1)		12 (4.4)		14 (4.2)	
	[3] \$12,000 - \$15,999	33 (5.5)	25 (9.1)		8 (2.4)		15 (5.5)		18 (5.4)	
	[4] \$16,000 - \$24,999	29 (4.8)	20 (7.3)		9 (2.7)		14 (5.1)		15 (4.5)	
	[5] \$25,000 - \$34,999	44 (7.3)	32 (11.6)		12 (3.6)		19 (7)		25 (7.5)	
	[6] \$35,000 - \$49,999	44 (7.3)	27 (9.8)		17 (5.2)		16 (5.9)		28 (8.4)	
	[7] \$50,000 - \$74,999	86 (14.2)	43 (15.6)		43 (13)		35 (12.8)		51 (15.3)	
	[8] \$75,000 - \$99,999	89 (14.7)	40 (14.5)		49 (14.9)		40 (14.7)		49 (14.7)	
	[9] ≥ \$100,000	227 (37.5)	50 (18.1)		177 (53.6)		111 (40.7)		116 (34.8)	
	Don't know	1 (0.2)	1 (0.4)		0 (0)		1 (0.4)		0 (0)	
	No response	5 (0.8)	3 (1.1)		2 (0.6)		1 (0.4)		4 (1.2)	
	Total	606 (100)	276 (100)		330 (100)		273 (100)		333 (100)	

1) Not including low-carbohydrate diet

2) For participants have been on a weight reducing diet during the past five years (N=147)

* Center for epidemiologic studies-depression (CES-D) in race ($p=0.003$)

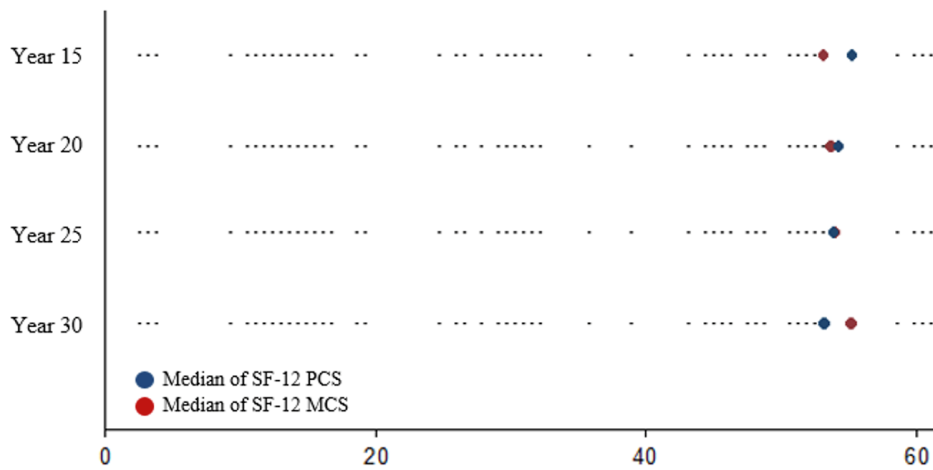
† $p < 0.001$ for race

††† $p < 0.001$ for race and sex, respectively

Appendix Table 2 The 12-item short form health survey (SF-12) at Year 15 to Year 30 (by race and sex)

		Total	Mean (SD)	Black	Mean (SD)	White	Mean (SD)	Male	Mean (SD)	Female	Mean (SD)
SF-12 PCS	Year 15	565	52.9 (6.5)	248	51.9 (6.6)	317	53.7 (6.4)	250	53.4 (6.2)	315	52.4 (6.8)
	Year 20	558	52 (7)	245	50.4 (8.1)	313	53.3 (5.7)	246	52.6 (6.7)	312	51.6 (7.3)
	Year 25	584	51.1 (7.8)	262	49 (8.4)	322	52.8 (6.8)	263	51.7 (7.1)	321	50.6 (8.3)
	Year 30	603	50.2 (8.6)	274	48.4 (9.3)	329	51.7 (7.6)	271	50.8 (7.8)	332	49.6 (9.1)
SF-12 MCS	Year 15	565	50.6 (8.6)	248	50.7 (8.9)	317	50.5 (8.4)	250	51.7 (8.1)	315	49.7 (8.8)
	Year 20	558	50.8 (8.8)	245	50.7 (9.7)	313	50.9 (8.1)	246	51.7 (7.8)	312	50.1 (9.5)
	Year 25	584	51.1 (8.6)	262	50.4 (9.2)	322	51.6 (8.1)	263	51.5 (8.7)	321	50.6 (8.6)
	Year 30	603	52 (8.8)	274	52 (9.5)	329	52.1 (8.2)	271	52.5 (7.9)	332	51.7 (9.4)

Appendix Figure 1 Scatter plots of the mental component score (MCS) and the physical component score (PCS)



Appendix Table 3 Socio-demographic exposures at Year 30 (by race and sex)

		Total, Mean (SD)	Black, Mean (SD)	White, Mean (SD)	Male, Mean (SD)	Female, Mean (SD)
Age †		606 55.2 (3.5)	276 54.5 (3.7)	330 55.9 (3.2)	273 55.1 (3.4)	333 55.3 (3.6)
		Total, N(%)	Black, N(%)	White, N(%)	Male, N(%)	Female, N(%)
Centre	Birmingham	97 (16.0)	59 (21.4)	38 (11.5)	39 (14.3)	58 (17.4)
	Chicago	300 (49.5)	136 (49.3)	164 (49.7)	142 (52)	158 (47.5)
	Minneapolis	108 (17.8)	41 (14.9)	67 (20.3)	47 (17.2)	61 (18.3)
	Oakland	101 (16.7)	40 (14.5)	61 (18.5)	45 (16.5)	56 (16.8)
	Total	606 (100)	276 (100)	330 (100)	273 (100)	333 (100)
Educational attainment maximum years up to Y30 †	<12	10 (1.7)	9 (3.2)	1 (0.3)	5 (1.8)	5 (1.5)
	12-<16	241 (39.8)	154 (55.8)	87 (26.4)	114 (41.8)	127 (38.1)
	16+	355 (58.7)	113 (40.9)	242 (73.4)	154 (56.5)	201 (60.3)
	Total	606 (100)	276 (100)	330 (100)	273 (100)	333 (100)
Smoking status †	Never	387 (64.5)	169 (61.9)	218 (66.7)	177 (65.6)	210 (63.6)
	Former	136 (22.7)	46 (16.9)	90 (27.5)	53 (19.6)	83 (25.2)
	Current	77 (12.8)	58 (21.3)	19 (5.8)	40 (14.8)	37 (11.2)
	Total	600 (100)	273 (100)	327 (100)	270 (100)	330 (100)
		Total, N(%)	Black, N(%)	White, N(%)	Male, N(%)	Female, N(%)
Have a usual source of medical care	No	33 (5.5)	15 (5.4)	18 (5.5)	19 (7.0)	14 (4.2)
	Yes	573 (94.6)	261 (94.6)	312 (94.6)	254 (93.0)	319 (95.8)
	Total	606 (100)	276 (100)	330 (100)	273 (100)	333 (100)

† $p < 0.001$ for race

Appendix Table 4 Clinical exposures at Year 30 (by race and sex)

		Total, N(%)	Black, N(%)	White, N(%)	Male, N(%)	Female, N(%)
Immediate family member diagnosed with dementia? †	No	461 (77.4)	229 (85.1)	232 (71)	223 (82.9)	238 (72.8)
	Yes	135 (22.7)	40 (14.9)	95 (29.1)	46 (17.1)	89 (27.2)
	Total	596 (100)	269 (100)	327 (100)	269 (100)	327 (100)
Family member diagnosed with dementia before 65 years? (for immediate family member diagnosed with dementia)	No	124 (91.9)	36 (90.0)	88 (92.6)	41 (89.1)	83 (93.3)
	Yes	11 (8.2)	4 (10.0)	7 (7.4)	5 (10.9)	6 (6.7)
	Total	135 (100)	40 (100)	95 (100)	46 (100)	89 (100)
Gone through menopause, past 12 months	No	60 (18.0)	36 (22.8)	24 (13.7)		60 (18.0)
	Yes	229 (68.8)	98 (62.0)	131 (74.9)		229 (68.8)
	Don't Know	44 (13.2)	24 (15.2)	20 (11.4)		44 (13.2)
	Total	333 (100)	158 (100)	175 (100)		333 (100)
Kidney urine infection? (for participants with kidney problems)	Yes	13 (24.1)	1 (6.3)	12 (31.6)	2 (7.1)	11 (42.3)
	Don't Know	2 (3.7)	1 (6.3)	1 (2.6)	2 (7.1)	0 (0)
	NA	39 (72.2)	14 (87.5)	25 (65.8)	24 (85.7)	15 (57.7)
	Total	54 (100)	16 (100)	38 (100)	28 (100)	26 (100)
Currently taking medications for HBP †	No	423 (69.9)	151 (54.9)	272 (82.4)	189 (69.2)	234 (70.5)
	Yes	181 (29.9)	124 (45.1)	57 (17.3)	83 (30.4)	98 (29.5)
	Don't Know	1 (0.2)	0 (0)	1 (0.3)	1 (0.4)	0 (0)
	Total	605 (100)	275 (100)	330 (100)	273 (100)	332 (100)
Currently taking medications to lower choles †	No	473 (78.2)	199 (72.4)	274 (83)	211 (77.3)	262 (78.9)
	Yes	125 (20.7)	74 (26.9)	51 (15.5)	57 (20.9)	68 (20.5)
	Don't Know	7 (1.2)	2 (0.7)	5 (1.5)	5 (1.8)	2 (0.6)
	Total	605 (100)	275 (100)	330 (100)	273 (100)	332 (100)
Currently taking medications for heart condition †	No	574 (94.9)	252 (91.6)	322 (97.6)	256 (93.8)	318 (95.8)
	Yes	30 (5.0)	23 (8.4)	7 (2.1)	17 (6.2)	13 (3.9)
	Don't Know	1 (0.2)	0 (0)	1 (0.3)	0 (0)	1 (0.3)
	Total	605 (100)	275 (100)	330 (100)	273 (100)	332 (100)
Currently taking medications for asthma	No	545 (90.1)	237 (86.2)	308 (93.3)	255 (93.4)	290 (87.4)
	Yes	60 (9.9)	38 (13.8)	22 (6.7)	18 (6.6)	42 (12.7)
	Total	605 (100)	275 (100)	330 (100)	273 (100)	332 (100)
Currently taking medications for depression*	No	531 (87.8)	239 (86.9)	292 (88.5)	252 (92.3)	279 (84)
	Yes	73 (12.1)	36 (13.1)	37 (11.2)	20 (7.3)	53 (16)
	Don't Know	1 (0.2)	0 (0)	1 (0.3)	1 (0.4)	0 (0)
	Total	605 (100)	275 (100)	330 (100)	273 (100)	332 (100)
Currently taking medications for diabetes †	No	546 (90.3)	231 (84.0)	315 (95.5)	249 (91.2)	297 (89.5)
	Yes	58 (9.6)	44 (16.0)	14 (4.2)	23 (8.4)	35 (10.5)
	Don't Know	1 (0.2)	0 (0)	1 (0.3)	1 (0.4)	0 (0)
	Total	605 (100)	275 (100)	330 (100)	273 (100)	332 (100)
Currently taking any other medications ††	No	370 (61.2)	178 (64.7)	192 (58.2)	191 (70)	179 (53.9)
	Yes	232 (38.4)	96 (34.9)	136 (41.2)	79 (28.9)	153 (46.1)
	Don't Know	3 (0.5)	1 (0.4)	2 (0.6)	3 (1.1)	0 (0)
	Total	605 (100)	275 (100)	330 (100)	273 (100)	332 (100)
Gone through menopause, past 12 months	No	60 (18.0)	36 (22.8)	24 (13.7)		60 (18.0)
	Yes	229 (68.8)	98 (62.0)	131 (74.9)		229 (68.8)
	Don't Know	44 (13.2)	24 (15.2)	20 (11.4)		44 (13.2)
	Total	333 (100)	158 (100)	175 (100)		333 (100)
	Total	Mean (SD)	Black, Mean (SD)	White, Mean (SD)	Male, Mean (SD)	Female, Mean (SD)
Serum Creatinine (mg/dL) †††	601	0.9 (0.2)	0.9 (0.3)	0.8 (0.2)	1 (0.2)	0.8 (0.2)
Urinary Creatinine (mg/dL) †††	606	124.6 (84.6)	152.2 (91.7)	101.5 (70.5)	148.5 (94.3)	105 (70.1)

* For sex ($p=0.002$)† $p < 0.001$ for race†† $p < 0.001$ for sex††† $p < 0.001$ for race and sex, respectively

Appendix Table 5 Cardio-metabolic risk factors at Year 30 (by race and sex)

		Total	Mean (SD)	Black	Mean (SD)	White	Mean (SD)	Male	Mean (SD)	Female	Mean (SD)
Hypertension	SBP †	606	119.3 (16.1)	276	124.7 (16.9)	330	114.8 (13.8)	273	120.7 (13.5)	333	118.1 (17.9)
	DBP †	606	72.8 (11.4)	276	76.6 (10.8)	330	69.6 (10.9)	273	74.1 (9.9)	333	71.7 (12.3)
Dyslipidaemia	HDL-C ††	601	60.2 (20.5)	271	60.8 (22.9)	330	59.6 (18.3)	271	50.6 (13.0)	330	68 (22.1)
	LDL-C	593	110.6 (32.7)	268	108.4 (35.1)	325	112.5 (30.5)	265	110.5 (32.4)	328	110.8 (32.9)
	Triglycerides (mg/dL) ††	599*	105.8 (66.7)	270	101.7 (64.8)	329	109.1 (68.1)	269	118.5 (72.7)	330	95.5 (59.5)
Hyperglycaemia	Fasting glucose (mg/dL) ††	601	100.5 (25.2)	271	103 (30.9)	330	98.4 (19.2)	271	104.8 (30.8)	330	97 (18.9)
	HOMA-IR †††	593	3.1 (2.8)	266	3.4 (2.7)	327	2.8 (2.8)	265	3.5 (3.2)	328	2.8 (2.4)
	eGFR †	596	91.3 (15.5)	268	94.4 (18.6)	328	88.8 (12)	266	89.5 (15.1)	330	92.8 (15.7)
	Fasting insulin †	598	11.9 (8.7)	269	13.2 (9.3)	329	10.8 (8)	270	12.8 (9.0)	328	11.2 (8.4)
(Abdominal) Obesity	BMI †	605	29.4 (6.2)	275	31.4 (6.5)	330	27.8 (5.4)	272	29.6 (5.2)	333	29.3 (7.0)
	WC †††	605	94.6 (15.2)	276	97.5 (14.6)	329	92.1 (15.2)	273	100.2 (13.5)	332	89.9 (14.8)
CVD	10-Year Framingham risk, male	261	7.7 (4.7)	112	7.8 (5.4)	149	7.7 (4.1)	261	7.7 (4.7)		
	10-Year Framingham risk, female †	323	4.7 (3.7)	153	5.8 (4.5)	170	3.7 (2.4)			323	4.7 (3.7)

*Omitted two distinctive observations of triglycerides > 600 (2366 and 3234 mg/dL)

† $p < 0.001$ for race†† $p < 0.001$ for sex††† $p < 0.001$ for race and sex, respectively

Appendix Table 6 Cardio-metabolic risk factors by risk threshold at Year 30 (by race and sex)

