Characterising cognition in later life and its relationship with biomarkers of Alzheimer's disease: a study of members of the National Survey of Health and Development (the British 1946 birth cohort)

Kirsty Lu

Dementia Research Centre

Queen Square Institute of Neurology

University College London

Thesis submitted to University College London for the degree of Doctor of Philosophy

2019

I, Kirsty Lu, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

ABSTRACT

Alzheimer's disease (AD) has a long preclinical stage characterised by the accumulation of brain pathology, which is estimated to begin several decades before the onset of symptoms. A significant proportion of older adults harbour such pathology, although many of them may not develop dementia during their lifetimes. Growing evidence suggests that subtle cognitive decline occurs during this preclinical period, but many unanswered questions remain about the nature and timing of changes in different cognitive domains, and associations with life-course predictors.

This thesis is based on data from Insight 46, a neuroimaging sub-study of the MRC National Survey of Health and Development (the British 1946 birth cohort). In this population-based sample of 502 adults aged ~70 years, cognitive performance was assessed using standard and novel tests, and associations were investigated between cognition, life-course predictors, genetic risk factors for AD and brain pathologies, with a particular focus on β-amyloid.

The key finding was that participants with elevated levels of β-amyloid showed poorer performance across a range of cognitive domains – some of which have received little attention in previous studies – including non-verbal reasoning, intra-individual variability in reaction time, visuomotor integration and memory. Other important results include: independent effects of childhood cognitive ability, educational attainment and adult socioeconomic position on later-life cognition; an association between white matter pathology and slower processing speed; associations between larger whole brain volume and faster performance on several diverse timed measures; and evidence that *APOE*-ε4 carriers may be advantaged on tests of short-term memory after accounting for the detrimental effect of β-amyloid.

These results have implications for the interpretation of cognitive data measured in later life, and for the use of cognitive assessments to detect and track subtle cognitive decline in clinical trials that seek to delay or prevent the onset of AD dementia.

IMPACT STATEMENT

This thesis is based on data from Insight 46, a neuroimaging study of 502 members of the MRC National Survey of Health and Development (NSHD, the British 1946 birth cohort). All participants were born during the same week in March 1946 and have been studied ever since, with a rich dataset of measures of physical and mental health, cognition and lifestyle. As the NSHD is the world's longest continuously-running birth cohort study, it is a unique resource for the scientific community and its outputs have had a substantial influence on health policy, and will continue to do so. With participants now in their early 70s, this is an ideal time to investigate the emergence of neurodegenerative diseases, as participants are predominantly cognitively healthy but a significant proportion are expected to show evidence of accumulating brain pathologies. This study focuses on β-amyloid pathology, which is critical to the development of Alzheimer's disease (AD), the most common cause of dementia. As the prevalence of this devastating disease continues to rise rapidly, there is a pressing need for better understanding of its early preclinical stage – the stage during which future disease-modifying treatments are most likely to be effective. In particular there is a need for greater insight into the nature and timing of the earliest subtle changes in cognition, and how these changes can best be measured.

One of the main aims of this thesis was to investigate whether subtle differences in cognition could be detected between individuals with and without β-amyloid pathology (measured by β-amyloid positron emission tomography (PET) imaging). The results provide novel evidence that such differences are indeed detectable in cognitively-normal 70-year-olds across a variety of cognitive domains, including non-verbal reasoning, consistency of reaction time, visuomotor integration and memory. This has implications for our understanding of the earliest cognitive changes associated with AD, as the dominant narrative is that memory is the earliest domain to be affected, whereas these results suggest that other cognitive domains may also see very early changes. The results may also influence the design of future clinical trials, as sensitive cognitive tests are required to judge the effectiveness of potential disease-modifying treatments, and some of the novel computerised cognitive tests described here may be good candidates for outcome measures in such trials.

The results also showed that childhood cognitive ability, educational attainment and socioeconomic position each have independent effects on cognitive performance at age 70. This is consistent with previous reports that education and other cognitively-stimulating activities have an influence on cognitive trajectories across adulthood, and may have implications for public health efforts to reduce risk of later-life cognitive decline.

Dissemination of results to the research community is in progress and will continue through publication in scientific journals and presentation at international conferences, and public engagement activities will allow research outputs to be shared more widely. Data-sharing agreements are in place so that other researchers can request Insight 46 data to investigate their own research questions.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF FIGURES IN APPENDIX

LIST OF ABBREVIATIONS

A4 = Anti-Amyloid Treatment in Asymptomatic Alzheimer's Disease trial

Aβ = β-amyloid

Aβ+ = β-amyloid positive

Aβ- = β-amyloid negative

AD = Alzheimer's disease

ADNI = Alzheimer's Disease Neuroimaging Initiative

AIC = Akaike's Information Criterion

APOE = apoliprotein gene

AUC = Area under the curve

CI = confidence interval

CSF = cerebrospinal fluid

FAD = familial Alzheimer's disease

FLAIR = fluid attenuated inversion recovery MRI

FNAME = face-name associative memory exam

GEE = generalised estimating equations

IWG = International Working Group for New Research Criteria for the Diagnosis of AD

IQR = interquartile range

MCI = mild cognitive impairment

MMSE = Mini Mental State Examination

MRI = magnetic resonance imaging

MRC = Medical Research Council

NIA-AA = National Institute on Aging – Alzheimer's Association

NSHD = National Survey of Health and Development (the British 1946 Birth Cohort)

PACC = Preclinical Alzheimer Cognitive Composite

PET = positron emission tomography

ROC = Receiver Operating Characteristic

RT = reaction time

SD = standard deviation

SEP = socioeconomic position

SUVR = standard uptake volume ratio

UCL = University College London

WAIS-R = Wechsler Adult Intelligence Scale - Revised

WASI = Wechsler Abbreviated Scale of Intelligence

WMHV = white matter hyperintensity volume

WMS-R = Wechsler Memory Scale - Revised

1. INTRODUCTION

1.1. Alzheimer's Disease

There are 50 million people living with dementia globally (World Health Organization, 2019). In addition to the incalculable human impact, the economic cost is staggering – in the UK alone the estimated current cost is £26 billion per year, and this is projected to more than double by 2040 (Prince *et al.*, 2014). The most common cause of dementia is Alzheimer's disease (AD), accounting for about two-thirds of cases (*ARUK Dementia Statistics Hub*).

AD was first described in 1907 by Alois Alzheimer, a psychiatrist in Frankfurt, who reported memory loss, paranoia and disorientation in his patient Auguste D. After her death he produced a detailed characterisation of the neurofibrillary tangles he observed in her neurons, and described aggregations of a "peculiar substance" we now know to be β-amyloid, hereafter referred to as Aβ (Ryan, Rossor and Fox, 2015). These pathological hallmarks – extracellular β-amyloid plaques and intracellular neurofibrillary tangles containing hyperphosphorylated tau – are the basis for histopathological diagnosis of AD today, and are accompanied by atrophy (loss of neuronal cells) in selective brain regions which mainly correspond to the clinical presentation of symptoms. A typical presentation is characterised by progressive memory impairment and disorientation, corresponding to medial temporal lobe atrophy (Braak and Braak, 1991), but a variety of atypical non-amnestic presentations occur (Jones and Thompson, 2017). As the disease progresses and atrophy spreads, the cognitive and functional impairment becomes more global. While post-mortem confirmation of pathology is still the gold standard for a definitive diagnosis, new technologies have emerged which allow detection of Aβ, tau pathology and neurodegeneration *in vivo*. These innovations have made it possible to study the progression of pathological changes, and revealed that these changes begin around 20-30 years before the onset of symptoms, with Aβ pathology being the first to accumulate (Jack *et al.*, 2013; Palmqvist *et al.*, 2018). This period of pathological changes in the absence of symptoms is referred to as the preclinical stage of AD, and is discussed in greater detail later (section 2.2).

The causes of AD on an individual level are poorly understood, apart from for a very small minority of patients who carry autosomal dominant mutations for familial Alzheimer's disease (FAD), accounting for less than 1% of cases of AD (Bateman *et al.*, 2010). For sporadic AD, the biggest known risk factor is age: 95% of people living with dementia in the UK are over the age of 65, and the prevalence rises steadily in older

age, reaching about 40% among those aged 95 and above (Prince *et al.*, 2014). Other risk factors include the *APOE* gene, which is discussed later (section 2.6.1) and lifestyle factors such as smoking, hypertension and obesity (Norton *et al.*, 2014).

There are currently no disease-modifying treatments available for AD and the field has suffered from a series of disappointing failures from clinical trials of drugs that have targeted Aβ pathology.

1.2. Preclinical Alzheimer's Disease

As AD has such a long preclinical window, this may be the most beneficial time to provide disease-modifying therapies. While some secondary prevention trials (i.e. targeting asymptomatic individuals with evidence of Aβ deposition) are already underway, many important questions remain. We still lack understanding of the timing and sequence of the pathological changes that occur, influences on their progression, and factors affecting how and when this leads to cognitive decline – especially on an individual level. Over the last decade, much effort has been directed towards creating standardised criteria for defining preclinical AD. These criteria are designed to serve as a framework for research, with the ultimate aim of providing an evidence base for designing appropriate and successful clinical trials in this population. These criteria are discussed in more detail in section 2.2.