			Total, N(%)	Black, N(%)	White, N(%)	Male, N(%)	Female, N(%)
Hypertension	SBP †	< 140	545 (89.9)	231 (83.7)	314 (95.2)	251 (91.9)	294 (88.3)
		140+	61 (10.1)	45 (16.3)	16 (4.9)	22 (8.1)	39 (11.7)
		Total	606 (100)	276 (100)	330 (100)	273 (100)	333 (100)
	DBP †	< 90	569 (93.9)	243 (88)	326 (98.8)	257 (94.1)	312 (93.7)
		90+	37 (6.1)	33 (12)	4 (1.2)	16 (5.9)	21 (6.3)
		Total	606 (100)	276 (100)	330 (100)	273 (100)	333 (100)
Dyslipidaemia	HDL–C, male	35+ mg/dL	246 (90.8)	106 (91.4)	140 (90.3)	246 (90.8)	
		< 35 mg/dL	25 (9.2)	10 (8.6)	15 (9.7)	25 (9.2)	
		Total	271 (100)	116 (100)	155 (100)	271 (100)	
	HDL–C, female	45+ mg/dL	303 (91.8)	141 (91)	162 (92.6)		303 (91.8)
		< 45 mg/dL	27 (8.2)	14 (9)	13 (7.4)		27 (8.2)
		Total	330 (100)	155 (100)	175 (100)		330 (100)
	LDL–C	< 160 mg/dL	548 (92.4)	244 (91)	304 (93.5)	250 (94.3)	298 (90.9)
		160+ mg/dL	45 (7.6)	24 (9)	21 (6.5)	15 (5.7)	30 (9.2)
		Total	593 (100)	268 (100)	325 (100)	265 (100)	328 (100)
	Triglycerides ††	< 200 mg/dL	552 (92.2)	253 (93.7)	299 (90.9)	236 (87.7)	316 (95.8)
		200+ mg/dL	47 (7.9)	17 (6.3)	30 (9.1)	33 (12.3)	14 (4.2)
		Total	599 (100)	270 (100)	329 (100)	269 (100)	330 (100)
Hyperglycaemia	Fasting glucose ††	< 100 mg/dL	391 (65.1)	161 (59.4)	230 (69.7)	150 (55.4)	241 (73)
		100+ mg/dL	210 (34.9)	110 (40.6)	100 (30.3)	121 (44.7)	89 (27)
		Total	601 (100)	271 (100)	330 (100)	271 (100)	330 (100)
	HOMA-IR †††	< 2.3	298 (50.3)	106 (39.9)	192 (58.7)	111 (41.9)	187 (57)
		2.3+	295 (49.8)	160 (60.2)	135 (41.3)	154 (58.1)	141 (43)
		Total	593 (100)	266 (100)	327 (100)	265 (100)	328 (100)
	eGFR	< 90	264 (44.3)	107 (39.9)	157 (47.9)	135 (50.8)	129 (39.1)
		90+	332 (55.7)	161 (60.1)	171 (52.1)	131 (49.3)	201 (60.9)
		Total	596 (100)	268 (100)	328 (100)	266 (100)	330 (100)
	Fasting insulin	< 20 micro U/ML	518 (86.6)	223 (82.9)	295 (89.7)	227 (84.1)	291 (88.7)
		≥20 micro U/ML	80 (13.4)	46 (17.1)	34 (10.3)	43 (15.9)	37 (11.3)
		Total	598 (100)	269 (100)	329 (100)	270 (100)	328 (100)
(Abdominal) Obesity	BMI †	< 30	356 (58.8)	124 (45.1)	232 (70.3)	162 (59.6)	194 (58.3)
		30+	249 (41.2)	151 (54.9)	98 (29.7)	110 (40.4)	139 (41.7)
		Total	605 (100)	275 (100)	330 (100)	272 (100)	333 (100)
	WC, male	< 102 cm	168 (61.5)	75 (63.6)	93 (60)	168 (61.5)	
		102+ cm	105 (38.5)	43 (36.4)	62 (40)	105 (38.5)	
		Total	273 (100)	118 (100)	155 (100)	273 (100)	
	WC, female †	< 88 cm	161 (48.5)	44 (27.9)	117 (67.2)		161 (48.5)
		88+ cm	171 (51.5)	114 (72.2)	57 (32.8)		171 (51.5)
		Total	332 (100)	158 (100)	174 (100)		332 (100)
CVD	10-Year Framingham Risk, Male	< 10	194 (74.3)	84 (75)	110 (73.8)	194 (74.3)	
		10+	67 (25.7)	28 (25)	39 (26.2)	67 (25.7)	
		Total	261 (100)	112 (100)	149 (100)	261 (100)	
	10-Year Framingham Risk, female †	< 8	277 (85.8)	118 (77.1)	159 (93.5)		277 (85.8)
		8+	46 (14.2)	35 (22.9)	11 (6.5)		46 (14.2)
		Total	323 (100)	153 (100)	170 (100)		323 (100)

† $p < 0.001$ for race†† $p < 0.001$ for sex††† $p < 0.001$ for race and sex, respectively

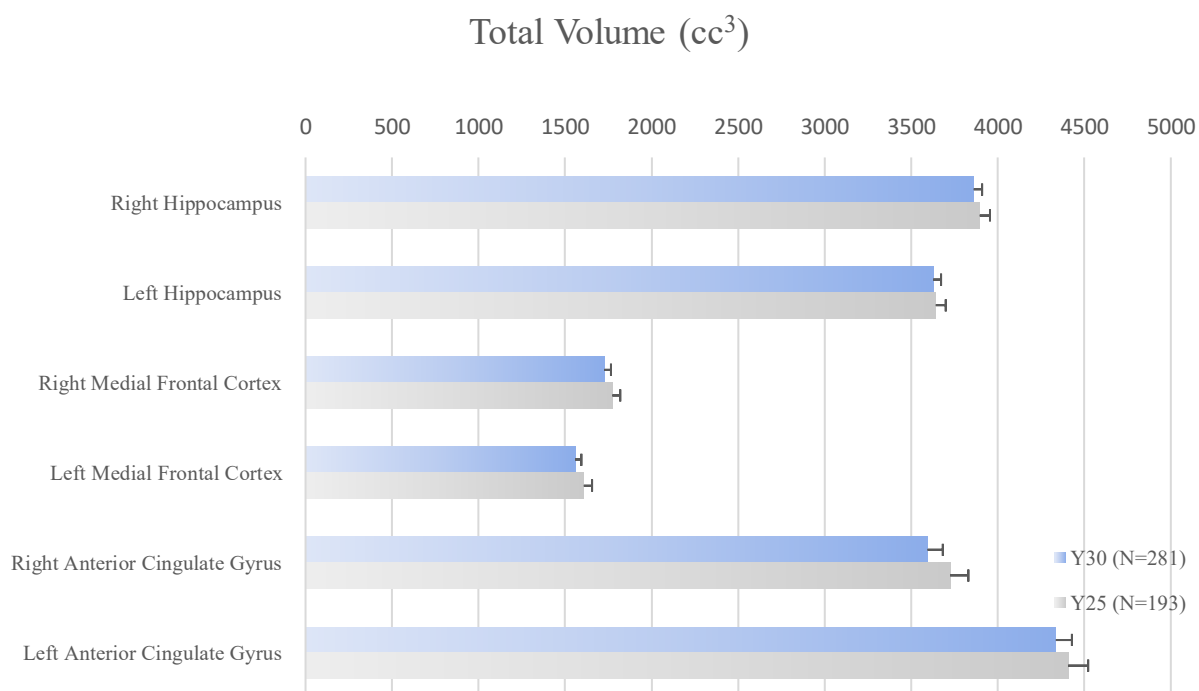
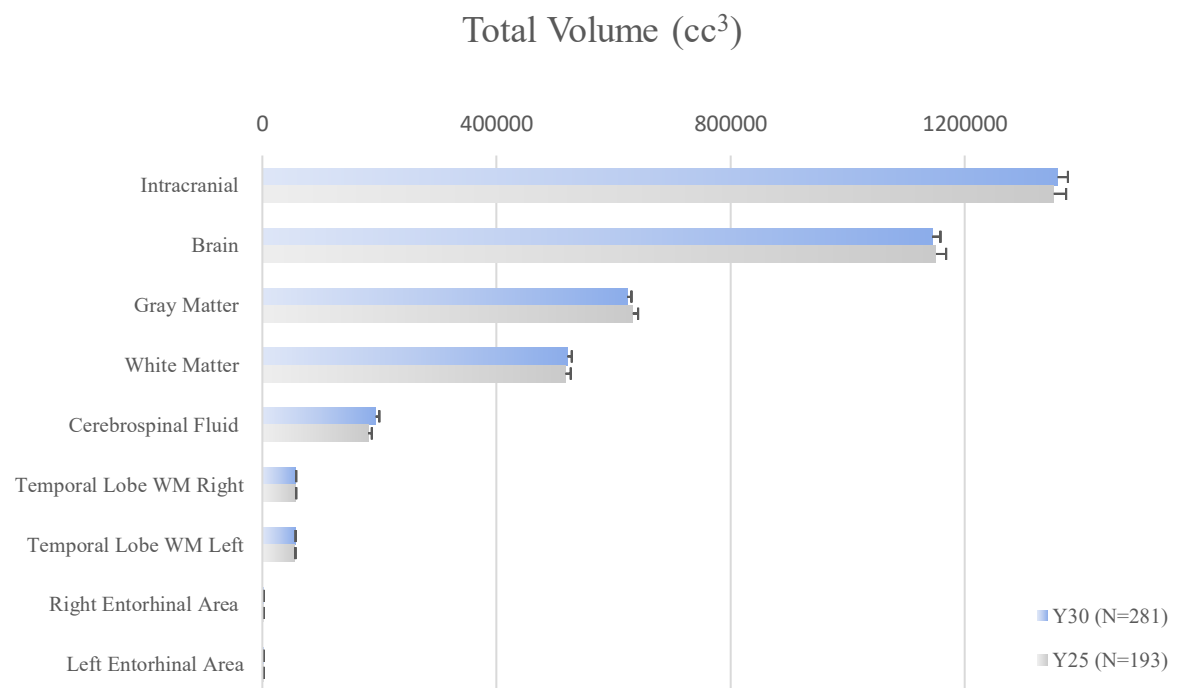
Appendix Table 7 The brain MRI parameters at Year 25 and Year 30

		Year 25					Year 30				
		N	Mean	SD	Min	Max	N	Mean	SD	Min	Max
Total Volume (cc ³)	Intracranial	193	1352273.00	148269.80	1040564.00	1907914.00	281	1358952.00	148628.50	1051819.00	1891086.00
	Brain	193	1150976.00	122181.90	894025.80	1585268.00	281	1144541.00	119533.40	867855.10	1559794.00
	Gray Matter	193	632945.10	64142.34	477411.30	836045.00	281	623221.60	63058.89	466054.00	803872.00
	White Matter	193	518030.80	62881.74	380995.00	749223.00	281	521319.40	62733.39	382918.00	755922.00
	Cerebrospinal Fluid	193	180589.60	45054.46	89707.37	303635.00	281	193641.40	52775.82	86706.16	380149.00
	Right Hippocampus	193	3895.87	417.23	2765.66	5445.00	281	3861.10	411.42	2782.82	5490.00
	Left Hippocampus	193	3641.88	405.60	2609.25	5081.00	281	3626.64	389.11	2653.12	5063.00
	Right Medial Frontal Cortex	193	1772.56	323.40	1127.24	2959.25	281	1726.72	325.59	1020.00	3136.64
	Left Medial Frontal Cortex	193	1607.96	335.96	937.46	3051.00	281	1557.16	309.90	896.00	2649.00
	Right Anterior Cingulate Gyrus	193	3724.62	740.95	2057.00	6712.91	281	3594.38	749.78	1825.00	6601.33
	Left Anterior Cingulate Gyrus	193	4406.59	814.56	2672.20	7496.00	281	4335.75	789.72	2701.00	6999.00
	Temporal Lobe WM Right	193	56931.33	7104.52	42731.29	79261.00	281	57050.93	7145.02	40656.00	79377.00
	Temporal Lobe WM Left	193	55591.79	6988.20	40215.00	80278.00	281	55974.60	6815.33	40448.00	80636.00
	Right Entorhinal Area	193	2241.94	308.43	1533.00	3134.73	281	2282.60	334.94	1263.00	3153.80
	Left Entorhinal Area	193	2250.25	303.59	1548.77	3242.49	281	2288.54	343.19	1111.00	3412.25
Abnormal Tissue Volume (cc ³)	Brain	193	1117.49	1231.35	26.70	10186.20	276	1173.61	1909.52	0.00	17504.69
	Gray Matter	193	119.84	147.03	0.00	1050.95	276	67.01	103.03	0.00	632.00
	White Matter	193	997.65	1183.47	22.89	10097.50	276	1106.60	1866.14	0.00	17207.15
	Cerebrospinal Fluid	193	202.24	283.37	3.81	2634.00	276	145.12	238.24	0.00	2376.00
	Right Hippocampus	193	12.33	25.74	0.00	237.00	276	6.93	17.35	0.00	186.00
	Left Hippocampus	193	7.41	12.68	0.00	67.71	276	3.22	9.32	0.00	100.14
	Right Medial Frontal Cortex	193	0.05	0.35	0.00	3.81	276	0.01	0.12	0.00	1.00
	Left Medial Frontal Cortex	193	0.10	0.71	0.00	6.68	276	0.03	0.31	0.00	4.00
	Right Anterior Cingulate Gyrus	193	15.42	51.77	0.00	528.34	276	2.61	17.02	0.00	180.24
	Left Anterior Cingulate Gyrus	193	8.56	34.23	0.00	305.18	276	3.27	14.20	0.00	177.38
	Temporal Lobe WM Right	193	66.14	207.98	0.00	2550.13	276	79.75	272.89	0.00	3009.80
	Temporal Lobe WM Left	193	62.91	191.77	0.00	1955.03	276	60.84	249.94	0.00	3090.86
	Right Entorhinal Area	193	0.31	3.34	0.00	45.78	276	0.08	0.48	0.00	5.72
	Left Entorhinal Area	193	0.86	4.39	0.00	38.15	276	0.15	0.95	0.00	11.00

(Table Continued)

		Year 25					Year 30				
		N	Mean	SD	Min	Max	N	Mean	SD	Min	Max
(Corrected Mean)	Brain	193	0.22	0.01	0.18	0.29	277	0.21	0.01	0.17	0.24
Fractional	Gray Matter	193	0.15	0.01	0.12	0.24	277	0.13	0.02	0.08	0.17
Anisotropy	White Matter	193	0.31	0.02	0.22	0.35	277	0.30	0.02	0.24	0.36
(Normalized)	Cerebrospinal fluid	193	0.12	0.02	0.09	0.22	277	0.09	0.03	0.03	0.16
	Right Hippocampus	193	0.14	0.02	0.10	0.25	277	0.13	0.02	0.07	0.18
	Left Hippocampus	193	0.14	0.02	0.09	0.27	277	0.13	0.02	0.07	0.19
	Right Medial Frontal Cortex	193	0.13	0.02	0.08	0.21	277	0.11	0.03	0.05	0.16
	Left Medial Frontal Cortex	193	0.15	0.03	0.09	0.23	277	0.12	0.04	0.05	0.21
	Right Anterior Cingulate Gyrus	193	0.12	0.02	0.08	0.23	277	0.11	0.02	0.07	0.18
	Left Anterior Cingulate Gyrus	193	0.13	0.02	0.08	0.21	277	0.11	0.02	0.07	0.17
	Temporal Lobe WM Right	193	0.30	0.02	0.22	0.35	277	0.30	0.02	0.24	0.36
	Temporal Lobe WM Left	193	0.32	0.02	0.22	0.36	277	0.31	0.02	0.24	0.37
	Right Entorhinal Area	193	0.14	0.02	0.10	0.23	277	0.14	0.02	0.08	0.21
	Left Entorhinal Area	193	0.14	0.02	0.11	0.23	277	0.13	0.02	0.07	0.19
(Mean) Cerebral	Brain	185	40.67	9.94	3.56	73.64	173	41.55	8.64	7.86	68.49
Blood Flow	Gray Matter	185	43.34	10.41	3.71	75.92	173	36.32	7.89	5.71	63.43
(mm ³ /100gms/min)	White Matter	185	37.61	9.54	3.35	70.95	173	41.59	8.70	10.12	67.72
	Right Hippocampus	185	44.54	11.53	7.23	104.52	173	42.51	9.99	9.98	82.49
	Left Hippocampus	185	43.63	10.91	2.23	81.69	173	42.97	9.49	7.56	75.14
	Right Medial Frontal Cortex	185	53.15	16.94	6.67	140.70	173	47.56	17.61	-5.94	105.59
	Left Medial Frontal Cortex	185	51.30	16.14	7.11	126.07	173	47.62	16.79	-4.81	111.16
	Right Anterior Cingulate Gyrus	185	53.50	15.61	12.59	131.17	173	46.93	10.55	6.23	84.14
	Left Anterior Cingulate Gyrus	185	52.84	14.79	8.76	119.45	173	47.51	10.20	8.56	77.49
	Temporal Lobe WM Right	185	37.45	9.40	3.13	73.30	173	35.72	8.91	5.36	77.20
	Temporal Lobe WM Left	185	35.40	9.00	3.58	74.27	173	34.36	8.33	5.59	64.31
	Right Entorhinal Area	185	33.87	12.09	3.64	77.01	173	31.63	12.60	-11.23	89.32
	Left Entorhinal Area	185	34.62	12.85	2.98	82.33	173	29.36	12.40	-26.01	85.41

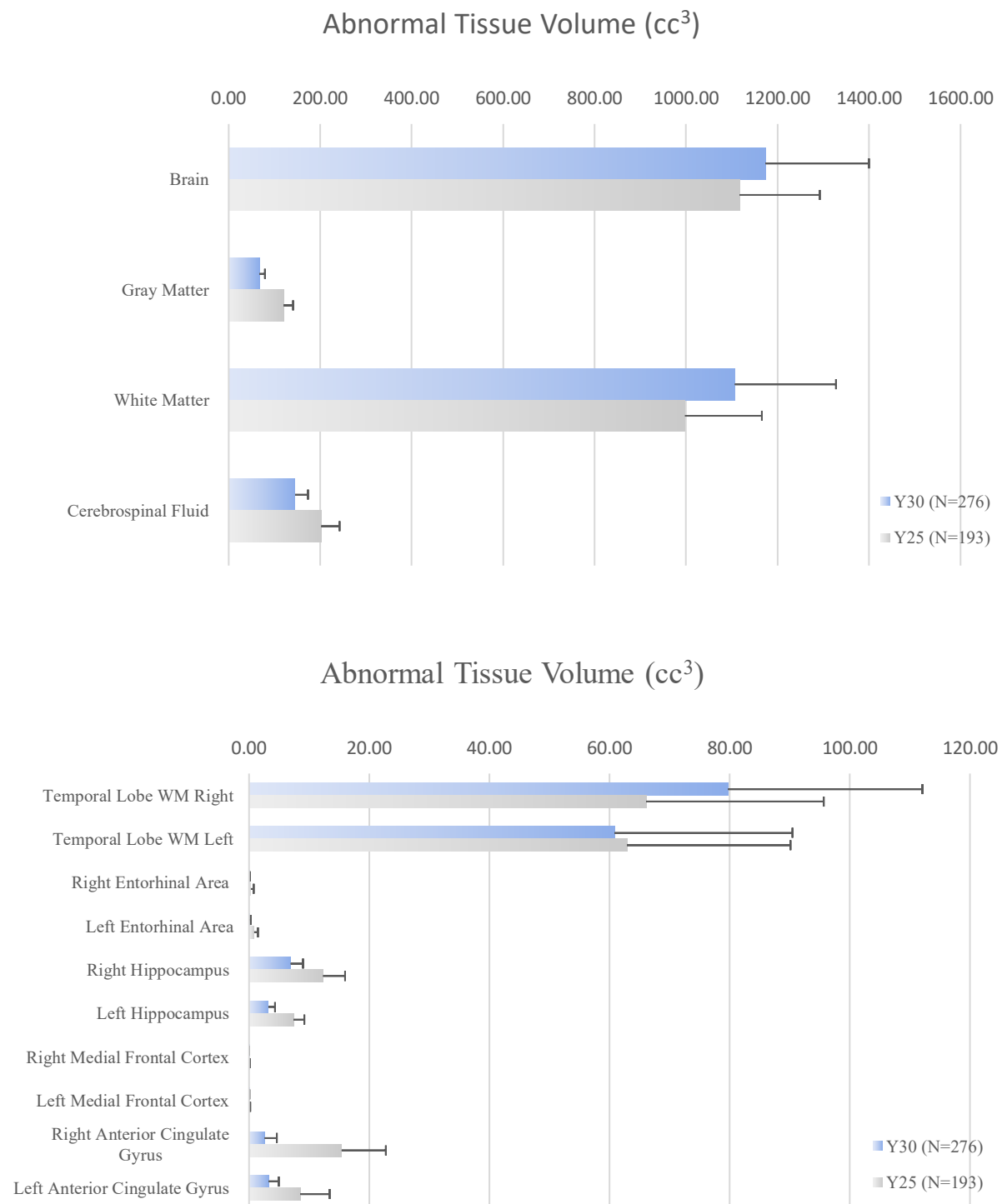
Appendix Figure 2 Bar charts of total brain volumes, abnormal tissue volumes, fractional anisotropy (normalized) and cerebral blood flow at Year 25 and Year 30



Error bar = Mean $\pm$ The 95% confidence intervals

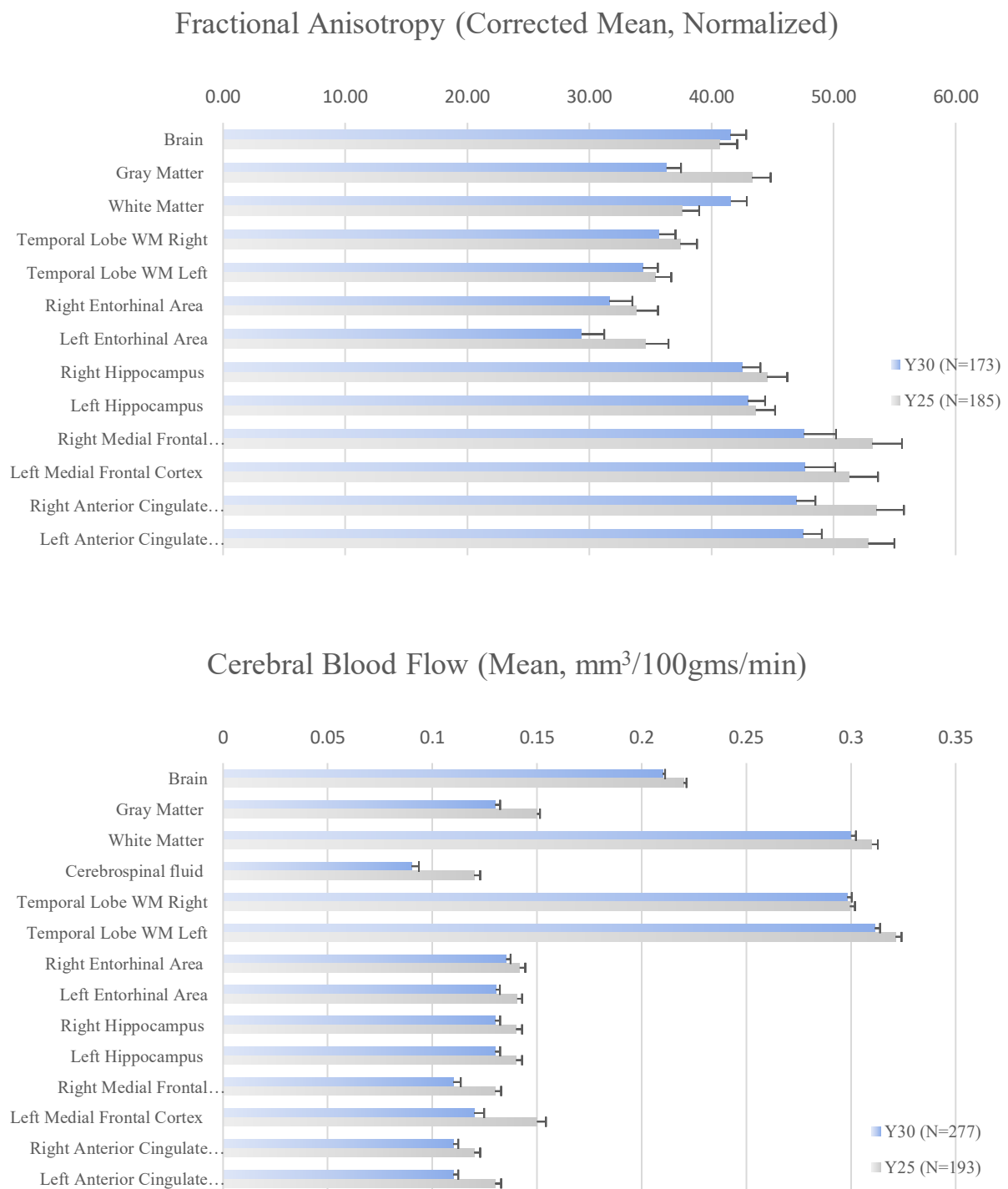
Intracranial volumes at Years 25 and 30 showed the largest volumes (cc³). Among the brain MRI parameters associated with successive stages of memory processing and early markers of AD, MFC showed the smallest volumes. Mean of entorhinal area right and left volumes are 2282.6 (1263, 3153.8) and 2288.5 (1111, 3412.3) at Year 30 (N=281), respectively.

(Figure Continued)



Abnormal WM tissue volumes are noticeable along with abnormal total brain tissue volumes (cc³). Mean of abnormal temporal lobe WM right tissue volume is apparently larger at Year 30 (79.7 (0, 3009.8), N=276) than the volume at Year 25 (66.1 (0, 2550.1), N=193).

(Figure Continued)



CBF in temporal lobe WM, and WM (mean, mm³/100gms/min) at Years 25 and 30 showed large values.

Appendix Table 8 Process of the probabilistic quotient normalisation (PQN) NMR data, normalized using the one-dimensional (1D) NMR data

(1) Calculate the Median of the intensities in each point spectrum

→ Median Y1, Y2, Y3 ...

ppm	10	9.99945	9.9989	...
Subject 1	253	407	322	
Subject 2	51	957	238	
...				
	Median Y1	Median Y2	Median Y3	...

(2) Divide the original intensity value by each Median Y1, Y2, Y3 ...

ppm	10	9.99945	9.9989	...
Subject 1	253/ Median Y1	407/ Median Y2	322/ Median Y3	
Subject 2	51/ Median Y1	957/ Median Y2	238/ Median Y3	
...				

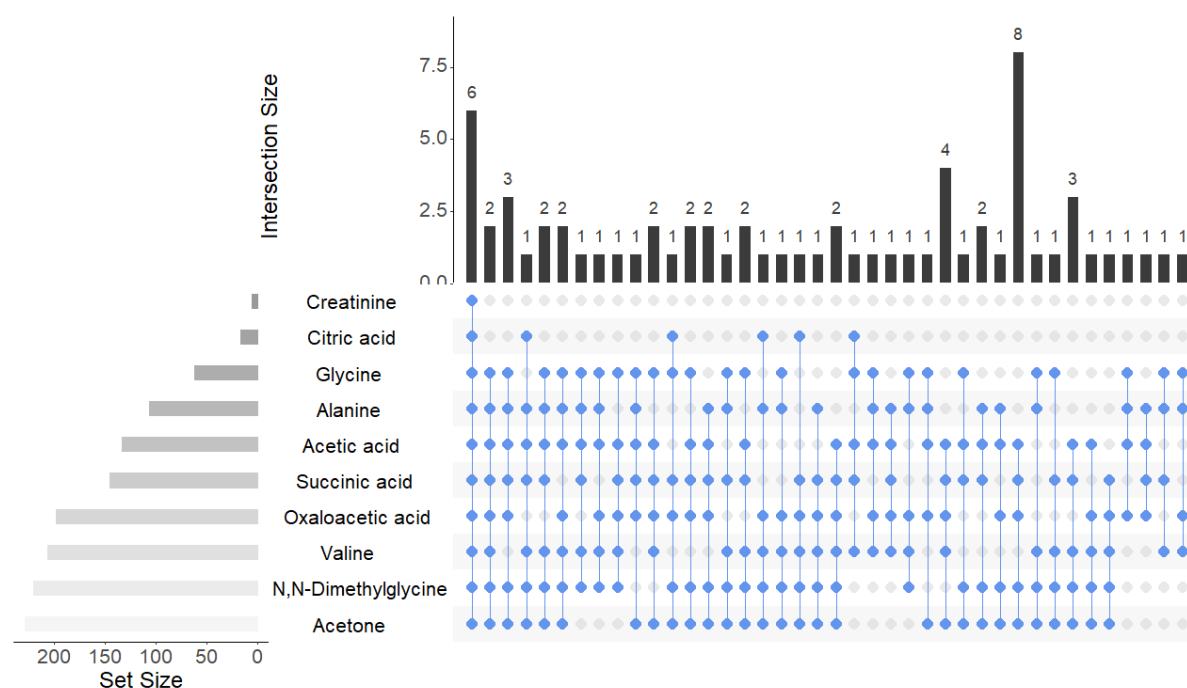
(3) Calculate the Median of the ratios per subject -> Median Z1, Z2, ...

ppm	10	9.99945	9.9989	...	Normalisation factor
Subject 1	253/ Median Y1	407/ Median Y2	322/ Median Y3		Median Z1
Subject 2	51/ Median Y1	957/ Median Y2	238/ Median Y3		Median Z2
...					

(4) Divide the quantified metabolite value by each Median Z1, Z2, ...