Although individuals with preclinical AD are, by definition, cognitively normal, there is increasing evidence that subtle changes in cognition can be detected during this period (see section 2.3). A key research question is how these changes relate to the preclinical disease process and how they can best be measured. Sensitive cognitive tests are crucial for identifying individuals at risk of AD and for use as outcome measures in clinical trials. Indeed, the US Food and Drug Administration have recently approved cognition as a sole end-point for such trials, where previously they required evidence of functional improvement as well – a requirement clearly unworkable for a population with no functional impairment (Kozauer and Katz, 2013). In addition, using sensitive cognitive tests in observational studies could increase our understanding of the factors which affect an individual's risk of cognitive decline – such as demographic factors, physical and mental health – which could inform interventions to reduce risk.

1.3. The MRC National Survey of Health and Development

My PhD project sits within the Insight 46 study, a neuroimaging sub-study of the MRC National Survey of Health and Development (NSHD, also known as the British 1946 Birth Cohort). The NSHD was established in 1946 to address concerns about infant and maternal health, specifically why the national fertility rate was falling and whether this could be attributed to the quality and cost of obstetric and midwifery services. Ninety-one percent of all mothers who gave birth in England, Scotland and Wales during one week in 1946 (n = 13687) were interviewed at their 8-week check-up and the findings were influential in designing maternity services within the new NHS. 5362 babies from this sample were enrolled to form the original NSHD cohort, stratified by socioeconomic position in order to investigate the effects of social inequality (Wadsworth *et al.*, 2006).

Between 1946 and 2015 there have been 24 data collection waves, focussing initially on childhood development and educational attainment, then shifting to physical and mental health, lifestyle and cognition across adulthood. Further details on the cognitive measures – of particular relevance to this thesis – are provided in section 3.2.1. The results of the NSHD have had a substantial impact on policy, including influencing the debate about comprehensive education and social mobility, providing pivotal evidence that smoking during pregnancy is harmful, and drawing attention to the unfolding obesity crisis (Pearson, 2016). High rates of participation have been maintained through the use of home visits for data collection and by keeping in regular contact with study members (Stafford *et al.*, 2013; Kuh *et al.*, 2016). At the most recent data collection wave in 2014-2015, the target sample was 2816 study members, i.e. 52% of the original cohort. The remainder were no longer active for the following reasons: died (18%), permanently withdrawn (12%), living abroad (11%), lost to follow-up (7%) (Kuh *et al.*, 2016).

The NSHD is the world's longest continuously-running birth cohort and aims to become a complete cradle-to-grave study. With participants now entering their eighth decade, the cohort provides a unique opportunity for understanding neurodegeneration in the context of ageing, and the complex factors and interactions that influence its progression.

1.4. Insight 46

Insight 46 is a neuroscience sub-study of the NSHD, run jointly by the Dementia Research Centre and the MRC Lifelong Health and Ageing Unit (both at University College London). Combining intensive cognitive and clinical assessment with collection of neuroimaging, blood and genetic biomarkers, Insight 46 aims to identify investigate life-course and genetic influences on brain health and cognition, with a particular focus

on AD. This will provide a critical evidence base for future therapeutic trials in the preclinical stage of AD. As the study members were aged ~70 at the time of recruitment into Insight 46, the prevalence of dementia was expected to be low – around 3% (Prince *et al.*, 2014) – but a sizeable minority of participants were expected to be in the preclinical stages of AD, with meta-analytical data suggesting significant Aβ pathology in around 15-25% of individuals at this age (Jansen *et al.*, 2015).

Participants recruited to Insight 46 were invited to University College London (UCL) for assessments at two time-points with an interval of approximately two years. Data collection included cognitive tests, clinical history and examination, Aβ positron emission tomography (PET), brain magnetic resonance imaging (MRI), and other biomarker and genetic measures. Further details on recruitment and data collection are provided in sections 3.1and 3.2 and in the protocol paper (Lane *et al.*, 2017). Baseline assessments were conducted between May 2015 and January 2018. Follow-up assessments began in January 2018 and will be completed around the summer of 2020.

1.5. Scope of PhD

My research focuses on analysing the cognitive data collected during the baseline Insight 46 assessments, with the ultimate aim of understanding more about subtle cognitive changes that may be associated with preclinical AD pathology, in particular Aβ plaques. The following chapter introduces the background to the unanswered questions in this field, reviews the relevant literature, and leads to a statement of my specific research questions and hypotheses.

2. BACKGROUND AND RESEARCH QUESTIONS

2.1. Rationale

Research into the relationship between preclinical AD pathology and changes in cognition is based on the understanding that individuals do not suddenly become cognitively impaired, but rather there is a gradual process of cognitive decline, during which time cognitive performance initially remains within the normal range. Jessen *et al.* produced a figure to illustrate this, which I have used as a basis for structuring my review of the literature (Figure 2-1): I have annotated the figure with the letters A-D to highlight four important questions, which are explained below.

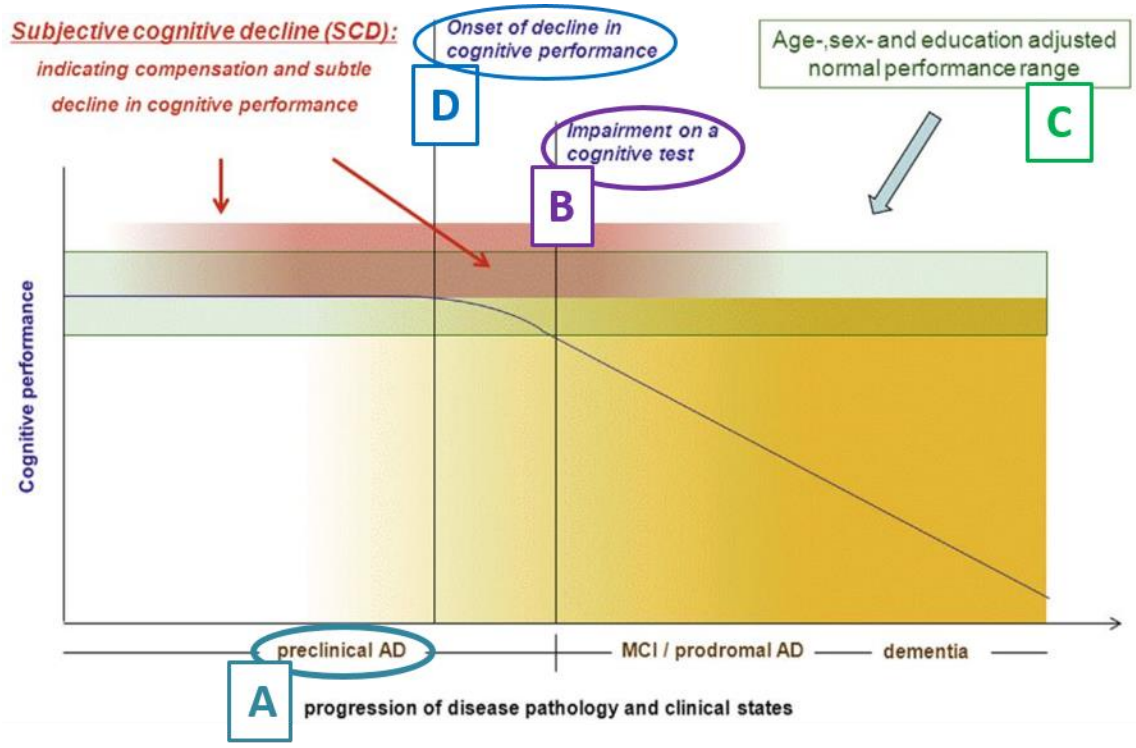


Figure 2-1. The course of cognitive decline through preclinical AD to dementia

The questions for my literature review are indicated by the labels A to D – see text for a full explanation. This figure is reprinted from Jessen et al. (2014) with permission from Elsevier. (The letters A to D are my own addition.) Jessen et al. designed this figure to draw attention to the concept of Subjective Cognitive Decline (SCD), which describes the phenomenon of some older adults reporting concerns about declining memory and cognition, despite showing no evidence of any objective cognitive impairment. My research focuses on objective measures of cognition, so SCD is not discussed further.

A – What is preclinical AD? To investigate whether the cognitive tests used in Insight 46 can detect subtle cognitive decline in the preclinical stage of AD, preclinical AD must be defined. The main scope of this question is to review the development of criteria for defining preclinical AD in clinical and pathological terms, as well as to review evidence relevant to the application of these criteria to the Insight 46 cohort, such as the prevalence of preclinical AD at age ~70 years. However, as becomes clear from the debates and controversies surrounding these criteria, this question also has a more philosophical dimension relating to classification of disease in apparently asymptomatic individuals. This question is reviewed in section 2.2.

B – What is the evidence that subtle cognitive changes are detectable in preclinical AD? Cognitive impairment can be defined as the point when a participant's cognition is below the normal range for their age, sex and education. This is represented in Figure 2-1 by the vertical line labelled "Impairment on a cognitive test" which intersects the cognitive trajectory (smooth purple line) at the point where it drops out of the green shaded area. But there is evidence that cognition begins to decline among so-called cognitively-normal individuals who are in the preclinical stage of AD, before reaching the point of impairment, as indicated by the vertical line labelled "Onset of decline in cognitive performance". In aiming to detect and measure subtle cognitive changes in preclinical AD, the Insight 46 cognitive battery is targeting the region between these two vertical lines. Evidence from previous studies with similar aims is reviewed in section 2.3.

C – Which predictors of individual differences in cognitive performance are most important to account for? Some of the variation between individuals in terms of cognitive performance can be explained by factors such as age and education. These factors need to be accounted for in order to define the normal range of cognitive performance for a certain population (represented in Figure 2-1 by the green band). Accounting for this predictable variation between individuals should increase the sensitivity of cognitive tests to detect subtle cognitive changes associated with preclinical AD within this normal range. This question is reviewed in section 2.4.