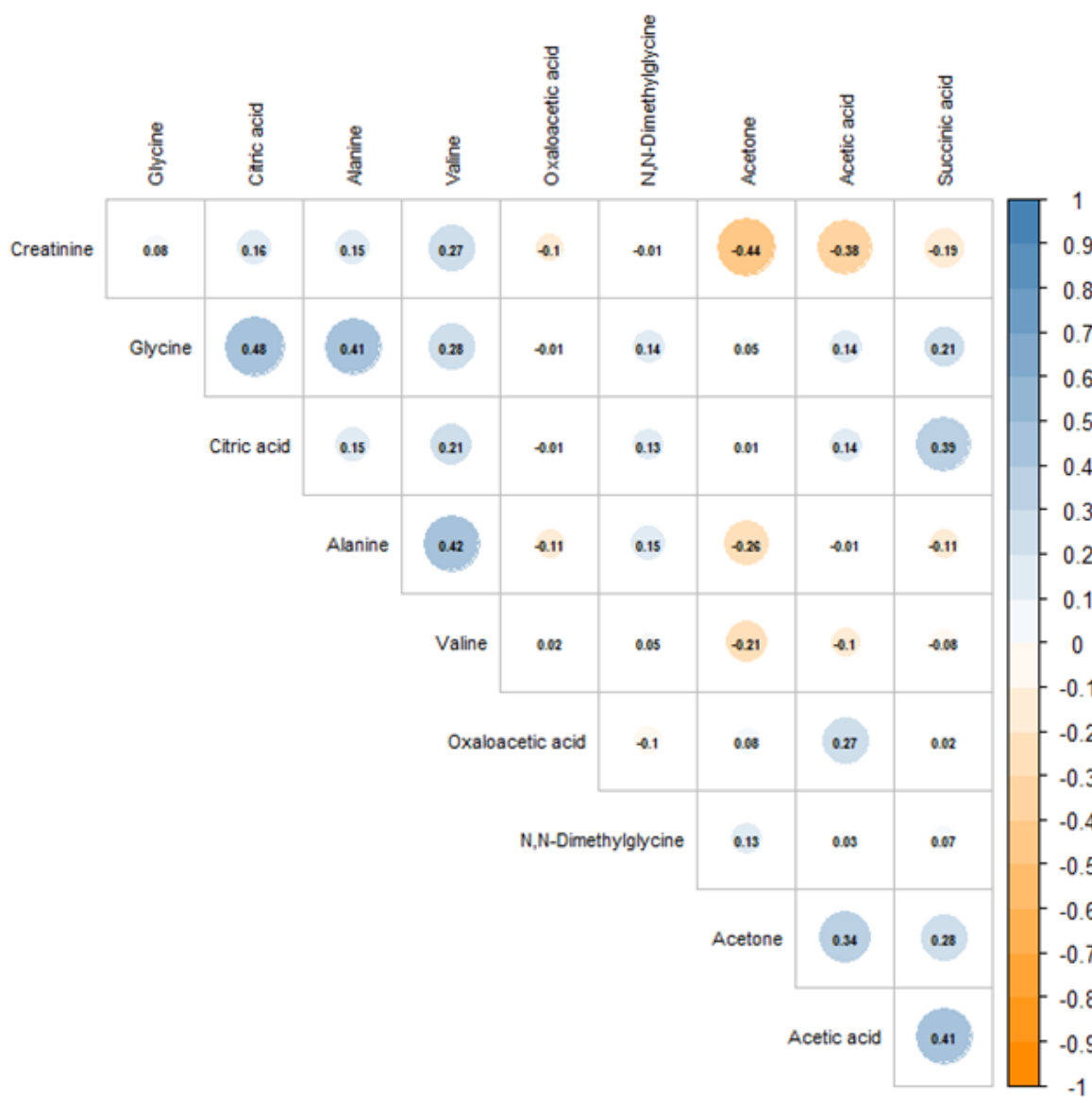
ppm	10	9.99945	9.9989	...
Subject 1	0.4/ Median Z1	0.2/ Median Z1	0.7/ Median Z1	
Subject 2	0.1/ Median Z2	0.3/ Median Z2	0.5/ Median Z2	
...				

Appendix Figure 3 Visualization of intersecting sets – Missing cases of the metabolites by NMR

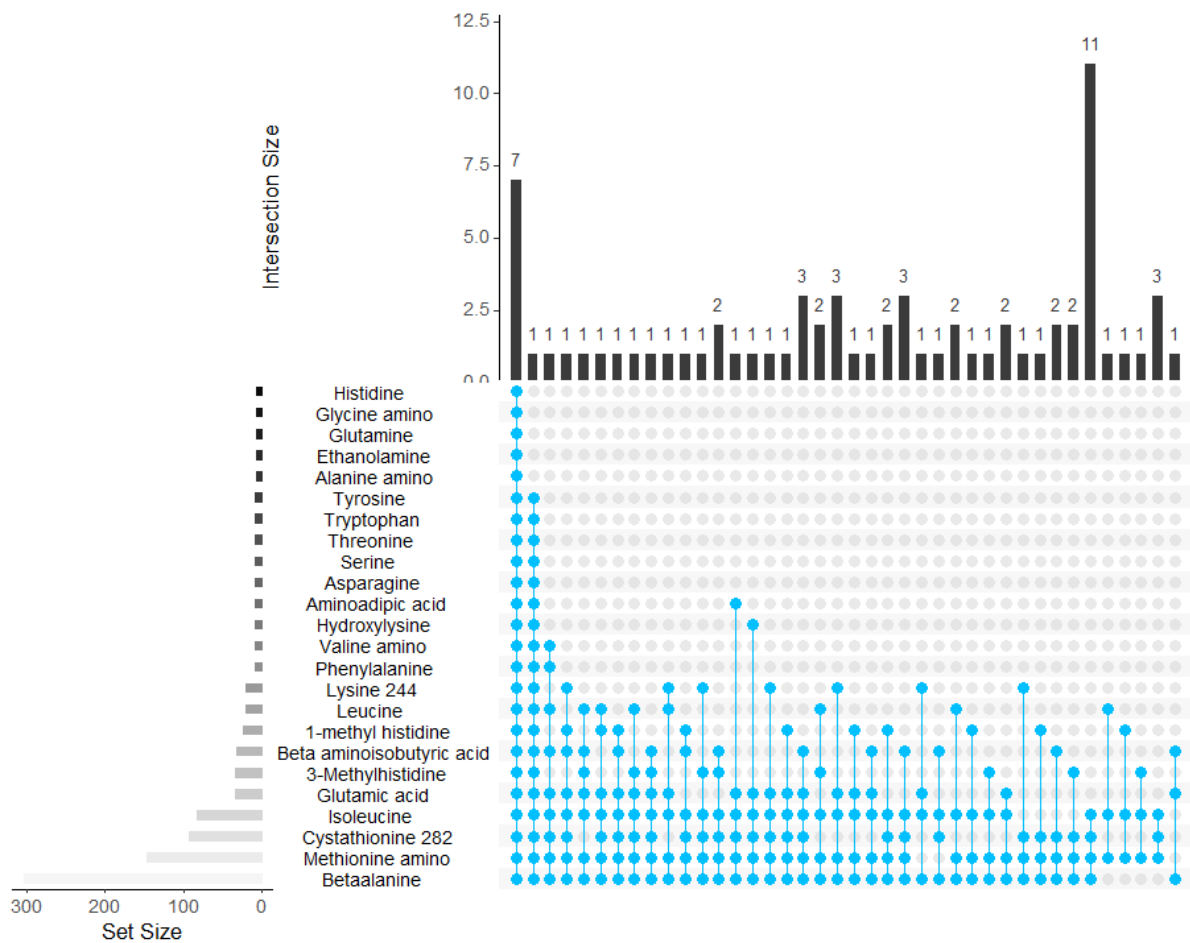


The number of individuals with missing cases in 10 metabolites N (%), up to 2 missing are 398 (66.3), up to 6 missing are 590 (98.3), 7 missing are 8 (1.3) and 8 missing are 2 (0.3) cases. Creatinine has 6 missing cases in 606 participants, and all other 9 metabolites (which the missing cases are not exceed 50%) shared those missing cases (i.e. individuals) as reported in Appendix Figure 3. Among the 600, 83 individuals have 36 missing NMR metabolites (13.83); 71 individuals have 37 missing NMR metabolites (11.83); 64 individuals have 38 missing.

Appendix Figure 4 The upper triangle of a Spearman's rank correlation coefficient matrix – Metabolites by NMR

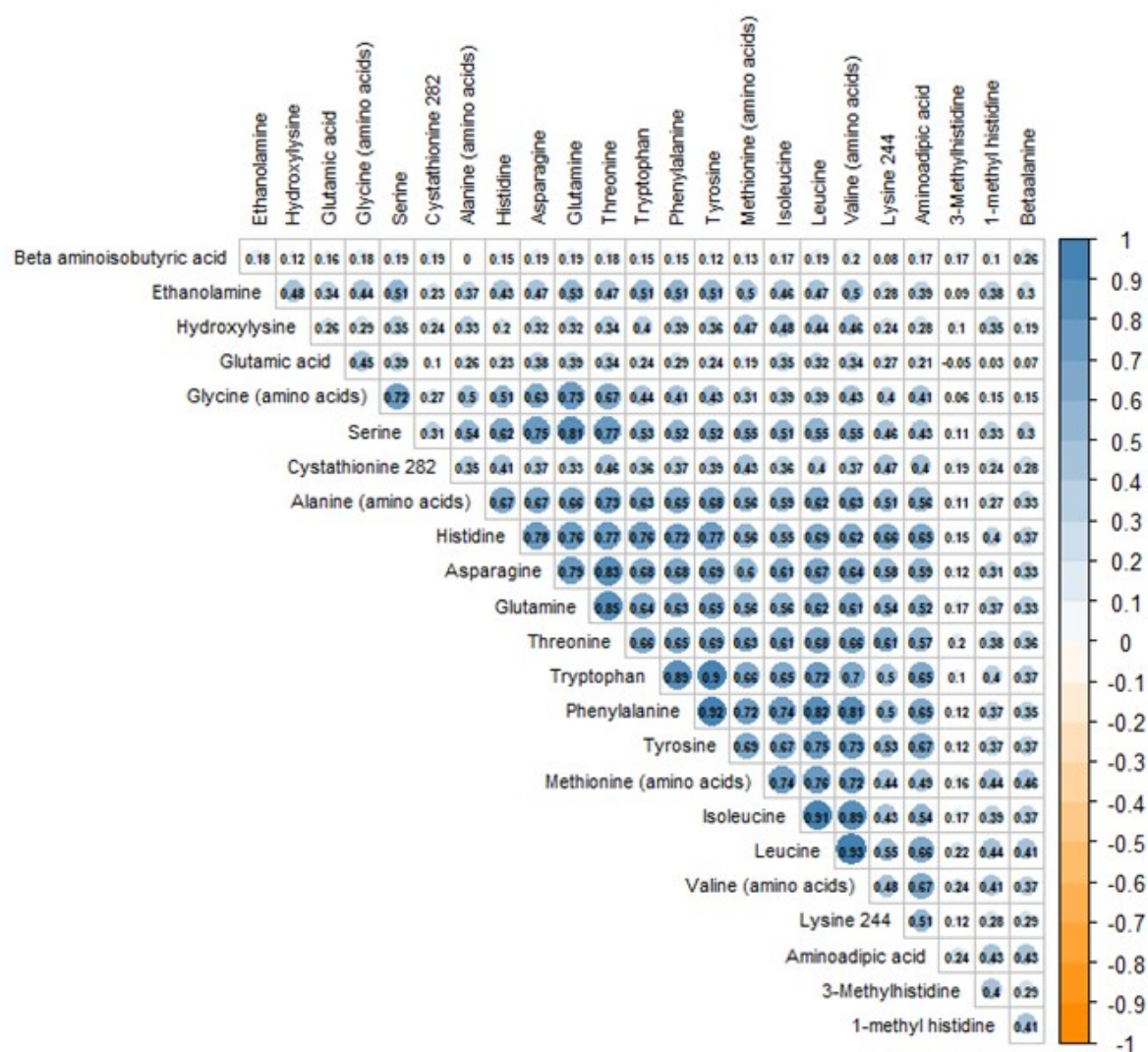


Appendix Figure 5 Visualization of intersecting sets – Missing cases of amino acids by LC–MS

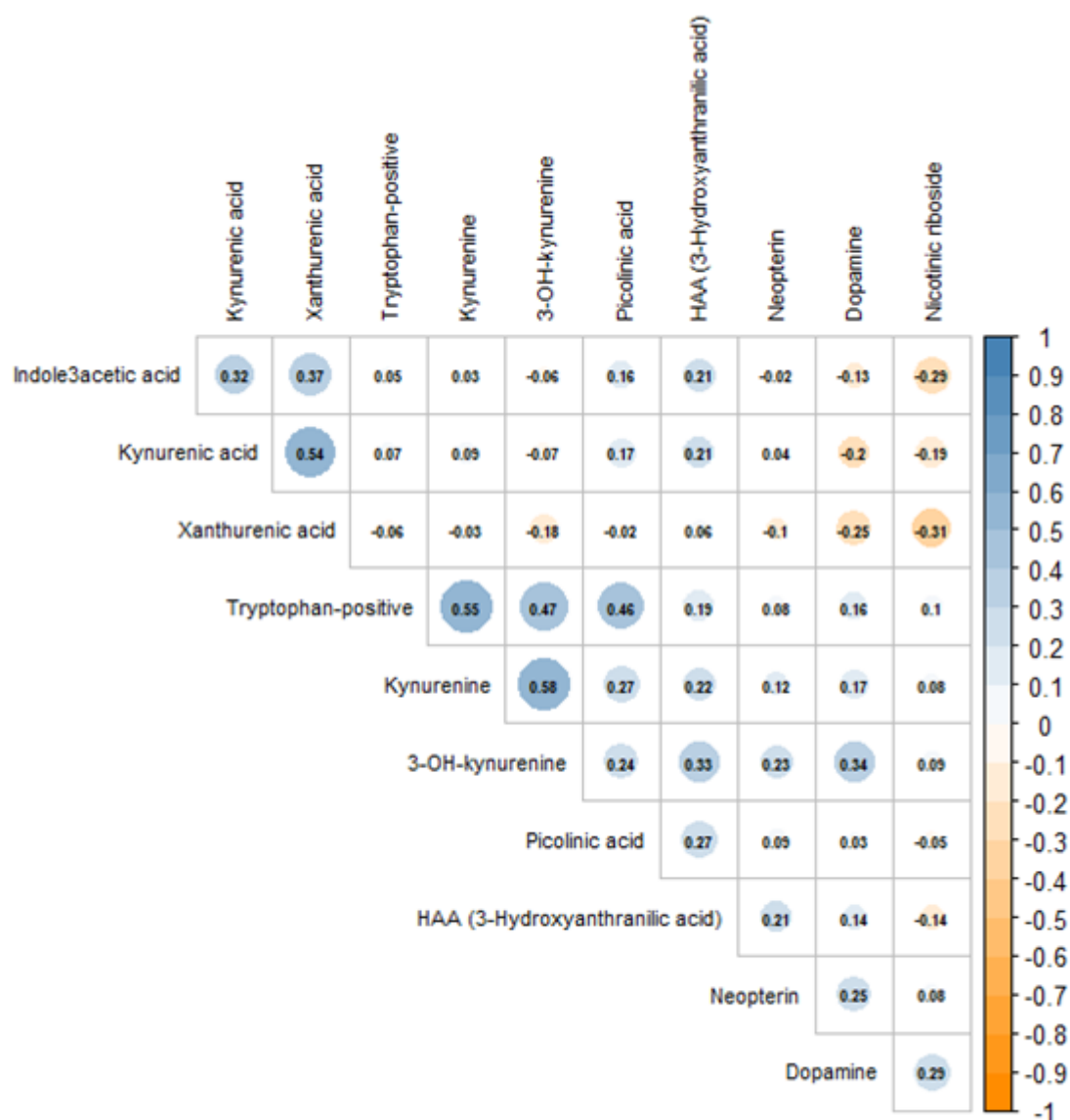


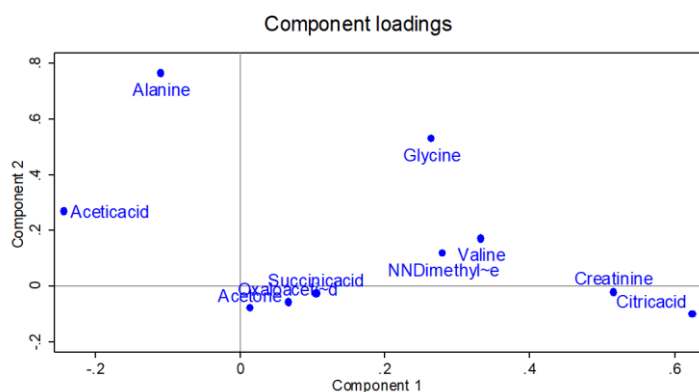
Histidine, glycine, glutamine, ethanolamine, and alanine have 7 missing cases in the total study population (N=606).

Appendix Figure 6 The upper triangle of a Spearman's rank correlation coefficient matrix – Amino acids by LC–MS

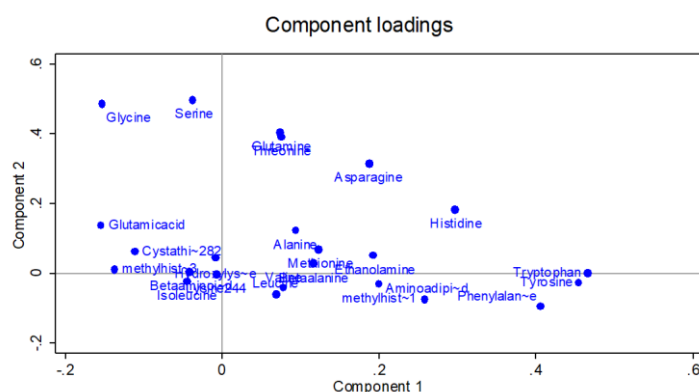
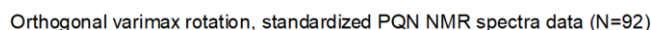


Appendix Figure 8 The upper triangle of a Spearman's rank correlation coefficient matrix – Tryptophan pathway metabolites by LC–MS

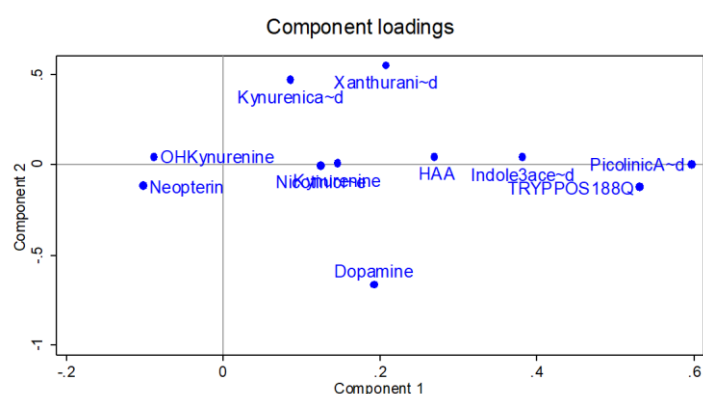
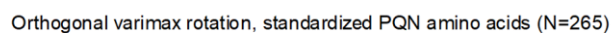




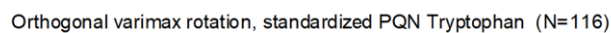
Appendix Figure 9 PCA of the probabilistic quotient normalisation (PQN) applied NMR spectra data (10 metabolites < 50% missing cases)

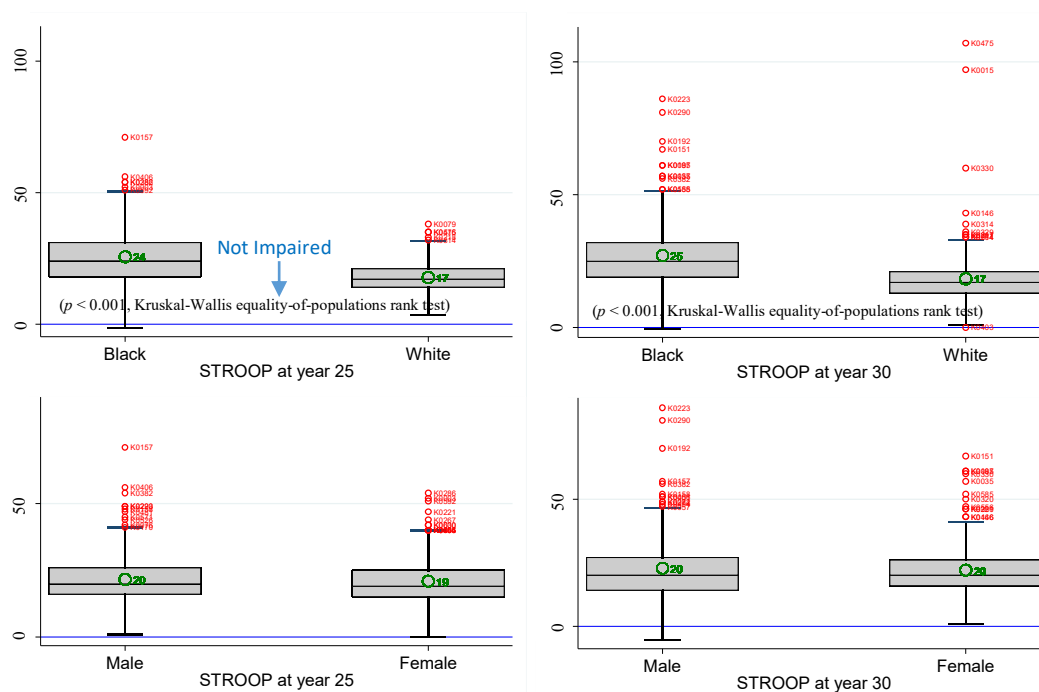
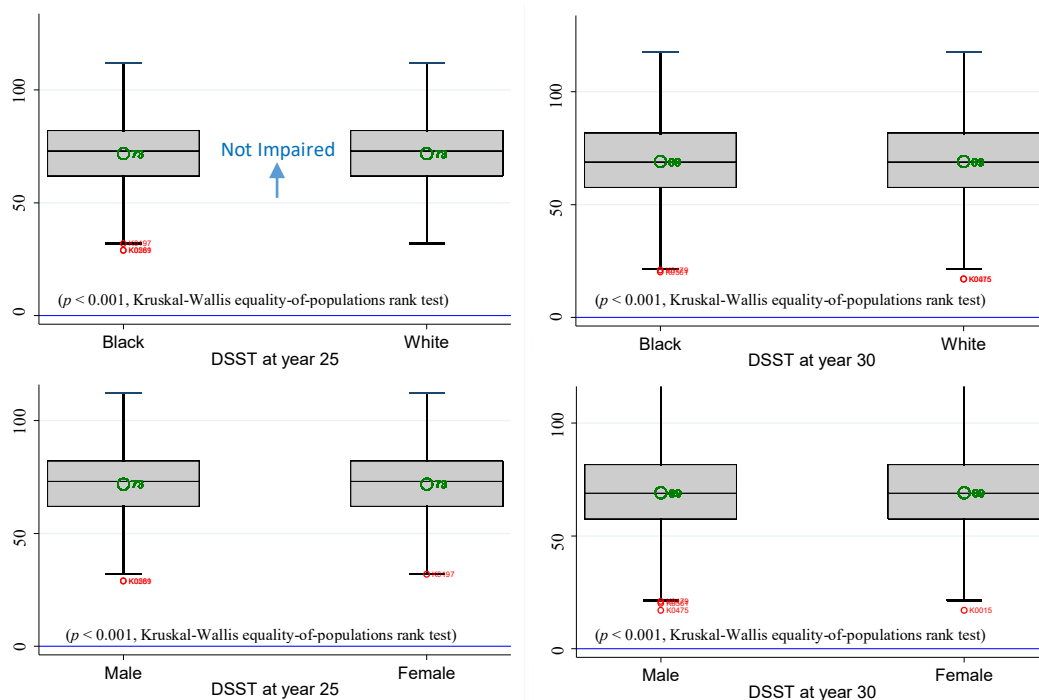


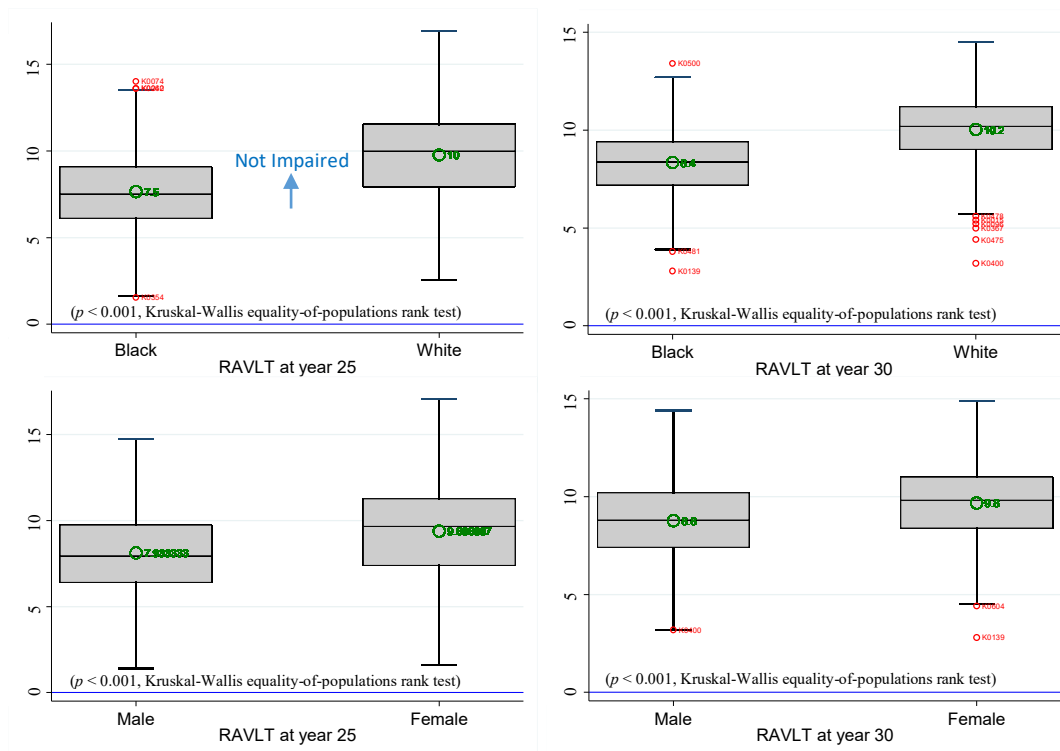
Appendix Figure 10 PCA of the probabilistic quotient normalisation (PQN) applied amino acids (24 metabolites < 50% missing cases).



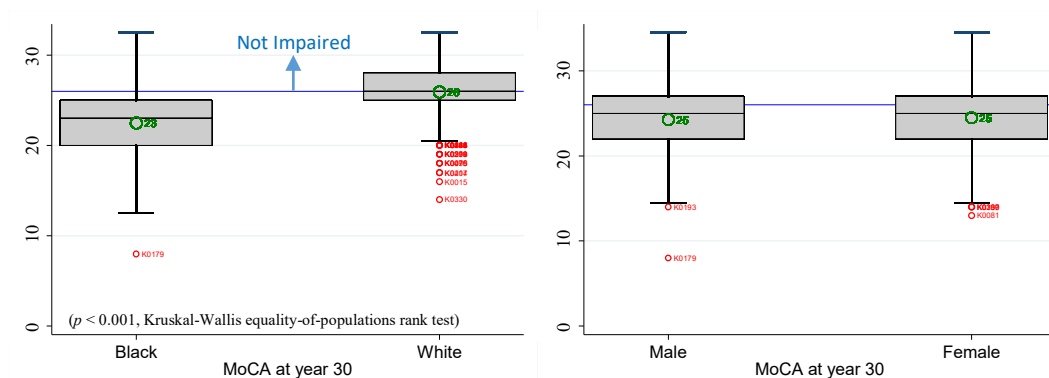
Appendix Figure 11 PCA of the probabilistic quotient normalisation (PQN) applied amino acids (24 metabolites < 50% missing cases).



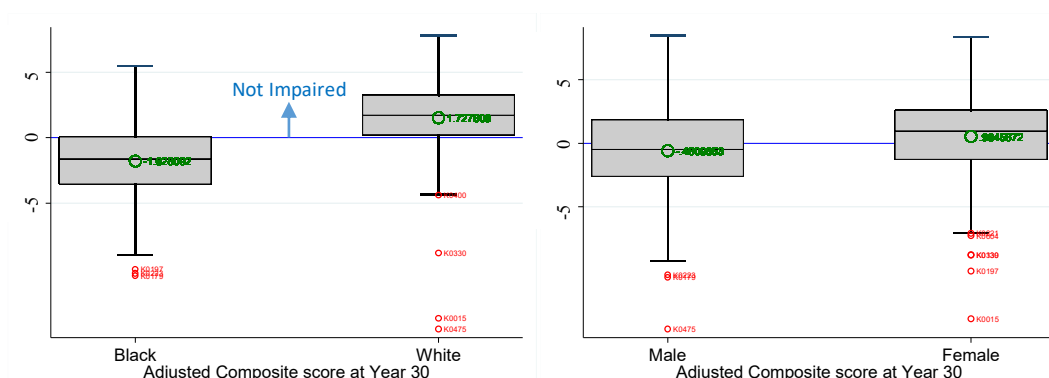




Appendix Figure 14 The RAVLT at Y25 and Y30 (by race and sex)



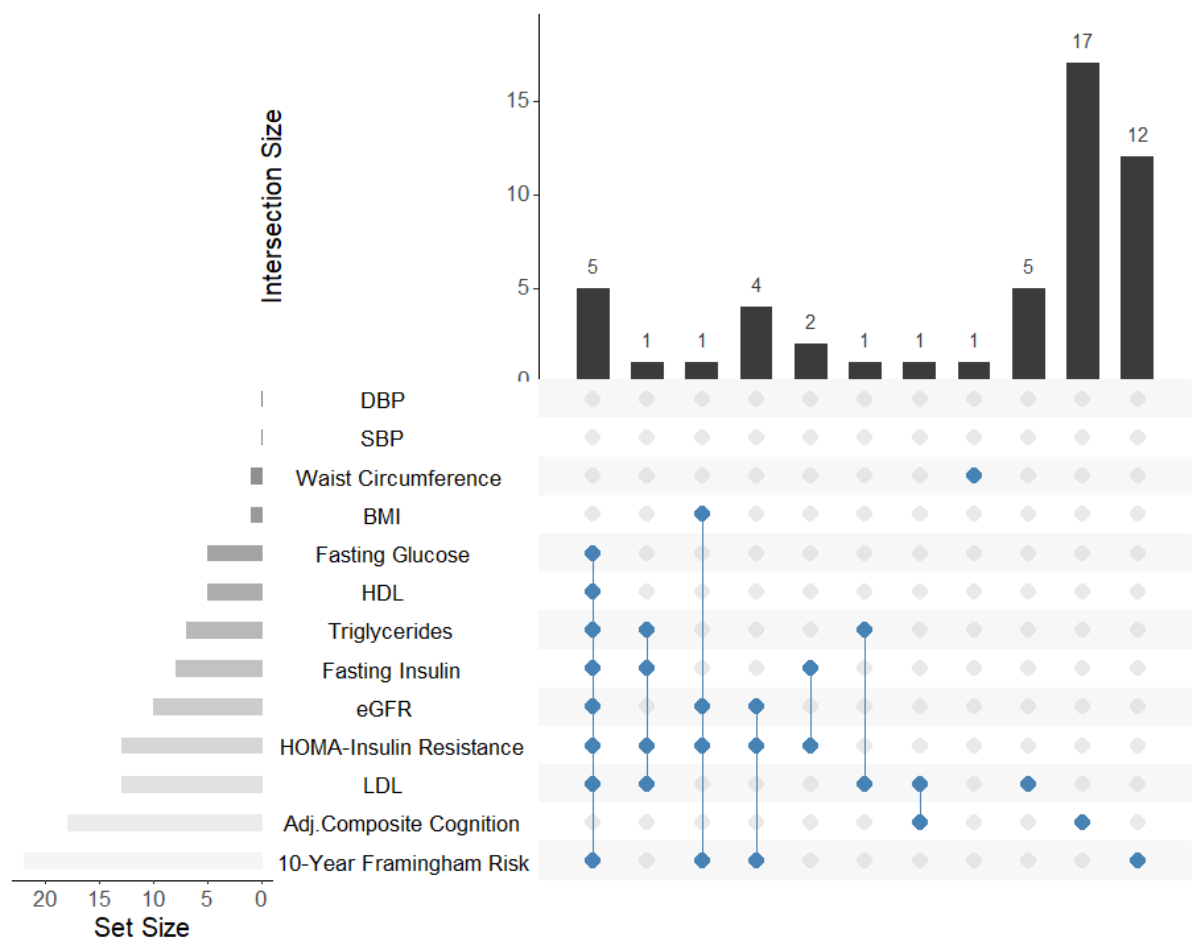
Appendix Figure 15 The MoCA at Y30 (by race and sex)