D – Which cognitive measures are most sensitive to subtle cognitive decline in preclinical AD? In Figure 2-1, the vertical line labelled 'Onset of decline in cognitive performance' illustrates the point at which subtle cognitive decline becomes detectable, even though performance is still within the normal range. With more sensitive cognitive measures, this vertical line could move to the left as subtle differences could be detected earlier. Some approaches to enhancing the sensitivity of cognitive measures are reviewed in section 2.5.

These four questions are now reviewed in the following four sub-sections. Each sub-section ends with a summary of the implications for my research.

2.2. What is preclinical AD?

2.2.1. Development of research criteria for preclinical AD

As mentioned in section 1.1, the advent of technologies to measure *in vivo* biomarkers of AD pathology has led to a reconceptualization of AD as a continuum with a long preclinical stage. Efforts to create standardised biomarker-based criteria for preclinical AD have been led by two main groups: the International Working Group for New Research Criteria for the Diagnosis of AD (IWG) and the US National Institute on Aging – Alzheimer's Association (NIA-AA). The evolution of these criteria over the last ten years reflects the progress that has been made in understanding the biology of AD, the new biomarker measures that have become available, and the ongoing debates about how the disease should be conceptualised. A chronological summary of the various iterations of the IWG and NIA-AA criteria for preclinical AD is provided in Table 2-1. It is important to note that both sets of criteria have been conceived as frameworks for research and are not recommended for use in clinical practice.

All the biomarkers currently included in criteria for preclinical AD are derived from neuroimaging or cerebrospinal fluid sampling. Blood-based biomarkers for AD are in development but require further work (Zetterberg, 2019) and are not discussed further. Technical details about the biomarkers referred to in Table 2-1, and their relative advantages and disadvantages, are not discussed here, but further details are provided in section 3.2.2 for the biomarker measures that are used in the analyses presented in this thesis.

Table 2-1. Development of IWG and NIA-AA criteria for preclinical AD

		AD is…	AD begins…	Definitions of preclinical AD	Biomarkers of AD pathology
IWG	IWG-1 criteria (Dubois *et al.*, 2007)	...a dual clinicopathological entity. Evidence of progressive episodic memory impairment plus at least one abnormal biomarker.	*(not explicitly stated)*	No operational definition, but a descriptive definition: Preclinical AD = "The long asymptomatic period between the first brain lesions and the first appearance of symptoms and which concerns normal individuals that later fulfil AD diagnostic criteria."	• abnormal Aβ, t-tau or p-tau in CSF • medial temporal lobe atrophy on MRI • hypometabolism on FDG-PET
	Update and clarification of the lexicon (Dubois *et al.*, 2010)	...a dual clinicobiological entity, with symptoms classified as either typical (episodic memory) or atypical (e.g. language or visual). Biomarkers support the diagnosis.	...with the first symptoms.	• Asymptomatic at risk for AD = cognitively-normal individuals with biomarker evidence of AD pathology. • Presymptomatic AD = cognitively-normal individuals who carry a proven autosomal dominant mutation for AD	• abnormal Aβ-PET • abnormal Aβ, t-tau or p-tau in CSF • medial temporal lobe atrophy on MRI • hypometabolism on FDG-PET
	IWG-2 criteria (Dubois *et al.*, 2014)	*(as above)*	*(as above)*	*(as above)*	• abnormal Aβ-PET • abnormal Aβ in CSF together with abnormal t-tau or p-tau in CSF

IWG (continued)	Updated criteria for preclinical AD (Dubois *et al.*, 2016)	…a pathological entity defined by amyloid and tau pathology.	...when there is evidence of both tau and amyloid pathology.	Preclinical AD = cognitively-normal individuals with biomarker evidence of abnormal amyloid and tau The following two classifications are no longer considered as preclinical AD but may precede it: • Asymptomatic at risk for AD = cognitively-normal individuals with biomarker evidence of abnormal amyloid or tau (but not both) • Presymptomatic AD = cognitively-normal individuals who carry a proven autosomal dominant mutation for AD	• abnormal Aβ-PET or abnormal Aβ in CSF • abnormal tau-PET or abnormal tau in CSF
NIA-AA	criteria (Sperling *et al.*, 2011)	…a pathological entity defined by amyloid and tau pathology, based on a model where amyloid becomes abnormal first, then tau, then cognition.	...when there is evidence of amyloid pathology.	Preclinical AD has three stages: • Stage 1 = cognitively-normal individuals with biomarker evidence of abnormal amyloid (but normal tau) • Stage 2 = cognitively-normal individuals with biomarker evidence of abnormal amyloid and neuronal injury • Stage 3 = individuals with biomarker evidence of abnormal amyloid and neuronal injury, plus evidence of subtle cognitive decline†	• abnormal Aβ-PET or abnormal Aβ in CSF Neuronal injury defined as any of the following: • abnormal t-tau or p-tau in CSF • hypometabolism on FDG-PET • a particular distribution of cortical thinning • hippocampal atrophy on MRI
	Proposed update to criteria (Jack *et al.*, 2012)	*(as above)*	*(as above)*	As above for Stages 1-3, but with two additional categories: • Stage 0 = cognitively-normal individuals with no abnormal biomarkers • Suspected Non-Alzheimer Pathology (SNAP) = cognitively-normal individuals with normal amyloid but abnormal markers of neuronal injury	*(as above)*

table continued on next page

		AD is…	AD begins…	Definitions of preclinical AD	Biomarkers of AD pathology
NIA-AA (continued)	A/T/N biomarker classification system (Jack *et al.*, 2016)	*This paper does not seek to define AD or any other disease, but makes an important contribution to the development of the biomarker framework used by NIA-AA. Amyloid (A), tau (T) and neurodegeneration (N) are all dichotomised into normal (-) or abnormal (+), giving rise to 8 possible biomarker profiles: A+T+N+, A+T+N-, A+T-N+, A+T-N-, A-T+N+, A-T-N+, A-T+N-, A-T-N-.*			• A is defined by either abnormal Aβ-PET or abnormal Aβ in CSF • T is defined by either abnormal tau-PET or abnormal p-tau in CSF • N is defined by either hypometabolism on FDG-PET, abnormal t-tau in CSF, or atrophy in regions characteristic of AD on MRI
	Updated criteria (Jack *et al.*, 2018)	…a purely pathophysiologic entity, with no reference to clinical symptoms	...when there is evidence of both tau and amyloid pathology.	Using the ATN framework above, individuals are placed into five categories based on their biomarker profiles. Three of these categories form the Alzheimer's continuum: • Alzheimer's pathologic change (A+T-N-) • AD (A+T+N-, A+T+N+) • Alzheimer's and concomitant suspected non-Alzheimer's pathologic change (A+T-N+) The other two categories are not part of the continuum: • Normal AD biomarkers (A-T-N-) • Non-Alzheimer's pathologic change (A-T+N-, A-T-N+, A-T+N+) Any of these biomarker profiles may be combined with one of three cognitive stages: cognitively unimpaired, MCI or dementia. Preclinical AD = cognitively-unimpaired individuals with an "AD" biomarker profile Preclinical Alzheimer's pathologic change = cognitively-unimpaired individuals with an "Alzheimer's pathologic change" biomarker profile.	*(as above)*

Some ideas for the structure of this table came from Table 1 in (Dubois et al., 2016). † No operationalised definition of subtle cognitive decline was proposed but it was noted that cognition may still be within the normal range, individuals may report subjective cognitive decline, or there may be subtle neurobehavioural changes. Other studies have attempted to define cut-offs for subtle cognitive impairment (Jack et al., 2012; Vos et al., 2013; Edmonds et al., 2015).

Aβ = β-amyloid; AD = Alzheimer's disease; CSF = cerebrospinal fluid; FDG = fluorodeoxyglucose; IWG = International Working Group for New Research Criteria for the Diagnosis of AD; MCI = mild cognitive impairment; MRI = magnetic resonance imaging; NIA-AA = US National Institute on Aging – Alzheimer's Association; PET = positron emission tomography; p-tau = phosphorylated tau; t-tau = total tau.

The impact of biomarkers on conceptualisation of AD is immediately obvious from Table 2-1, as it can be seen that the IWG's 2007 diagnostic framework was the first to require the presence of one or more abnormal biomarkers in addition to evidence of an episodic memory impairment for a diagnosis of AD (Dubois *et al.*, 2007), and by 2018 the influential NIA-AA criteria have now removed clinical symptoms from the framework completely.

It can be seen from Table 2-1 that both the IWG and NIA-AA have arrived at a very similar definition of AD, stating that it begins when there is evidence of both tau and Aβ pathology. One key area of difference has been around the classification of cognitively-normal individuals who show only evidence of Aβ pathology (but normal tau). The influential NIA-AA 2011 criteria designated these individuals as being in Stage 1 of preclinical AD, whereas all iterations of the IWG criteria have opted for the more cautious designation of "Asymptomatic at risk for AD". However, the gap between these two positions has been narrowed by the updated NIA-AA 2018 criteria, as such individuals are now referred to as showing "preclinical Alzheimer's pathologic change" and the label of "preclinical AD" is reserved for those with evidence of both Aβ and tau pathology (as in the IWG criteria). Another important difference is the classification of cognitively-normal individuals who show only evidence of tau pathology (but normal Aβ). IWG includes these individuals in their "Asymptomatic at risk for AD" category whereas NIA-AA considers such individuals to show "Non-Alzheimer's pathologic change".

It is no coincidence that these differences relate to the earliest stages of pathologic change in cognitively-normal individuals, as the earliest stages are obviously particularly challenging to characterise, and are the focus of the most contentious ongoing debates within AD research. Major issues include 1) to what extent the amyloid hypothesis should remain the dominant model of AD and how it may be integrated with other models; 2) the significance of terminology. These issues are discussed in the following sub-sections, with a focus on those which are most relevant to the investigation of associations between biomarkers of Aβ pathology and cognition in Insight 46. As the NIA-AA 2018 criteria are the most recent and most influential, the current debates and discussions focus more on them.

2.2.1.1. Dominance of the amyloid hypothesis

The amyloid hypothesis contends that the accumulation of Aβ plaques between neurons – arising from an imbalance between the production and clearance of β-amyloid – is the primary cause of AD. This hypothesis has dominated the field for the last 25 years (Selkoe and Hardy, 2016). Pivotal evidence for the amyloid hypothesis came from the

discovery of the genetic mutations causing FAD, all of which are in the amyloid precursor protein (*APP*) or the presenilin genes (*PSEN1* and *PSEN2*) involved in generating Aβ. These mutations result in over-production of Aβ, and a similar phenomenon is seen in Down's Syndrome due to the duplication of chromosome 21 which contains the *APP* gene. Further compelling evidence for the amyloid hypothesis will not be described in detail here, but includes 1) the observation that accumulation of Aβ pathology begins several years before the appearance of tau pathology and neurodegeneration (Bateman *et al.*, 2012; Jack *et al.*, 2013; Pletnikova *et al.*, 2018); 2) the well-documented neuronal toxicity of Aβ in animal studies (Selkoe and Hardy, 2016); 3) evidence that the *APOE*-ε4 allele – the biggest risk factor for AD after age (see section 2.6.1) – impairs clearance of Aβ (Selkoe and Hardy, 2016).

Criticism of the amyloid hypothesis is rooted in the universal failure of drug trials targeting Aβ in patients with AD or mild cognitive impairment (MCI). Some argue that the amyloid hypothesis is too linear and simplistic (e.g. Edmonds *et al.*, 2015), and the time has come – or indeed is long overdue – to embrace other models of AD. Proposed alternatives will not be discussed here, but some examples are listed to illustrate the complexity and diversity of models of AD, and to highlight the contrast with the apparent simplicity of the IWG and NIA-AA criteria, which are firmly based on the amyloid hypothesis: 1) systems-based models that focus on the wider consequences of pathology for brain systems such as cholinergic deficits (Tang, Lutz and Xing, 2018) and the effects of these "systems failures" on diverse clinical symptoms such as cognition, sleep and depression (Medina *et al.*, 2017); 2) cellular models which focus on the complex cellular alterations which underline the long preclinical phase (De Strooper and Karran, 2016); 3) holistic approaches that focus on the biological mechanisms of diverse risk factors for AD such as ageing, genetics and lifestyle (Morris, Clark and Vissel, 2018); 4) the vascular hypothesis which focuses on vascular dysfunction including changes to the integrity of the blood-brain barrier and cerebral blood flow (de la Torre, 2018; Sweeney *et al.*, 2019); 5) the calcium hypothesis which focuses on the consequences of alterations to neuronal calcium signalling (Alzheimer's Association Calcium Hypothesis Workgroup, 2017); 6) the neuroinflammation hypothesis which focuses on dysfunction of microglial cells (Heneka *et al.*, 2015).

This decision to focus on AD pathology in isolation from other processes and systems causes some to question the validity of the criteria in real life, and also raises concerns that the dominance of these criteria may divert research effort away from other potentially fruitful areas (McCleery *et al.*, 2018, 2019; Louie, 2019).

2.2.1.2. Significance of terminology

The criteria for preclinical AD raise an immediate issue about terminology. As noted earlier, the term "preclinical AD" does not have the same meaning in the IWG and NIA-AA criteria, and it can be seen in Table 2-1 that both groups have evolved their definition of this term over the last few years. Consequently, the literature on preclinical AD shows a lack of consistency, with the terms "preclinical", "prodromal", "presymptomatic" and "at-risk asymptomatic" being used with overlapping but non-equivalent meanings by different authors. "Prodromal" is usually applied to cover the period immediately preceding the onset of dementia, when patients might meet criteria for MCI (e.g. Visser *et al.*, 2012), whereas "preclinical" generally covers the cognitively-normal stage before this. Although both sets of criteria now require evidence of both Aβ and tau pathology for a classification of preclinical AD, the fact that that NIA-AA 2011 criteria required only evidence of Aβ pathology has resulted in numerous studies describing a "preclinical AD" group based purely on elevated levels of amyloid (e.g. Dang *et al.*, 2018; Harrington *et al.*, 2018; Slot *et al.*, 2018). The term has also been used by some authors to describe *APOE*-ε4 homozygotes (Caselli *et al.*, 2014) or individuals with MCI (Kirova *et al.*, 2015).

The issue of terminology is important because of concerns about the consequences of labelling people with "preclinical Alzheimer's Disease". While the criteria are explicitly for research purposes only, it has been argued that they will inevitably filter through into clinical practice, which could result in a significant proportion of older adults being diagnosed with preclinical AD, even though many of them may never go on to develop clinical symptoms (Boenink, 2018; McCleery *et al.*, 2018, 2019; Morris, Clark and Vissel, 2018). This would raise a number of ethical, societal and economic issues, especially given the current dearth of evidence-based treatment or management options available to individuals meeting the criteria. Some of the authors behind the NIA-AA criteria have pointed out that there is not the same concern over the terms "precancerous lesion" or "pre-diabetes", arguing that the unease over the term "preclinical AD" reflects the continued fear and stigma surrounding dementia, as well as the fact that progression from preclinical AD to dementia cannot currently be predicted with any accuracy on an individual level (Sperling, Mormino and Johnson, 2014). The prospect of causing significant anxiety to otherwise healthy people is concerning, and its implications are hard to predict (Stites, Milne and Karlawish, 2018).

Another contentious issue of terminology is raised by the definition of "Alzheimer's disease" in the NIA-AA criteria. The authors make clear that their usage of this term is purely biological and implies nothing about clinical symptoms. Although they provide alternative terminology to describe the clinical presentation of AD – either "AD with dementia" for individuals meeting the biomarker criteria or "Alzheimer's clinical

syndrome" for individuals with unknown biomarker profiles – critics argue that the common usage of the term AD (as a descriptor for the clinical symptoms) is likely to persist, so it would be more appropriate to create a new term for the biological entity or simply use the A/T/N categories (Louie, 2019; McCleery *et al.*, 2019). In addition, it is important to remember that most clinics around the world do not have the resources to investigate biomarkers. Also, the value of a broad commonly-understood term in public perception should not be underestimated as it underpins efforts to raise awareness and support patients and carers.

While attending a feedback session on the proposed NIA-AA 2018 criteria at the Alzheimer's Association International Conference 2017, I observed that the debates around terminology were marked by a strength of feeling that stems from the fundamental nature of the issues raised. It may be informative to note that these issues are rooted in a wider question that has received much attention from philosophers of science, namely "What is disease?". The extensive literature on this question will not be reviewed here, but some points of interest are listed below to indicate why the controversies surrounding criteria for preclinical AD have no easy resolution.

i) The identification of AD pathology in a substantial proportion of older adults raises the issue of the blurred boundary between ageing and disease. If such pathology is common, then to what extent can it be considered distinct from so-called "normal" ageing (Lock, 2013)?

ii) Related to this, gerontologists debate the question of whether there is such a thing as "healthy ageing" or whether senescence (biological ageing) itself should be considered as a disease process (Bulterijs *et al.*, 2015; Gladyshev and Gladyshev, 2016; Janac, Clarke and Gems, 2017).

iii) This debate follows from the fact that it has proved impossible to agree on a definition of disease. No theory can draw a meaningful distinction between diseases and other states of sub-optimal health (e.g. injury, disability, malnutrition, frailty) (Murphy, 2015) or account satisfactorily for presymptomatic disease (Broadbent, 2014). Competing concepts of disease broadly fall into naturalist and constructivist camps, which disagree in essence over whether diseases exist in nature or are constructed based on human normative judgements (Murphy, 2015).

iv) As illustrated by the concerns around of the NIA-AA criteria, these abstract concepts have potentially far-reaching implications for the way individuals view their health and identity, and the priorities of medical research, treatment and care (Nordenfelt, 2007).

2.2.2. Expected prevalence of preclinical AD in Insight 46

A significant proportion of cognitively-normal older adults meet criteria for preclinical AD, with one meta-analysis putting the figure at 22% (Parnetti *et al.*, 2019). The increasing accumulation of brain pathology throughout older age is well-documented. For example, one study estimated that the prevalence of Aβ pathology increases from 10% at age 50 to 44% at age 90 (Jansen *et al.*, 2015), and another study based on the ATN biomarker framework (see Table 2-1) found that more than 90% of people have at least one abnormal biomarker by the age of 85 (Jack *et al.*, 2017). Previous studies provide a basis for estimating the expected prevalence of preclinical AD in Insight 46 participants (aged ~70 years at the time of recruitment) and the estimated risk of progression to cognitive impairment.