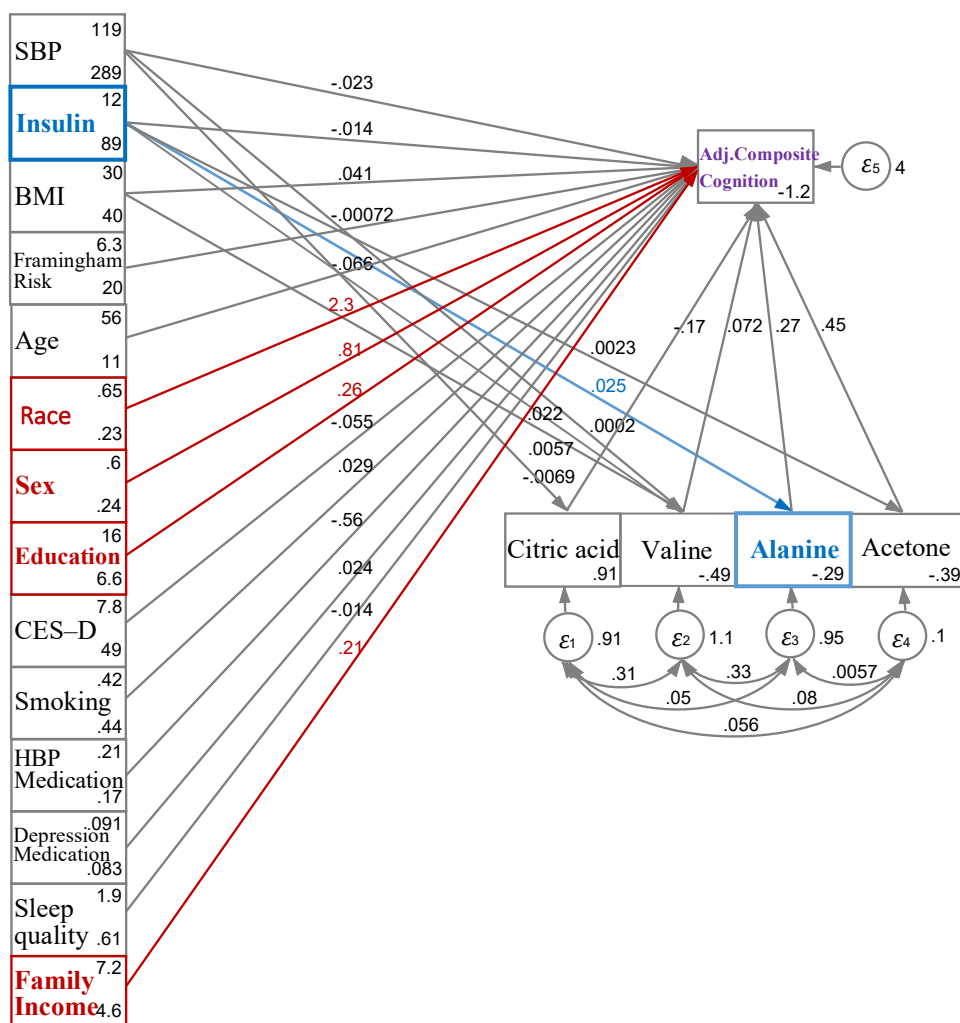


Appendix Figure 16 Adjusted Composite cognition at Y30 (by race and sex)

Appendix Figure 17 Visualization of intersecting sets - Missing cases of cardio-vascular risk factors and the adjusted composite cognitive score

The adjusted composite cognitive score has 18 missing cases (out of 606 participants), and it shares 1 missing case with LDL. Fasting glucose, HDL, Triglycerides, fasting insulin, eGFR, HOMA-IR, LDL, and 10-Year Framingham Risk share 5 missing cases.





Appendix Figure 18 Total multivariate regression path diagram for the adjusted composite cognitive score on cardio-metabolic risk factors (SBP, Insulin, BMI, and Framingham risk) with NMR metabolites (N=187)

Note: The variables inside the rectangles are observed variables in this path diagram. The figures inside the rectangles are mean and variance of the variables. For instance, 119 and 289 indicate mean and variance of SBP. Paths are direct relationships between variables. Estimated path coefficients are represented by straight arrows. Curved arrows specify that error terms covariance.

Appendix Table 9 Structural equation model for the adjusted composite cognitive score on cardio-metabolic risk factors (SBP, Insulin, BMI, and Framingham risk) with the metabolites by NMR

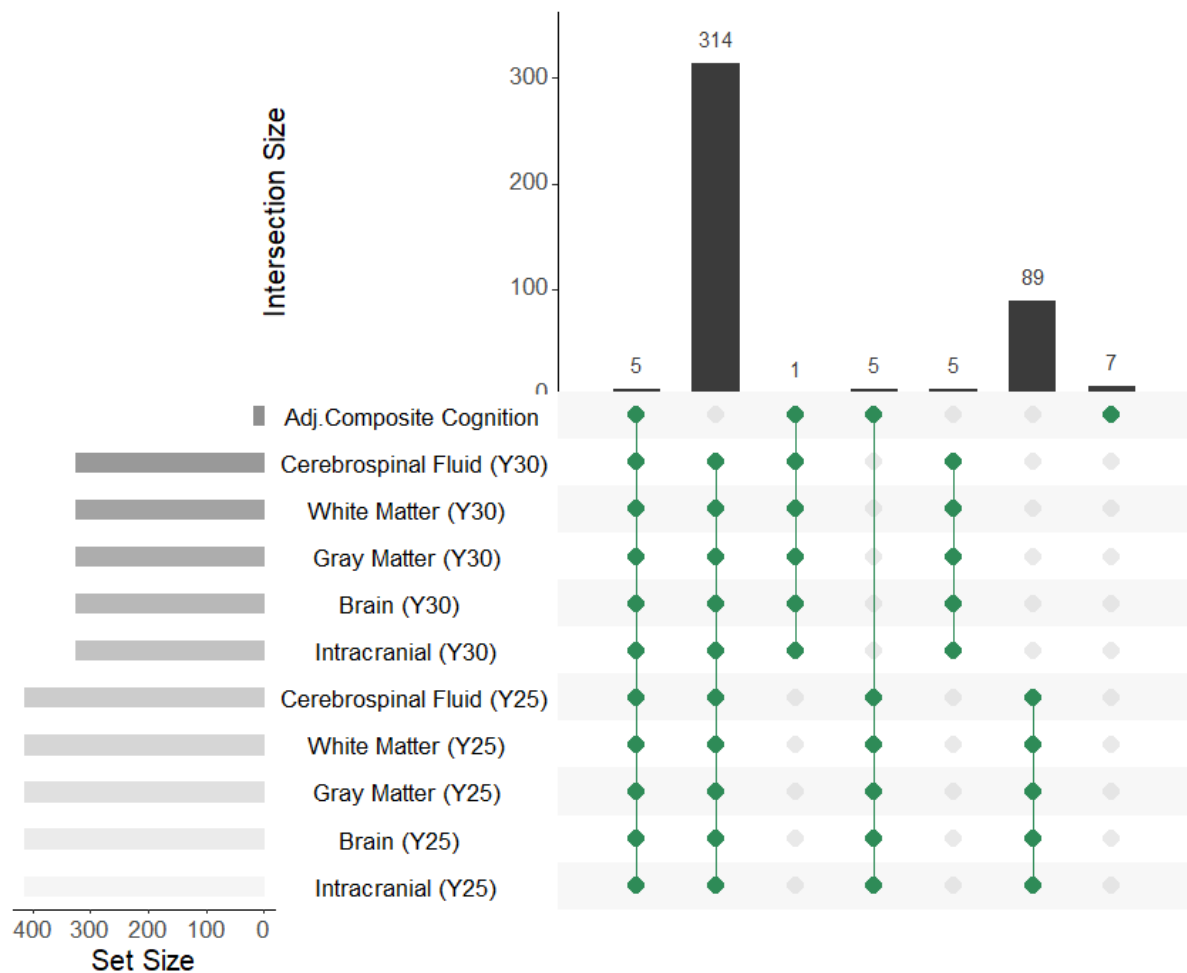
Structural equation model
Log likelihood = -6340.7313

Number of observations = 187
Replications = 500

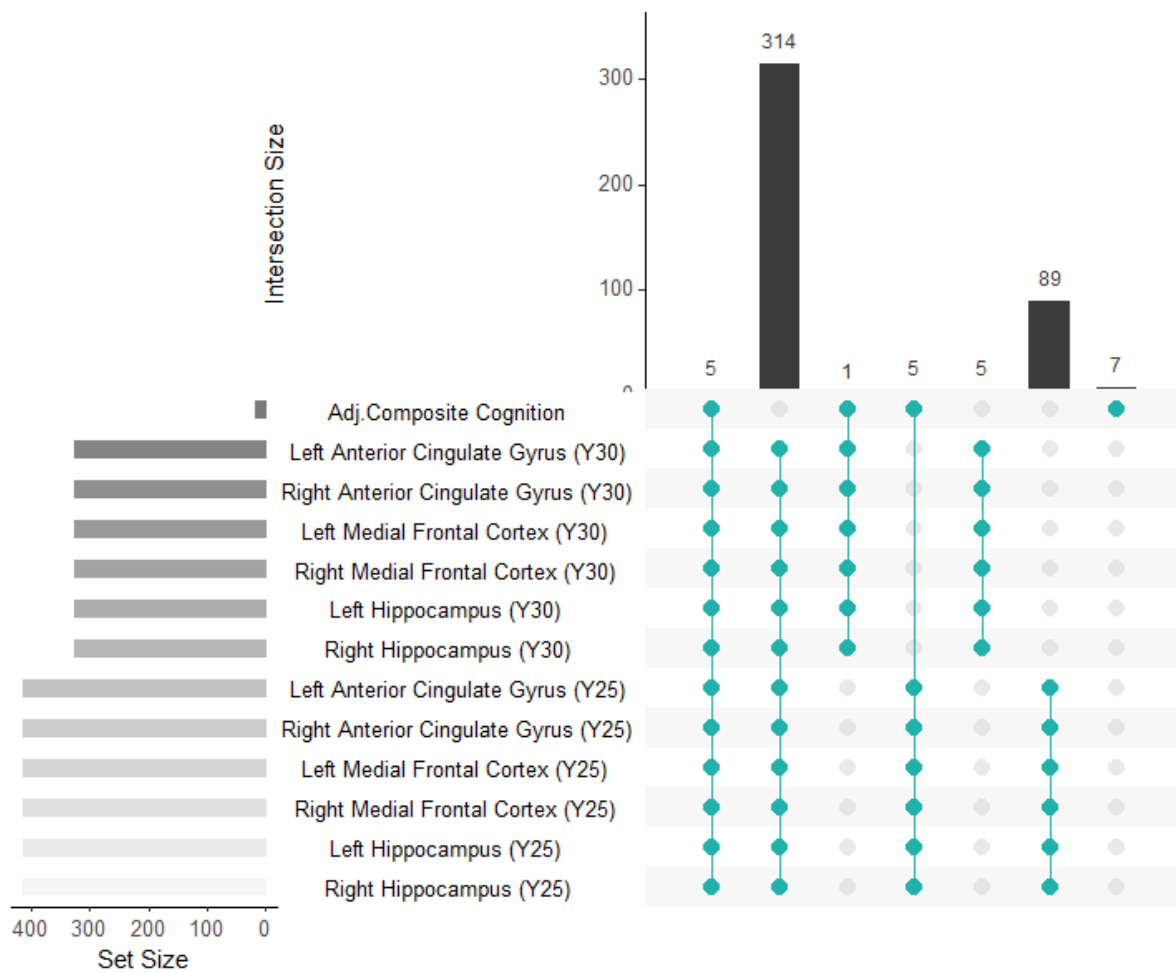
		Observed Coef.	Bootstrap Std. Err.	z	P > z	Normal-based [95% CIs]	
Adj. Composite Cognition ←	Citric acid	-0.171	0.19	-0.89	0.37	-0.55	0.21
	Valine	0.072	0.17	0.42	0.67	-0.26	0.40
	Alanine	0.270	0.16	1.64	0.10	-0.05	0.59
	Acetone	0.449	0.40	1.12	0.27	-0.34	1.24
	SBP	-0.023	0.01	-1.83	0.07	-0.05	0.002
	Insulin	-0.014	0.02	-0.57	0.57	-0.06	0.03
	BMI	0.041	0.03	1.28	0.20	-0.02	0.10
	Framingham risk	-0.0007	0.05	-0.02	0.99	-0.09	0.09
	Age	-0.066	0.05	-1.42	0.16	-0.16	0.03
	Race	2.287	0.45	5.04	< 0.001	1.40	3.18
	Sex	0.813	0.39	2.10	0.04	0.06	1.57
	Education	0.261	0.10	2.62	0.01	0.07	0.46
	CES-D	-0.055	0.04	-1.32	0.19	-0.14	0.03
	Smoking	0.029	0.28	0.11	0.92	-0.51	0.57
	HBP medication	-0.557	0.41	-1.35	0.18	-1.36	0.25
	Depression medication	0.024	0.70	0.03	0.97	-1.34	1.39
	Sleep quality	-0.014	0.24	-0.06	0.95	-0.48	0.45
	Family income	0.209	0.10	2.08	0.04	0.01	0.40
	constant	-1.207	3.96	-0.31	0.76	-8.96	6.55
Citric acid ←	SBP	-0.007	0.00	-1.50	0.13	-0.02	0.002
	constant	0.912	0.54	1.69	0.09	-0.15	1.97
Valine ←	SBP	0.000	0.00	0.04	0.97	-0.01	0.01
	Insulin	0.022	0.01	1.74	0.08	-0.003	0.05
	BMI	0.006	0.01	0.43	0.67	-0.02	0.03
	constant	-0.488	0.63	-0.78	0.44	-1.72	0.74
Alanine ←	Insulin	0.025	0.01	2.75	0.01	0.01	0.04
	constant	-0.288	0.12	-2.41	0.02	-0.52	-0.05
Acetone ←	Insulin	0.002	0.00	1.04	0.30	-0.002	0.01
	constant	-0.392	0.04	-11.02	< 0.001	-0.46	-0.32
var (e.Adj. Composite Cognition)		4.046	0.94			2.57	6.37
var (e.Citric acid)		0.914	0.11			0.71	1.17
var (e.Valine)		1.143	0.12			0.93	1.40
var (e.Alanine)		0.951	0.26			0.55	1.63
var (e.Acetone)		0.102	0.01			0.08	0.13
cov (e.Citric acid, e.Valine)		0.308	0.07	4.13	< 0.001	0.16	0.45
cov (e.Citric acid, e.Alanine)		0.050	0.07	0.73	0.46	-0.08	0.18
cov (e.Citric acid, e.Acetone)		0.056	0.02	2.56	0.01	0.01	0.10
cov (e.Valine, e.Alanine)		0.328	0.08	4.08	< 0.001	0.17	0.49
cov (e.Valine, e.Acetone)		0.080	0.03	3.03	0.002	0.03	0.13
cov (e.Alanine, e.Acetone)		0.006	0.03	0.20	0.84	-0.05	0.06

Appendix Figure 19 Visualization of intersecting sets - Missing cases of the total brain volumes and the adjusted composite cognitive score

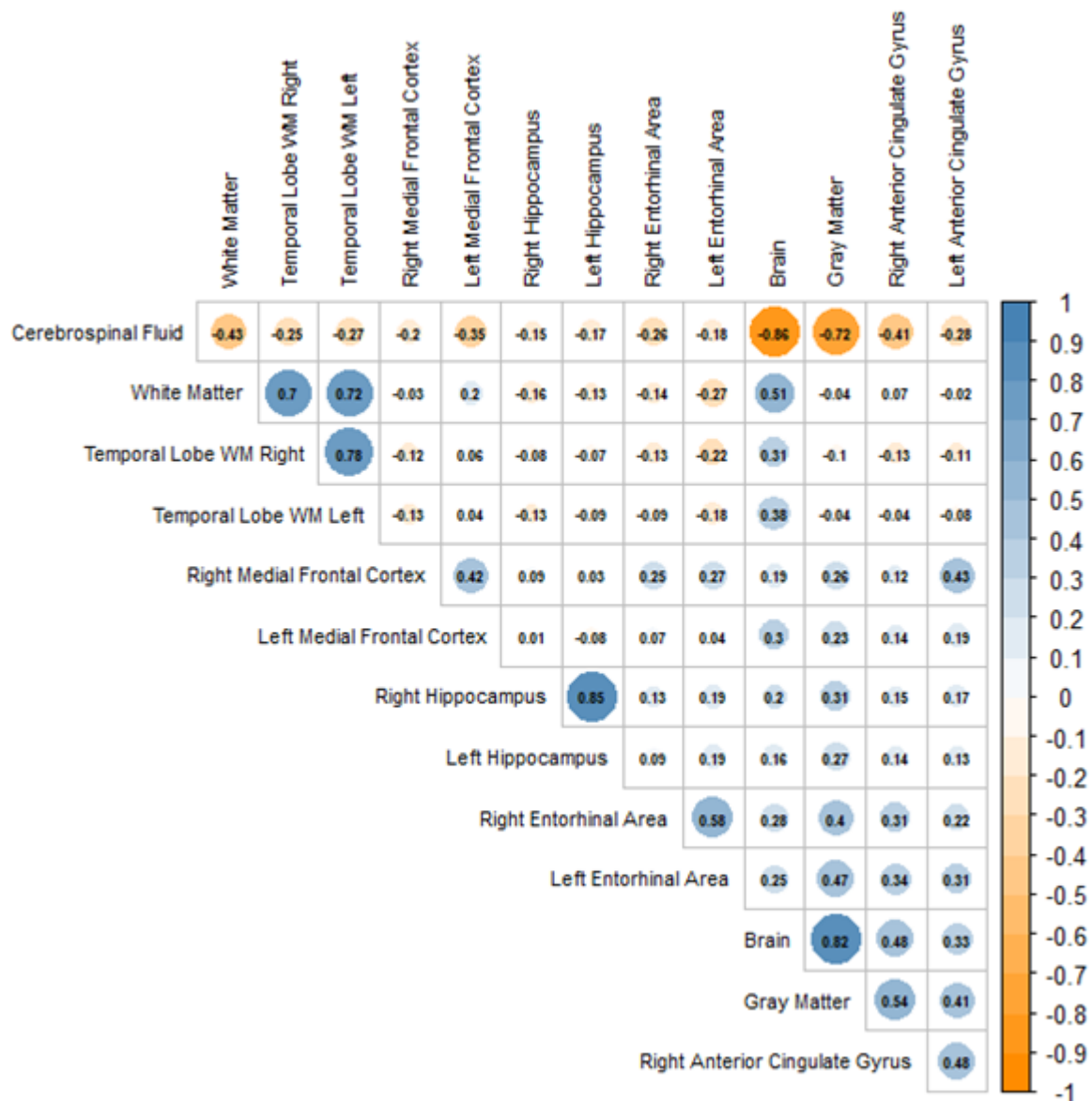
The adjusted composite cognitive score has 18 missing cases (out of 606 participants). Cerebrospinal fluid, white matter, gray matters, brain and intracranial volumes at Year 30 share 314 missing cases measured at Year 25. Eighty-nine more missing cases are in cerebrospinal fluid, white matter, gray matters, brain and intracranial volumes measured at Year 25.



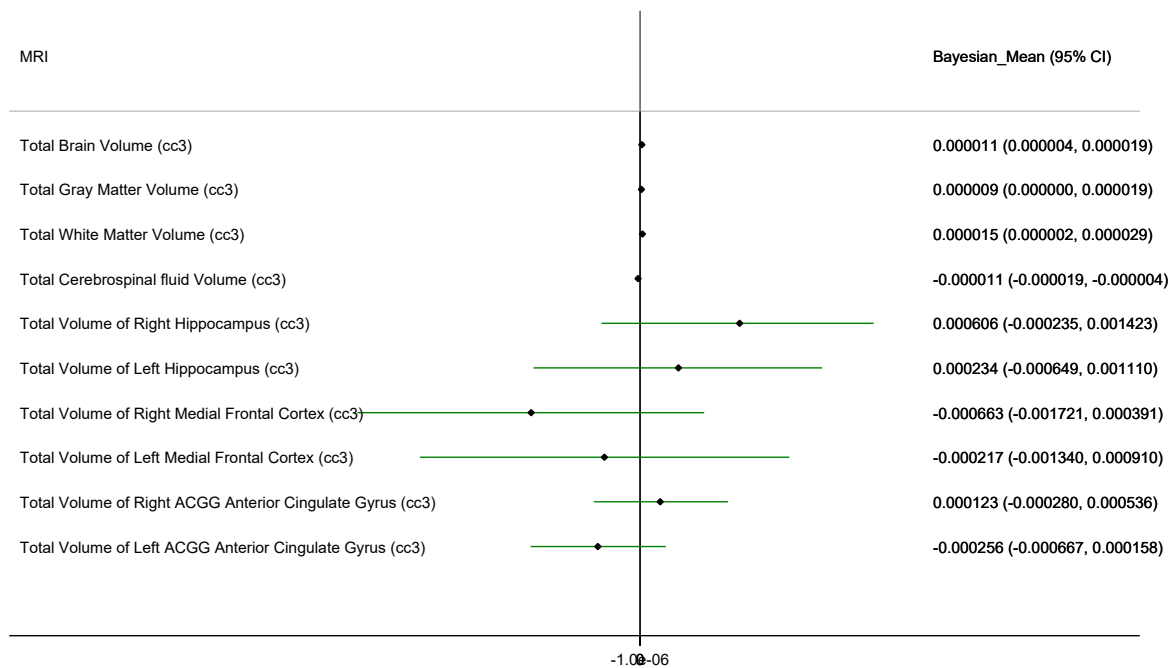
Appendix Figure 20 Visualization of intersecting sets - Missing cases of the brain structure (volumes) involved in the successive stages of memory processing and the adjusted composite cognitive score



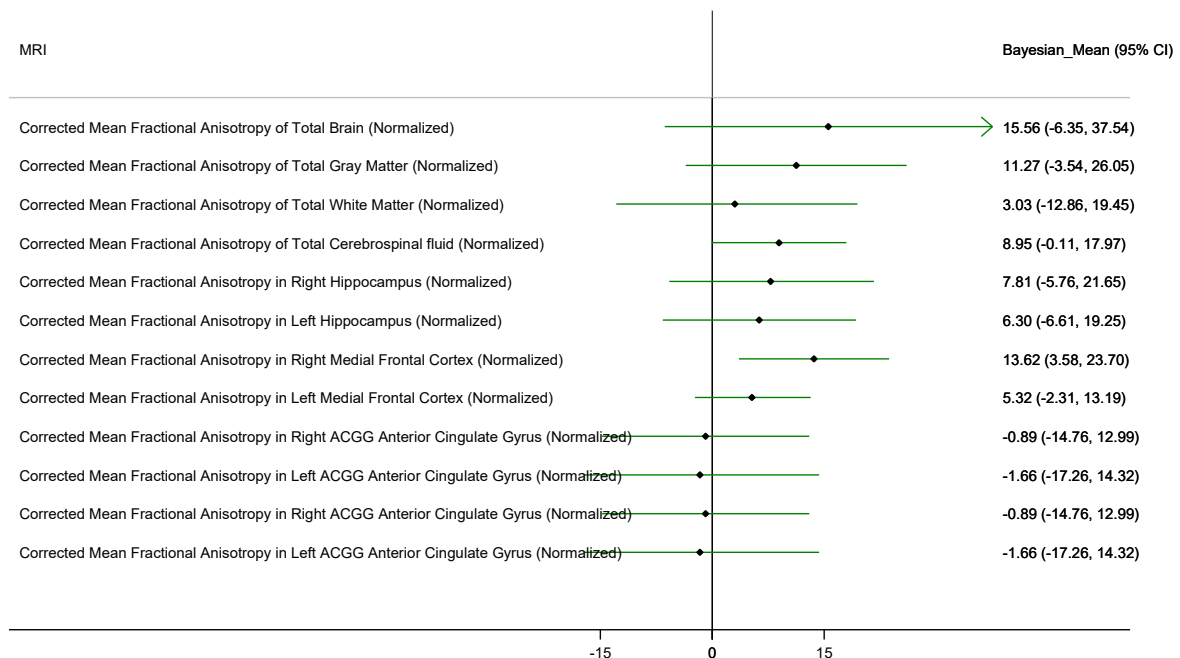
Appendix Figure 21 The upper triangle of a partial Spearman's rank correlation coefficient matrix – ICV adjusted brain MRI parameters



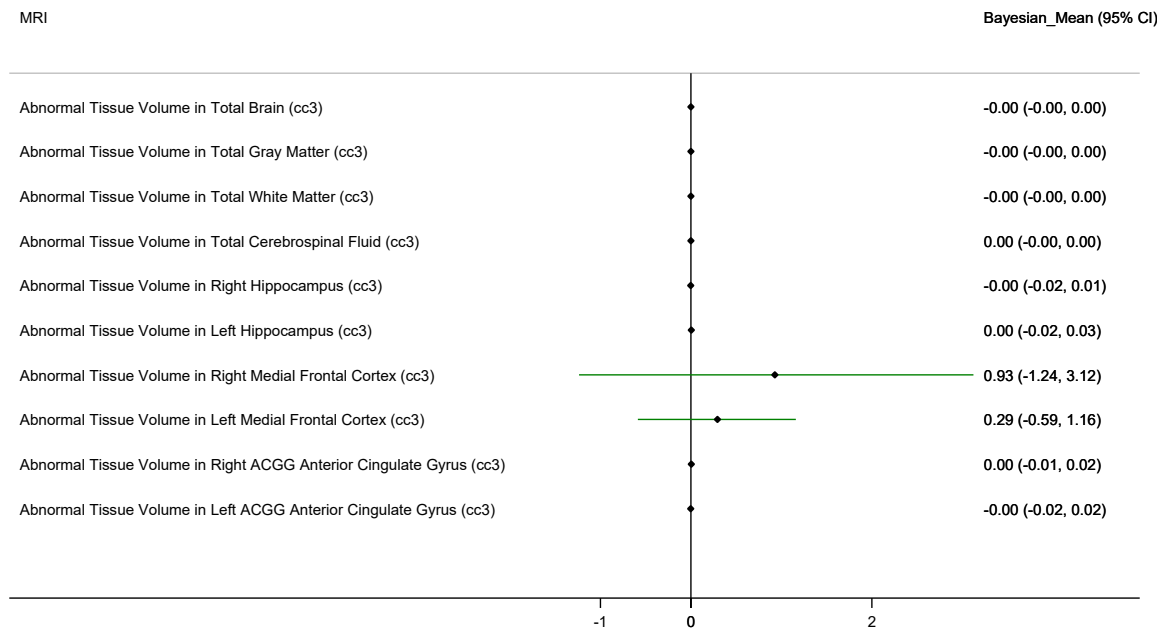
Appendix Figure 22 Total brain volume (cc³) regarding adjusted composite cognition using the Bayesian three-level random-intercept regression at Year 30



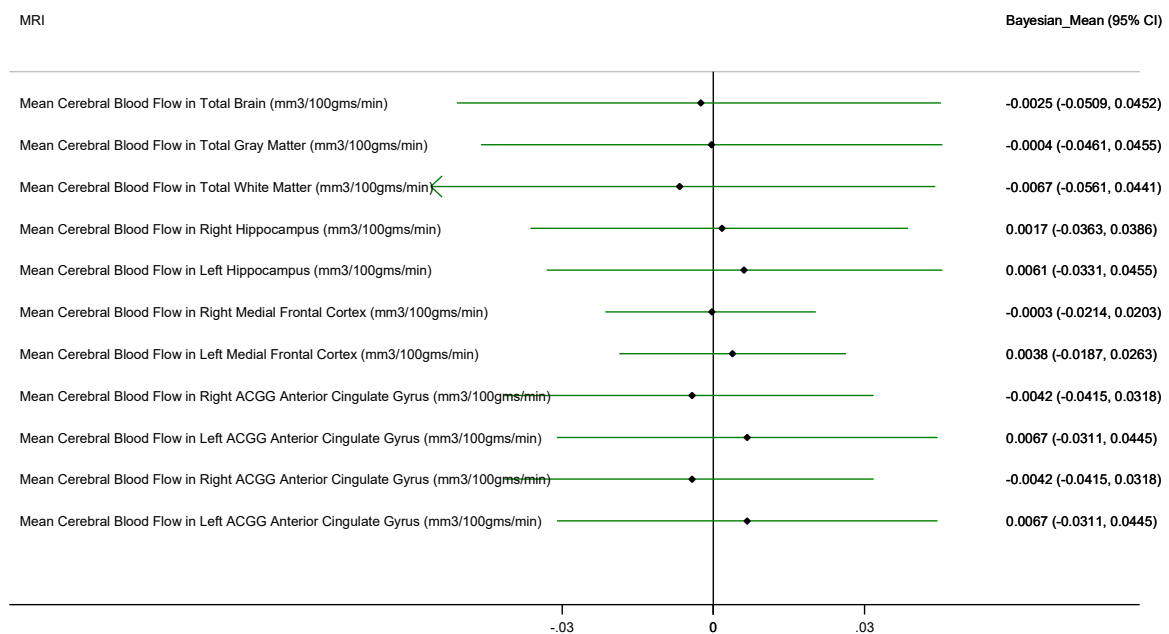
Appendix Figure 23 Corrected fractional anisotropy (normalized) regarding adjusted composite cognition using the Bayesian three-level random-intercept regression at Year 30



Appendix Figure 24 Abnormal tissue volume (cc³) regarding adjusted composite cognition using the Bayesian three-level random-intercept regression at Year 30



Appendix Figure 25 Mean cerebral blood flow (mm³/100gms/min) regarding adjusted composite cognition using the Bayesian three-level random-intercept regression at Year 30



Appendix Table 10 Structural equation model for the adjusted composite cognitive score on cardio-metabolic risk factors (SBP, Fasting glucose, WC, and Framingham risk) with brain MRI parameters