Based on a meta-analysis of nearly 3000 cognitively-normal older adults who underwent either Aβ-PET or cerebrospinal fluid (CSF) sampling, the prevalence of Aβ pathology at age 70 is around 15-25% (Jansen *et al.*, 2015). Consistent with this, in a more recent population sample of 322 70-year-olds from Sweden who underwent CSF sampling, 23% had evidence of Aβ pathology (Kern *et al.*, 2018). The prevalence of preclinical AD in the Swedish sample was 10%, according to IWG 2016 and NIA-AA 2018 criteria, meaning that less than half of individuals with Aβ pathology also showed evidence of tau pathology.

According to a meta-analysis of studies investigating risk of progression to MCI or dementia in older adults, over intervals ranging from 1.3 to 10.4 years, the estimated risk of progression for those with Aβ pathology alone was 20%, whereas the risk for those with both Aβ and tau pathology was 38% (Parnetti *et al.*, 2019). A study of participants (n=599) in the Australian Imaging Biomarkers and Lifestyle (AIBL) study found a similar result, with an 18% risk of progression for Aβ+ participants over 8 years (measures of tau were not available) (Dang *et al.*, 2018). These studies suggest that the risk of progression is low to moderate over a timescale of up to 10 years, and therefore it may be expected that many older adults with Aβ pathology will not develop symptoms in their lifetime. A recent study of cognitively-normal older adults stratified by sex and biomarkers of preclinical AD produced the following estimates of lifetime risk of AD dementia for 70-year-olds: females with no AD pathology = 17%; males with no AD pathology = 11%; females with Aβ pathology only = 27%; males with Aβ pathology only = 20%; females with Aβ pathology and neurodegeneration = 39%; males with Aβ pathology and neurodegeneration = 31% (Brookmeyer and Abdalla, 2018).

2.2.3. Implications for my research

The Insight 46 protocol includes Aβ-PET but does not include a measure of tau pathology, so current criteria for preclinical AD cannot be fully applied. When data collection and analysis began, it may have seemed reasonable to follow other studies in defining a "preclinical AD group" based on Aβ pathology alone, as per the NIA-AA 2011 criteria. However, with the publication of the updated 2018 criteria it has becomes clearer that this term is not justified according to current consensus. Therefore Insight 46 participants whose Aβ pathology is above normal levels will simply be referred to as Aβ+. The term suggested for Aβ+ in the NIA-AA 2018 criteria – "preclinical Alzheimer's pathologic change" – would not add additional clarity in this context, because it implies that tau pathology is absent, whereas in fact it is unknown. The same argument applies to the "asymptomatic at-risk" label favoured by the IWG group, which has the additional disadvantage of being a potential misnomer in the context of my aim to detect subtle cognitive changes. The concerns discussed in section 2.2.1.2 about the significance of terminology provide a further reason to stick to the simple descriptor of Aβ+, rather than applying disease terminology to healthy individuals.

However, preclinical AD remains the consensus term to describe the research area that my project focuses on, and, according to current understanding, Aβ pathology plays a critical – if not primary – role in the development of AD dementia. Therefore, if Aβ+ participants show cognitive differences from Aβ- participants in Insight 46, it will be reasonable to draw qualified conclusions about early cognitive changes in the preclinical AD continuum.

2.3. What is the evidence that subtle cognitive decline is detectable in preclinical AD?

As discussed above, the current criteria for preclinical AD are purely biological, but there is intense interest in understanding their relationship to cognitive decline. Being able to detect and track the earliest changes in cognition is of prime importance to clinical trials that seek to reduce the risk of conversion to MCI or dementia in individuals who are accumulating AD pathology.

Studies that have compared cognitive performance in older adults with and without preclinical AD pathology are reviewed below. This review only considers objective cognitive measures; there are reports that preclinical AD pathology may also be associated with subjective cognitive decline (Jessen et al., 2014) and subtle changes in

behaviour (Caselli *et al.*, 2018), but these are not discussed. Another important evidence-base relevant to subtle cognitive decline in preclinical AD comes from studies of presymptomatic FAD mutation carriers (e.g. Wang *et al.*, 2015), which are not reviewed here, but such studies are considered in subsequent chapters where relevant to specific tasks in the Insight 46 cognitive battery.

This section provides a general overview of the relationship between preclinical AD pathology and cognition, but does not discuss specific cognitive domains and tests. For the cognitive domains tested in Insight 46, more detailed discussions of the current evidence for pathology-related changes are provided in the introductions to the relevant chapters (4 to 8). As the Insight 46 cognitive battery was already in place before my involvement with the study, this review was not undertaken to inform the selection of cognitive tests, but rather to establish the consistency and magnitude of Aβ-related effects on cognition reported in other studies.

The evidence for differences in cognition between cognitively-normal older adults with and without preclinical AD pathology was comprehensively evaluated in 2017 in a systematic review (Mortamais *et al.*, 2017) and two meta-analyses (Baker *et al.*, 2017; Duke Han *et al.*, 2017). The findings of these three publications are summarised first, followed by a discussion of studies published between 2017 and April 2019.

2.3.1. Summary of systematic reviews and meta-analyses of cognitive changes in preclinical AD

Mortamais *et al.* (2017) summarised evidence of associations between cognition and the following three biomarker measures: structural brain changes, functional brain changes and amyloid burden. Most of the included studies did not claim to define a "preclinical AD" group, but investigated populations who were "at risk" of clinical AD due to the presence of the *APOE*-ε4 allele, amyloid deposition, or other markers of neurodegeneration such as hippocampal atrophy. The authors concluded that cross-sectional studies generally have not observed associations between cognition and biomarkers of preclinical AD, although there were some positive results for episodic memory tests. However, they found more consistent evidence for associations between cognitive decline and amyloid burden in longitudinal studies with follow-ups of at least 2 years, particularly for tests of episodic memory and for global cognitive composites.

Baker *et al.* (2017) conducted a more focused meta-analysis of the effect of Aβ pathology on cognition in cognitively-normal older adults, as an update to a previous meta-analysis (Hedden *et al.*, 2013). They concluded that in cross-sectional studies (n = 30) there was

evidence that Aβ+ groups showed small impairments (standardised effect sizes in the range *d* = 0.15 to 0.32) in global cognition, visuospatial function, processing speed, episodic memory and executive function. In longitudinal studies (n = 14) there was evidence that Aβ+ groups showed small to moderate cognitive decline (standardised effect sizes in the range *d* = 0.24 to 0.30) in episodic memory, visuospatial function, semantic memory and global cognition. Effect sizes were moderated by type of Aβ measure (PET or CSF), type of analysis, inclusion of covariates, and exclusion criteria used.

Duke Han *et al.* (2017) conducted a similar meta-analysis of 61 cross-sectional studies comparing cognitive performance of Aβ+ and Aβ- cognitively-normal older adults (classified using PET or CSF), and concluded that Aβ+ individuals showed evidence of small impairments in global cognitive function, memory, language, visuospatial ability, processing speed, attention, working memory and executive functions (standardised effect sizes in the range *d* = 0.04 to 0.20). They conducted a second meta-analysis of a subset of studies comparing Aβ+ individuals with and without tau pathology or neurodegeneration (corresponding to Stages 1 and 2 of the NIA-AA 2011 criteria – see Table 2-1) and concluded that Aβ+ individuals with tau pathology or neurodegeneration were more impaired on memory measures than those without the additional pathology (*d* = 0.46).

2.3.2. Studies published between 2017 and April 2019

Between 2017 and April 2019, I appraised new publications in the field on a weekly basis. Some key themes from studies published during this period are highlighted below.

Further studies have reported results consistent with the conclusions of the reviews discussed above, namely that cross-sectional differences in cognition between Aβ+ and Aβ- individuals are small and not observed in all studies, but Aβ+ individuals are consistently observed to show faster cognitive decline (Donohue, Sperling, *et al.*, 2017; Harrington, Lim, Ames, Hassenstab, Laws, *et al.*, 2017; Mormino *et al.*, 2017; Baker *et al.*, 2018; Rabin *et al.*, 2018). Several studies have replicated the finding of Duke Han *et al.* that the poorest cognition and greatest cognitive decline is seen in Aβ+ individuals who have additional tau pathology and/or neurodegeneration (Soldan *et al.*, 2016; Bilgel *et al.*, 2018; Ho and Nation, 2018; Sperling *et al.*, 2018).

However, several studies have provided evidence that even sub-threshold amounts of Aβ pathology (below the cut-off for Aβ+) may have a detectable effect on cognition. One approach to studying sub-threshold Aβ pathology is to identify “amyloid accumulators” –

initially Aβ- individuals whose Aβ levels are rising at repeated assessments. Three separate cohorts of amyloid accumulators have shown evidence of declining memory (but not executive function) (Farrell *et al.*, 2018; Landau *et al.*, 2018; Leal *et al.*, 2018), even in middle-aged adults (Farrell *et al.*, 2018). Grothe *et al.* (2017) found similar results with an alternative approach that does not require longitudinal assessment: they developed a 4-stage model of Aβ deposition based its pattern of progressive accumulation in different brain regions and found an association between higher stage and poorer episodic memory (but not executive function) in cognitively-normal older adults. They noted that the staging system accounts for sub-threshold amyloid accumulation, as most individuals in stages 1 and 2 would be classified as Aβ-. A third approach is to investigate associations between cognition and a continuous measure of Aβ, the standard uptake volume ratio derived from Aβ-PET (SUVR, see section 3.2.2). In a study of 1164 cognitively-normal adults aged 50-95 years, Knopman *et al.* (2018) reported that higher SUVR was associated with lower scores on a global cognitive composite across its full range, including at values below the cut-point for Aβ+. Similarly, in a study of adults aged 40-89, Farrell *et al.* (2017) reported that higher SUVR predicted greater cognitive decline over 4 years and was a better predictor than the dichotomous Aβ+/ Aβ- classification.

Taken together, these results suggest that subtle cognitive changes begin very early, decades before the onset of dementia. Indeed, recent preliminary estimates from the large Alzheimer's Disease Neuroimaging Initiative (ADNI) suggest that memory decline may begin about 20 years before dementia, at around the same time as amyloid accumulation reaches the threshold for positivity (which is about 10 years after changes are seen in CSF markers of Aβ and tau) (Palmqvist *et al.*, 2018). However, it is important to remember that these memory declines are likely to be so subtle that they would initially have little or no impact on daily life; a recent analysis of nearly 3000 cognitively-normal individuals concluded that it is not until age 70 that Aβ+ individuals are more likely to have a low score (defined as below the 10th percentile) on verbal memory tests, 10 to 15 years after the onset of Aβ+ (Jansen *et al.*, 2018).

2.3.3. Implications for my research

The above evidence suggests that subtle cognitive decline is detectable in cognitively-normal people with biomarker evidence of Aβ pathology. There is solid basis for the hypothesis that Aβ+ Insight 46 participants may have poorer cognitive performance at baseline, but previous cross-sectional studies have reported mixed results, so there is a need for greater understanding about the magnitude and nature of Aβ-related effects on

cognition. The evidence suggests that early changes may be detectable in multiple cognitive domains, with episodic memory receiving the most attention, but there is a need for better characterisation of the profile of subtle impairments that may emerge.

Given recent evidence that sub-threshold Aβ pathology may have detectable effects on cognition, I decided to use the continuous measure of Aβ (SUVR) in my analyses as well as the dichotomous Aβ+/ Aβ- classification (see section 3.2.2).

2.4. Which predictors of individual differences in cognitive performance are most important to account for?

As mentioned in section 2.1, accounting for factors associated with predictable variation in cognition between individuals should make it easier to detect subtle cognitive differences that may be associated with preclinical AD pathology. Three of the most obvious – age, sex and education – are highlighted in Figure 2-1, and discussed briefly below. As extensive research has been conducted on predictors of cognitive function across the life-course in the NSHD, Insight 46 has the advantage of being able to account for predictable variation in the cognitive performance of its participants; a summary of these key predictors follows in section 2.4.2.

2.4.1. Factors that have been shown to predict cognitive performance in other samples of older adults

This section contains some brief comments on the effects of age, sex and education on cognition in older adults. For each of the cognitive tests used within Insight 46, more detailed summaries of the literature on sex differences, ageing effects, and associations with education can be found in the introductions to the relevant chapters (4 to 8).

Most cognitive functions decline with age – notably memory, attention, processing speed and executive functioning – although there is considerable variability between individuals and some aspects of cognition are generally improved or maintained with age, such as semantic knowledge and procedural skills (Glisky, 2007). The association between age and accumulation of AD pathology means that these two factors can confound each other in analyses if not accounted for. For example, several studies have reported that the estimation of ageing effects on cognition in older adults is substantially reduced when accounting for brain pathology, particularly for measures of memory, although stronger ageing effects on executive function remain (Hassenstab *et al.*, 2016; Hedden *et al.*,

2016; Harrington, Lim, Ames, Hassenstab, Rainey-Smith, *et al.*, 2017; Hohman *et al.*, 2017; Harrington *et al.*, 2018).

Sex differences in cognition are a widely-researched and somewhat controversial topic. The commonly-held view is that males tend to perform better on spatial tasks and women on verbal tasks, although this is a simplification of a complex picture (Andreano and Cahill, 2009). Reports of superior verbal memory in women and superior visuospatial abilities in men are discussed in Chapters 4 and 5 respectively, in the context of verbal and visual memory tasks used in Insight 46. However, it is important to note here that sex differences in cognition are of particular interest to the field of Alzheimer's research because women are disproportionately affected by the disease, making up 65% of those living with dementia (Prince *et al.*, 2014). Whether or not this can be fully explained by women's longer life expectancy has not yet been conclusively established (Medeiros and Silva, 2019). There is evidence of sex differences in the relationships between risk factors and the development of dementia, including *APOE*-ε4 (Neu *et al.*, 2017), lifestyle factors (Norton *et al.*, 2014; Xu *et al.*, 2015; Podcasy and Epperson, 2016), and childhood intelligence (Snowdon *et al.*, 1996; Whalley *et al.*, 2000; McGurn *et al.*, 2008; Russ *et al.*, 2017; Huang *et al.*, 2018).

Associations between education and cognitive performance are widely observed across the life-span. For example, the US Alzheimer's Disease Centers' program has provided normative data for over 3000 clinically-normal older adults (~60-90 years) on cognitive tests that are widely-used in Alzheimer's research, showing that more years of education predicted better performance on all tests (Weintraub *et al.*, 2009; Shirk *et al.*, 2011). Similarly, a study of over 7000 adults with normal cognition or MCI (age range ~40-90 years) found a verbal memory advantage for those with higher education (Jansen et al., 2018). The degree to which educational attainment reflects cognitive ability is a matter of ongoing debate (see below), but it is clear that accounting for some measure of prior cognitive ability is important when seeking to identify subtle cognitive impairment or decline associated with preclinical AD. This may be particularly necessary when studying individuals with a high baseline level of cognitive ability, because they may experience significant decline over a long period before their performance falls below the normal range for their age (Rentz *et al.*, 2004, 2007).

2.4.2. Factors that have been shown to predict cognitive performance across adulthood in the NSHD cohort

Cognition has been assessed throughout childhood and adulthood in the NSHD. Details of the cognitive measures collected at different time-points are provided in section 3.2.1.

Measures of childhood cognitive ability have proved consistently predictive of cognition in middle age and later life (Richards and Sacker, 2003; Richards *et al.*, 2004, 2019; Davis *et al.*, 2017; Philippou *et al.*, 2018) as well as a range of physical health parameters such as grip strength, standing balance and chair rise speed (Kuh *et al.*, 2009; Cooper, Richards and Kuh, 2017), healthy dietary choice and exercise (Richards, Stephen and Mishra, 2010; Philippou *et al.*, 2018), and mental health outcomes (Koike *et al.*, 2017). Path models of cognition at ages 53 and 69 in the NSHD have shown that childhood cognitive ability has both direct and indirect effects on later-life cognition, with the indirect effects coming from its associations with educational attainment and adult socioeconomic position (defined according to occupational complexity), which are themselves predictors of cognition (Richards and Sacker, 2003; Richards *et al.*, 2019). Importantly, educational attainment and adult socioeconomic position also showed independent effects on later-life cognition at both ages, unaccounted for by childhood cognitive ability. Childhood cognitive ability, educational attainment and adult socioeconomic position were all partly explained by mother's educational attainment and father's socioeconomic position, but neither of these two variables showed a direct path to the later-life cognitive measures (Richards and Sacker, 2003; Richards *et al.*, 2019).

The finding of an independent effect of education on later-life cognition makes an important contribution to the debate about the extent to which general cognitive ability, or IQ, determines cognitive function throughout life, suggesting that education has a causal influence on subsequent cognition (Richards and Sacker, 2011). This issue is covered in more depth when discussing the results of my analyses (see Chapter 10).

2.4.3. Implications for my research

One of the key advantages of Insight 46 is that participants were all born during the same week, so the problem of disentangling the effects of age and brain pathology is essentially avoided. However, because there is a range in "age at assessment" due to data collection being carried out over 2.6 years (see section 3.6), potential ageing effects cannot be ruled out so should still be accounted for.

Insight 46 also has a unique advantage afforded by the availability of the life-course factors discussed above (childhood cognitive ability, educational attainment and adult socioeconomic position): predictable variation in cognition can be accounted for much more robustly than in most other studies which use educational attainment as a proxy for prior cognitive ability. This should increase the sensitivity of the cognitive measures to detect subtle differences that may be associated with Aβ pathology, as well as providing an opportunity to investigate and quantify the effects of childhood cognitive

ability, educational attainment and adult socioeconomic position on cognitive measures that are widely-used in research studies and clinical trials in preclinical AD.

2.5. Which cognitive measures are most sensitive to subtle cognitive decline in preclinical AD?

When the Insight 46 cognitive battery was designed, two approaches for improving the sensitivity of cognitive measures were adopted: 1) cognitive composites; 2) computerised tests with fine-grained outcome measures. A brief background to these two approaches is provided below.

2.5.1. Cognitive composites

A cognitive composite is a single score formed by combining scores from multiple cognitive tests. Cognitive composites are being used as outcome measures in several current clinical trials in preclinical AD (Weintraub *et al.*, 2018) and the US Food and Drug Administration has recently indicated openness to cognitive composite end-points (Kozauer and Katz, 2013). The rationale for their use in this context is that they may be sensitive to cognitive decline when effects are too small to be detectable on individual tests; several studies have demonstrated that composites are particularly sensitivity to cognitive decline in cognitively-normal individuals meeting criteria for preclinical AD (e.g. (Ayutyanont *et al.*, 2014; Langbaum *et al.*, 2014; Mormino *et al.*, 2017). In the context of clinical trials, an advantage of using a cognitive composite as a single outcome measure (rather than multiple cognitive outcomes) is that it reduces the likelihood of type 1 errors due to multiple comparisons (Ayutyanont *et al.*, 2014; Langbaum *et al.*, 2014; Jonaitis *et al.*, 2019).