Structural equation model
Log likelihood = -9919.213

Number of observations = 155
Replications = 500

		Observed Coef.	Bootstrap Std. Err.	z	P > z	Normal-based [95% CIs]	
Adj. Composite Cognition ←	Total Intracranial Volume	-20.937	19.51	-1.07	0.28	-59.17	17.29
	Total Cerebrospinal Fluid Volume	7.372	6.94	1.06	0.29	-6.23	20.97
	Total White Matter Volume	9.741	8.34	1.17	0.24	-6.60	26.09
	Total Gray Matter Volume	9.800	8.46	1.16	0.25	-6.79	26.39
	Total Right Medial Frontal Cortex Volume	0.020	0.28	0.07	0.94	-0.52	0.56
	Total Right ACG Volume	-0.007	0.29	-0.02	0.98	-0.58	0.57
	Abnormal Brain Tissue Volume	-0.158	5.01	-0.03	0.98	-9.98	9.66
	Abnormal White Matter Tissue Volume	-0.516	5.07	-0.10	0.92	-10.45	9.42
	CBF in Brain	-27.743	29.93	-0.93	0.35	-86.40	30.91
	CBF in White Matter	11.989	13.44	0.89	0.37	-14.35	38.32
	CBF in Gray Matter	15.106	16.93	0.89	0.37	-18.08	48.29
	CBF in Left Hippocampus	0.185	0.46	0.40	0.69	-0.72	1.09
	CBF in Right Medial Frontal Cortex	-0.480	0.40	-1.21	0.23	-1.25	0.30
	CBF in Right ACG	0.252	0.84	0.30	0.76	-1.39	1.89
	CBF in Left ACG	0.200	0.64	0.31	0.76	-1.06	1.46
	FA in White Matter	0.310	0.60	0.52	0.61	-0.86	1.48
	FA in Left ACG	0.058	0.33	0.18	0.86	-0.58	0.70
	FA in Right ACG	0.092	0.33	0.28	0.78	-0.55	0.74
	ATV in Temporal Lobe WM Right	0.297	0.69	0.43	0.67	-1.06	1.66
	ATV in Temporal Lobe WM Left	-0.735	1.18	-0.63	0.53	-3.04	1.57
	FA in Temporal Lobe WM Right	0.591	0.62	0.95	0.34	-0.63	1.81
	FA in Temporal Lobe WM Left	-0.842	0.59	-1.44	0.15	-1.99	0.31
	Total Temporal Lobe WM Left Volume	-0.054	1.00	-0.05	0.96	-2.02	1.91
	Total Temporal Lobe WM Right Volume	-0.791	0.95	-0.83	0.40	-2.65	1.07
	CBF in Temporal Lobe WM Left	0.921	0.78	1.18	0.24	-0.61	2.45
	FA in Left Entorhinal Area	0.242	0.21	1.13	0.26	-0.18	0.66
	CBF in Temporal Lobe WM Right	-0.307	0.79	-0.39	0.70	-1.85	1.24
	CBF in Left Entorhinal Area	0.429	0.28	1.55	0.12	-0.11	0.97
	Total Right Entorhinal Area Volume	-0.386	0.29	-1.33	0.18	-0.96	0.18
	SBP	-0.010	0.02	-0.56	0.58	-0.04	0.02
	Fasting Glucose	0.004	0.02	0.17	0.86	-0.04	0.05
	WC	0.026	0.02	1.26	0.21	-0.01	0.07
	10-Year Framingham Risk	0.022	0.10	0.23	0.82	-0.17	0.21
	ex9_age	-0.072	0.06	-1.13	0.26	-0.20	0.05
	Race	1.805	0.52	3.48	0.001	0.79	2.82
	SEX	2.389	0.69	3.48	0.001	1.04	3.74
	Education	0.302	0.08	3.61	<0.001	0.14	0.47
	HBP medication	-0.352	0.55	-0.64	0.52	-1.42	0.72
	Depression medication	-1.034	0.62	-1.67	0.10	-2.25	0.18
	Family income	0.160	0.12	1.34	0.18	-0.07	0.39
	constant	-5.838	5.07	-1.15	0.25	-15.77	4.10
Total Intracranial Volume ←	Fasting Glucose	0.014	0.01	1.97	0.05	0.00005	0.03
	WC	0.005	0.01	0.81	0.42	-0.01	0.02
	10-Year Framingham Risk	0.041	0.03	1.19	0.24	-0.03	0.11
	constant	-2.081	0.63	-3.29	0.001	-3.32	-0.84
Total Cerebrospinal Fluid Volume ←	Fasting Glucose	0.007	0.01	1.03	0.30	-0.01	0.02
	WC	-0.003	0.01	-0.41	0.68	-0.02	0.01
	10-Year Framingham Risk	0.048	0.02	1.92	0.06	-0.001	0.10
	constant	-0.740	0.72	-1.02	0.31	-2.16	0.68
Total White Matter Volume ←	Fasting Glucose	0.015	0.01	2.28	0.02	0.002	0.03
	10-Year Framingham Risk	0.040	0.03	1.29	0.20	-0.02	0.10
	constant	-1.775	0.60	-2.93	0.003	-2.96	-0.59
Total Gray Matter Volume ←	Fasting Glucose	0.012	0.01	1.70	0.09	-0.002	0.03
	WC	0.009	0.01	1.56	0.12	-0.002	0.02
	10-Year Framingham Risk	0.019	0.03	0.56	0.57	-0.05	0.08
	constant	-2.219	0.64	-3.46	0.001	-3.48	-0.96
Total Right Medial Frontal Cortex Volume ←	SBP	-0.009	0.01	-1.63	0.10	-0.02	0.00
	constant	1.069	0.66	1.63	0.10	-0.22	2.35

(Table Continued)

Total Right ACG Volume ←	SBP	-0.009	0.005	-1.89	0.06	-0.02	0.00
	constant	1.133	0.58	1.96	0.05	0.002	2.26
Abnormal Brain Tissue Volume ←	SBP	0.005	0.004	1.40	0.16	-0.002	0.01
	constant	-0.633	0.41	-1.53	0.13	-1.44	0.18
Abnormal White Matter Tissue Volume ←	SBP	0.005	0.003	1.44	0.15	0.00	0.01
	constant	-0.644	0.40	-1.60	0.11	-1.44	0.15
CBF in Brain ←	Fasting Glucose	-0.002	0.01	-0.21	0.83	-0.02	0.01
	WC	-0.010	0.01	-1.43	0.15	-0.02	0.004
	10-Year Framingham Risk	-0.033	0.03	-1.05	0.29	-0.09	0.03
	constant	1.294	0.61	2.12	0.03	0.10	2.49
CBF in White Matter ←	Fasting Glucose	-0.005	0.01	-0.81	0.42	-0.02	0.01
	WC	-0.013	0.01	-1.93	0.05	-0.03	0.0002
	constant	1.688	0.57	2.96	0.003	0.57	2.81
CBF in Gray Matter ←	Fasting Glucose	-0.004	0.01	-0.64	0.52	-0.02	0.01
	WC	-0.011	0.01	-1.59	0.11	-0.02	0.003
	constant	1.436	0.57	2.54	0.01	0.33	2.55
CBF in Left Hippocampus ←	WC	-0.012	0.01	-1.94	0.05	-0.02	0.00
	constant	1.080	0.56	1.94	0.05	-0.01	2.17
CBF in Right Medial Frontal Cortex ←	Fasting Glucose	-0.006	0.01	-1.08	0.28	-0.02	0.005
	WC	-0.011	0.01	-1.90	0.06	-0.02	0.0004
	constant	1.640	0.53	3.07	0.002	0.59	2.69
CBF in Right ACG ←	Fasting Glucose	0.005	0.01	0.79	0.43	-0.01	0.02
	WC	-0.011	0.01	-1.57	0.12	-0.02	0.003
	10-Year Framingham Risk	-0.043	0.02	-1.85	0.06	-0.09	0.003
	constant	0.750	0.59	1.28	0.20	-0.40	1.90
CBF in Left ACG ←	WC	-0.013	0.01	-2.40	0.02	-0.02	-0.002
	constant	1.202	0.49	2.47	0.01	0.25	2.16
FA in White Matter ←	SBP	-0.014	0.01	-2.71	0.01	-0.02	-0.004
	constant	1.913	0.63	3.04	0.002	0.68	3.15
FA in Left ACG ←	Fasting Glucose	-0.007	0.01	-1.15	0.25	-0.02	0.01
	constant	1.149	0.63	1.82	0.07	-0.09	2.38
FA in Right ACG ←	Fasting Glucose	-0.007	0.01	-1.12	0.26	-0.02	0.01
	constant	1.142	0.60	1.91	0.06	-0.03	2.31
ATV in Temporal Lobe WM Right ←	SBP	0.007	0.004	1.64	0.10	-0.001	0.02
	constant	-0.795	0.44	-1.79	0.07	-1.66	0.07
ATV in Temporal Lobe WM Left ←	SBP	0.003	0.003	0.79	0.43	0.00	0.01
	constant	-0.389	0.39	-1.00	0.32	-1.15	0.37
FA in Temporal Lobe WM Right ←	SBP	-0.017	0.01	-2.94	0.003	-0.03	-0.01
	constant	2.144	0.68	3.14	0.002	0.80	3.48
FA in Temporal Lobe WM Left ←	SBP	-0.014	0.01	-2.76	0.01	-0.02	-0.004
	constant	1.962	0.61	3.22	0.001	0.77	3.16
Total Temporal Lobe WM Left ←	Fasting Glucose	0.016	0.01	2.26	0.02	0.002	0.03
	10-Year Framingham Risk	0.046	0.03	1.36	0.18	-0.02	0.11
	constant	-1.818	0.63	-2.90	0.00	-3.05	-0.59
Total Temporal Lobe WM Right ←	Fasting Glucose	0.014	0.01	1.74	0.08	0.00	0.03
	WC	0.004	0.01	0.57	0.57	-0.01	0.02
	10-Year Framingham Risk	0.044	0.03	1.31	0.19	-0.02	0.11
	constant	-1.976	0.69	-2.88	0.004	-3.32	-0.63
CBF in Temporal Lobe WM Left ←	Fasting Glucose	-0.003	0.01	-0.47	0.64	-0.01	0.01
	WC	-0.014	0.01	-2.21	0.03	-0.03	-0.002
	constant	1.580	0.60	2.65	0.01	0.41	2.75
FA in Left Entorhinal Area ←	WC	-0.004	0.01	-0.68	0.50	-0.01	0.01
	10-Year Framingham Risk	-0.031	0.02	-1.66	0.10	-0.07	0.01
	constant	0.743	0.50	1.48	0.14	-0.24	1.73
CBF in Temporal Lobe WM Right ←	WC	-0.011	0.01	-1.92	0.06	-0.02	0.0002
	constant	1.075	0.53	2.01	0.04	0.03	2.12
CBF in Left Entorhinal Area ←	WC	-0.015	0.01	-2.51	0.01	-0.03	-0.003
	constant	1.344	0.52	2.57	0.01	0.32	2.37
Total Right Entorhinal Area Volume ←	10-Year Framingham Risk	0.067	0.02	2.83	0.01	0.02	0.11
	constant	-0.414	0.14	-2.95	0.003	-0.69	-0.14
var (e.Adj. Composite Cognition)		3.105	0.35			2.49	3.88
var (e.Total Right Medial Frontal Cortex Volume)		1.079	0.19			0.76	1.53
var (e.Total Right ACG Volume)		1.023	0.15			0.76	1.38
var (e.Abnormal Brain Tissue Volume)		0.385	0.07			0.27	0.55
var (e.Abnormal White Matter Tissue Volume)		0.363	0.07			0.25	0.53

(Table Continued)

var (e.FA in White Matter)	0.801	0.10	0.63	1.02
var (e.ATV in Temporal Lobe WM Right)	0.887	0.70	0.19	4.21
var (e.ATV in Temporal Lobe WM Left)	0.127	0.07	0.04	0.38
var (e.FA in Temporal Lobe WM Right)	0.899	0.12	0.70	1.16
var (e.FA in Temporal Lobe WM Left)	0.810	0.12	0.61	1.08
var (e.Total Intracranial Volume)	0.943	0.11	0.76	1.18
var (e.Total Cerebrospinal Fluid Volume)	0.785	0.10	0.61	1.01
var (e.Total White Matter Volume)	0.957	0.11	0.76	1.20
var (e.Total Gray Matter Volume)	0.964	0.10	0.79	1.18
var (e.CBF in Brain)	0.951	0.13	0.72	1.25
var (e.CBF in White Matter)	0.949	0.13	0.72	1.25
var (e.CBF in Gray Matter)	0.980	0.14	0.75	1.29
var (e.CBF in Right Medial Frontal Cortex)	0.902	0.13	0.68	1.20
var (e.CBF in Right ACG)	0.927	0.14	0.69	1.25
var (e.FA in Left ACG)	0.971	0.12	0.77	1.23
var (e.FA in Right ACG)	0.916	0.14	0.68	1.23
var (e.Total Temporal Lobe WM Left Volume)	0.988	0.12	0.78	1.26
var (e.Total Temporal Lobe WM Right Volume)	0.977	0.11	0.78	1.23
var (e.CBF in Temporal Lobe WM Left)	0.911	0.13	0.69	1.20
var (e.CBF in Left Hippocampus)	1.004	0.14	0.77	1.31
var (e.CBF in Left ACG)	0.936	0.13	0.72	1.22
var (e.FA in Left Entorhinal Area)	0.814	0.10	0.64	1.03
var (e.CBF in Temporal Lobe WM Right)	0.992	0.17	0.71	1.39
var (e.CBF in Left Entorhinal Area)	0.988	0.20	0.67	1.46
var (e.Total Right Entorhinal Area Volume)	0.964	0.11	0.77	1.20

* ACG – Anterior cingulate gyrus

* CBF – Cerebral blood flow

* FA – Fractional anisotropy

* ATV – Abnormal tissue volume

Appendix Table 11 Model selection based on the validation: misclassification rate for the metabolites

Model description	Validation data (30%)		Training data (70%)	
	Misclassification Rate	Average on Squared Error	Misclassification Rate	Average on Squared Error
Decision Tree 1	0.180	0.144	0.171	0.122
Decision Tree 2	0.180	0.145	0.171	0.124
Gradient Boosting	0.213	0.177	0.176	0.169
HP Forest 2 (Imputation)	0.216	0.156	0.000	0.026
HP Tree 2 (HP Forest 1)	0.227	0.175	0.240	0.182
HP Tree 3 (Imputation)	0.227	0.175	0.240	0.182
HP Tree 1	0.227	0.175	0.240	0.182
HP Tree 4 (Imputation + HP Forest 2)	0.227	0.175	0.240	0.182
HP Forest 1	0.256	0.160	0.000	0.026

* The arithmetic average (calculated as the sum of all values divided by the number of observations) was applied to only the missing cases of brain MRI parameters)

* HP Tree is created using the HPSPLIT procedure

Appendix Table 12 Model selection based on the validation: misclassification rate for the brain MRI parameters

Model description	Validation data (30%)		Training data (70%)	
	Misclassification Rate	Average on Squared Error	Misclassification Rate	Average on Squared Error
HP Forest 1	0.261	0.178	0.066	0.076
HP Tree 1	0.273	0.212	0.192	0.144
HP Tree 2 (HP Forest 1)	0.278	0.203	0.182	0.140
HP Tree 3 (Imputation)	0.290	0.206	0.301	0.209
HP Tree 4 (Imputation + HP Forest 2)	0.290	0.206	0.301	0.209
Decision Tree 1	0.292	0.195	0.298	0.193
Decision Tree 2	0.292	0.206	0.298	0.208
Gradient Boosting	0.292	0.211	0.276	0.207
HP Forest 2 (Imputation)	0.301	0.198	0.063	0.077

* The arithmetic average (calculated as the sum of all values divided by the number of observations) was applied to only the missing cases of brain MRI parameters