There are two main approaches to composites – either to design them based on *a priori* hypotheses about which cognitive domains are affected earliest in the disease process (e.g. Donohue *et al.*, 2014; Lim, Snyder, *et al.*, 2016; Soldan *et al.*, 2016; Bateman *et al.*, 2017; van Bergen *et al.*, 2018; Wang *et al.*, 2018) or to derive them empirically from cognitive data (e.g. (Ayutyanont *et al.*, 2014; Langbaum *et al.*, 2014; Donohue, Sun, *et al.*, 2017).

Further details of a widely-used composite – the Preclinical Alzheimer Cognitive Composite (PACC, (Donohue *et al.*, 2014)) – are provided in Chapter 4, along with a discussion of evidence for its sensitivity to subtle cognitive decline in preclinical AD.

2.5.2. Computerised tests

Computerised tests offer scope to generate a range of precise outcome measures such as reaction times in milliseconds. Other advantages of computerised tests are ease of scoring, greater standardisation between testers, and the potential for presentation of items to be adapted automatically according to a participant's performance (Silverberg *et al.*, 2011). Computerised assessments offer the potential for remote testing, which can be of great value in studies seeking to screen or recruit large numbers of people, such as the PROTECT study which aims to administer repeated online cognitive assessments to at least 50,000 older adults over ten years (http://www.protectstudy.org.uk). It also allows the measurement of cognitive function in real-world environments (Sliwinski *et al.*, 2018).

Even for relatively simple tasks where responses could be scored manually, computerised administration can provide additional information on the process by which an individual completes a task. For example, eye-tracking measures could be recorded during a visual recognition task (Bott *et al.*, 2018), or a digital pen could reveal organisational strategies for completing a drawing task (Davis *et al.*, 2014). Evidence suggests that the earliest stages of cognitive decline are characterised by the adoption of compensatory strategies (Jessen *et al.*, 2014), so these kind of 'process measures' are potentially useful for detecting individuals who may be working harder or thinking longer to achieve the same result, perhaps before showing more overt decline (Davis *et al.*, 2014).

The field of cognitive assessment in preclinical AD is increasingly moving towards using computerised assessments for the reasons mentioned above (Silverberg *et al.*, 2011; Rentz *et al.*, 2013; Ritchie *et al.*, 2017; Hassenstab *et al.*, 2018).

2.5.3. Implications for my research

Given that cognitive composites and computerised assessments are increasingly being used in preclinical AD research, it is important to evaluate how these measures compare to individual scores from standard paper-and-pencil cognitive tests in Insight 46. This should contribute to the evidence base on the most sensitive cognitive measures to use as outcomes in clinical trials.

2.6. Other influences on cognitive decline

It is important to be aware that other brain pathologies influence cognitive performance, as well as Aβ pathology. The rich dataset of biomarker and genetic measures collected in Insight 46 provides the opportunity to investigate some of these other pathologies and genetic risk factors. My aim in including some of these variables in my models of cognitive performance was partly to increase the sensitivity of my analyses to the effects of Aβ by accounting for potential confounding effects of other variables, but also to generate new evidence about their associations with cognition. The three variables that are explored throughout this thesis are *APOE*-ε4, whole brain volume and global white matter hyperintensity volume (WMHV). As these variables are not the primary focus of my research, I have not conducted a comprehensive review of evidence for their associations with cognition, but some introductory comments are provided below.

2.6.1. APOE-ε4

The strongest genetic risk-factor for sporadic AD is the apolipoprotein gene (*APOE*), which occurs in three different alleles: ε2, ε3 and ε4. Each person has 2 copies, with the most common combination being ε3/ε3 (60% of the population (Alzheimer's Society, 2016a)). The ε4 allele is associated with increased lifetime risk of developing AD in a dose-dependent manner: ε4-heterozygotes have a fourfold increase in risk, and ε4-homozygotes have a tenfold increase in risk (compared to ε3-homozygotes) (Alzheimer's Society, 2016a). The mechanism for this increased risk is understood to relate to reduced clearance of Aβ in *APOE*-ε4 carriers, resulting in accumulation of a higher burden of Aβ plaques (Kline, 2012). Among older adults, *APOE*-ε4 carriers are around twice as likely as non-carriers to be Aβ+, with around 50% of ε4 carriers aged ~50-90 years estimated to be Aβ+ (Rowe *et al.*, 2007; Jack *et al.*, 2017).

The interactions between *APOE*-ε4 and Aβ on cognition are yet to be fully understood. Many studies investigating associations between *APOE*-ε4 and cognition have not included measures of Aβ, making it impossible to know whether Aβ may account for cognitive deficits observed in ε4-carriers, although evidence for such deficits is mixed (see O'Donoghue *et al.* (2018) for a review). Several studies that were able to account for Aβ have reported that Aβ- ε4-carriers seem to experience normal cognitive aging, suggesting that Aβ is necessary for memory decline (Lim, Laws, *et al.*, 2016; Lim *et al.*, 2018). However, there is evidence that Aβ and *APOE*-ε4 may interact such that clinically-normal Aβ+ ε4-carriers experience faster memory decline than Aβ+ non-carriers (Mormino *et al.*, 2014; Lim *et al.*, 2015) and have a higher risk of progression to MCI or dementia (Dang *et al.*, 2018).

Beneficial effects of *APOE*-ε4 have also been reported, particularly during youth, with growing evidence that *APOE*-ε4 is associated with a diverse range of survival advantages including resistance to certain infections, increased fertility, increased fitness in infancy and slightly superiority in some aspects of cognition (Duke Han and Bondi, 2008; Zetterberg *et al.*, 2009; Tuminello and Duke Han, 2011; Smith, Ashford and Perfetti, 2019). This is an example of antagonistic pleiotropy – the principle that some genes have both beneficial and detrimental effects, with the detrimental effects generally manifesting after reproductive age (Austad and Hoffman, 2018). However, the putative beneficial effects of *APOE*-ε4 on cognition in earlier life are controversial and the literature has not yet reached a consensus (see reviews in (Tuminello and Duke Han, 2011; O'Donoghue *et al.*, 2018)).

2.6.2. White matter hyperintensity volume (WMHV)

White matter hyperintensities are commonly seen on brain MRI scans of older adults and reflect lesions caused by cerebral small vessel disease (Prins and Scheltens, 2015). White matter lesions are considered to be the primary pathology of vascular dementia – the second most common type of dementia after AD (Alzheimer's Society, 2018) – but are also commonly seen in patients with AD (Prins and Scheltens, 2015; Alosco *et al.*, 2018). Greater WMHV in cognitively-normal older people is associated with increased risk of dementia (Debette and Markus, 2010; Payton *et al.*, 2018) and with decline in cognition, particularly processing speed and executive function, although the associations with cognition generally appear to be weak (Gunning-Dixon and Raz, 2000; De Groot *et al.*, 2002; Oosterman *et al.*, 2004; Prins *et al.*, 2005; van Dijk *et al.*, 2008; Prins and Scheltens, 2015; Kaskikallio *et al.*, 2019).

2.6.3. Whole brain volume

Loss of brain volume, or atrophy, occurs gradually with age, but accelerated atrophy is a feature of neurodegenerative diseases that tracks closely with the progression of symptoms (Fox and Schott, 2004). Accelerated atrophy typically occurs relatively late in the AD pathological continuum, some years after accumulation of Aβ and tau pathology (Jack *et al.*, 2013). Therefore, evidence of significant atrophy is not expected among Insight 46 participants as they are at an age when those who are destined to develop dementia are still likely to be many years from symptoms (Prince *et al.*, 2014).

Aside from atrophy, another area of interest is whether brain volume is associated with cognitive ability at younger ages. This possibility has received attention in the long-

running debate about the neural basis for differences in intelligence (McDaniel, 2005), and two recent studies of large samples of young adults have reported that larger brain volume is weakly associated with better cognition, notably in terms of faster processing speed (Magistro *et al.*, 2015; Takeuchi *et al.*, 2017).

2.7. Research Questions and Hypotheses

2.7.1. Research Questions

My research sought to address three broad questions:

i) What are the patterns of performance on each cognitive test?

If the cognitive tests used in Insight 46 are to have future application as markers of preclinical AD, it is important to confirm that the tasks work appropriately in this age group (e.g. are there any floor or ceiling effects?) and to describe the normal range of performance. As the computerised tests have several different conditions (e.g. easier and harder levels) and outcomes, (e.g. speed and accuracy), identifying key outcome measures that may be particularly sensitive to subtle cognitive decline requires an understanding of patterns of performance across the various aspects of each task, such as "Is there a trade-off between speed and accuracy?" or "To what extent is memory recall affected by the number of items to be remembered?".

ii) What are the relationships between demographic and life-course predictors and performance on the cognitive tests?

As discussed in section 2.4, I sought to understand more about predictors of cognitive performance at age ~70 by investigating the effects of childhood cognitive ability, education, adult socioeconomic position, age at assessment, and sex. As well as aiming to generate novel evidence about predictors of performance on the specific cognitive tests used in Insight 46, the purpose of this was to be able to account for predictable variation between individuals, which may increase the sensitivity of the tests to detecting subtle cognitive changes associated with brain pathology.

iii) **What are the relationships between biomarkers and genetic risk factors for AD and performance on the cognitive tests?**

As discussed in section 2.6, I aimed to investigate whether the following measures were associated with cognitive performance: Aβ pathology (dichotomous amyloid status and continuous SUVR), whole brain volume, global WMHV and *APOE*-ε4.

2.7.2. Hypotheses

Overarching hypotheses are listed below. Specific hypotheses for each cognitive test are stated in the relevant chapters (4-8):

i) Higher childhood cognitive ability, educational attainment and adult socioeconomic position will show independent associations with better cognitive performance.
ii) As the sample size of Insight 46 (n = 502) is large for a neuropsychological study, subtle sex differences in cognition will be detectable.
iii) Participants with elevated Aβ deposition will show evidence of subtle cognitive deficits.
iv) WMHV and whole brain volume will show evidence of weak associations with cognition, particularly processing speed.
v) Composite and computerised measures will be more sensitive to brain pathology than standard paper-and-pencil cognitive tests.

Given the mixed evidence for effects of *APOE*-ε4 on cognition independent of Aβ (see 2.6.1), I did not make an overarching hypothesis about the effects of *APOE*-ε4, but I made a specific hypothesis for one of the cognitive tasks where previous studies have reported an effect (see Chapter 5).

3. GENERAL METHODOLOGY

3.1. Recruitment and data collection

The target sample for Insight 46 was NSHD participants who had a specific set of life-course data available including attendance at a clinic visit at age 60-64, at least one measure of childhood cognition, and various measures of physical health and lifestyle during adulthood – full details have been published in the protocol paper (Lane *et al.*, 2017). Additionally, participants were required to be willing to attend a clinic-based visit at UCL and to have no contraindications to MRI or PET, such as severe claustrophobia, or metal within the body (e.g. pacemakers and intracranial clips). Participants were sent an invitation by post and then screened by telephone if interested. A recruitment flow-chart is provided in Figure 3-1 and further details have been published (Lane *et al.*, 2017; James *et al.*, 2018).

502 participants were recruited to Insight 46 and attended a baseline assessment between May 2015 and January 2018. Ethical approval for Insight 46 was granted by the National Research Ethics Service (NRES) Committee London (14/LO/1173). All participants gave written informed consent.

The study protocol included: cognitive tests; a clinical interview with a neurologist; a structured physical and neurological examination; assessment of visual, auditory and olfactory function; self-administered questionnaires measuring subjective cognitive decline, anxiety, dental health, handedness and sleep; collection of blood and urine for clinical and genetic biomarker identification; neuroimaging comprising simultaneous acquisition of β-amyloid PET and MRI data. Each participant had an informant who completed the AD8 interview, a brief screening tool for dementia (Galvin *et al.*, 2005). Further details are provided in section 3.2 and in the protocol paper (Lane *et al.*, 2017).

While all assessments were typically completed on one day, 62 participants had to have their scans rescheduled for a later date, with a median interval of 49 days (IQR = 26 – 77; range = 1 – 216). For the purpose of calculating age at assessment, the date of cognitive testing was used.

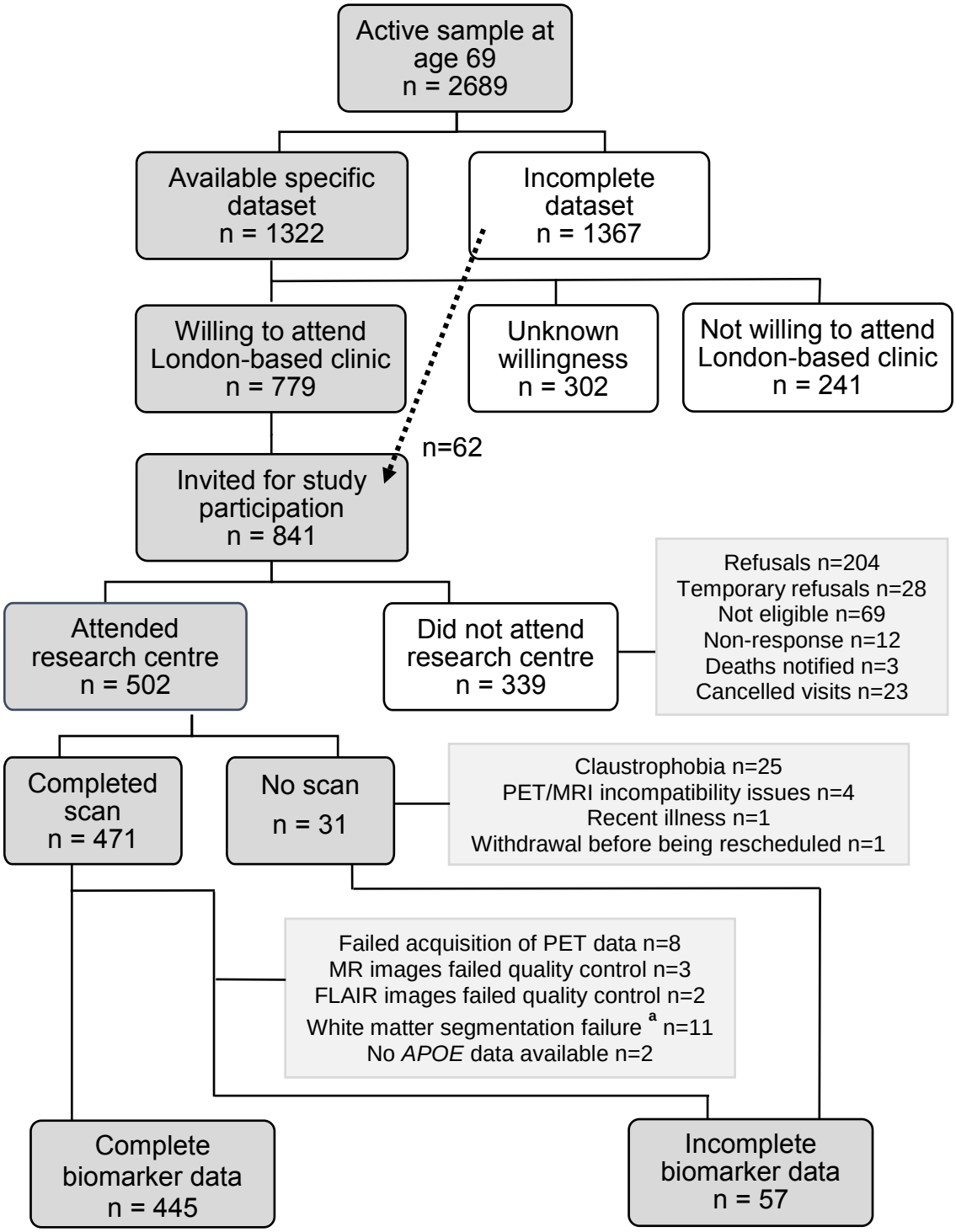


Figure 3-1. Flowchart of recruitment and data acquisition

The specific dataset refers to a set of life-course data which formed the original criteria for Insight 46 eligibility – see text for further details. To reach our target sample size, these criteria were relaxed to remove the requirement for a previous measure of lung function, smoking or physical exercise, enabling recruitment of a further 62 individuals. Details of the biomarker measures are provided in section 3.2.2.

[a] In most cases, this was due to erroneous segmentation of vascular abnormalities such as stroke or demyelination.

FLAIR = fluid attenuated inversion recovery MRI; MR = magnetic resonance; PET = positron emission tomography

3.2. Materials and measures

3.2.1. Cognitive Battery

As my research focuses on the cognitive data, the methods for its collection, processing and analysis are described in detail in this section.

The cognitive battery was designed to have the following characteristics: 1) to be complementary to prior NSHD cognitive assessments, the most recent of which included the Addenbrooke's Cognitive Examination III (Hsieh *et al.*, 2013), a word-list learning task and a timed letter search task (Silverwood *et al.*, 2014); 2) to have a total duration of less than 90 minutes; 3) to include standardised cognitive tests to allow results to be compared with normative data; 4) to include experimental computerised cognitive tests that might be more sensitive to subtle cognitive decline; 5) to be informed by a review of protocols and results of current large-scale initiatives and clinical trials in preclinical AD (including the A4 trial (Sperling *et al.*, 2014) and the Alzheimer's Prevention Initiative (Reiman *et al.*, 2011)).

An overview of the cognitive tests in the Insight 46 battery is provided in Table 3-1. Four of these tests are standardised clinical neuropsychological tests that have been widely used in studies of preclinical AD and the others are more novel tests. To facilitate comparison of the Insight 46 battery with the cognitive tests administered previously in NSHD, I compiled a table which was included in the Insight 46 protocol paper (Lane *et al.*, 2017); an updated copy is included here which contains more citations of recent relevant publications (Table 3-2).

The order of tests within the Insight 46 cognitive battery was designed to allow for the necessary delay times between the various recall trials of the memory tests (Table 3-3).

Chapters 4 to 9 of this thesis present my work analysing and interpreting the results of these cognitive tests, with the exception of the Instructionless Eyetracking and Irrelevant Distractor tests. These two tests were only performed on sub-samples of the Insight 46 cohort because the Irrelevant Distractor test was dropped from the battery and replaced by the Instructionless Eyetracking test in February 2017 (mid-way through data collection). This change was made because many participants found the Irrelevant Distractor test very difficult due to its fast presentation speed: 3% of participants declined to attempt the task after struggling with the practice, and 48% of those who attempted it had error rates of greater than or equal to 50% (the level that would be achieved by chance) in at least one part of the task. Therefore, in view of the fact that the task was

unpopular with participants and seemed unlikely to yield informative results, I decided along with my supervisors to remove it. The Instructionless Eyetracking test was chosen as a replacement based on evidence that eye movement patterns may reveal early changes in memory, attention, visuospatial and executive processes (Pereira *et al.*, 2014; Primativo *et al.*, 2017) Data from the Instructionless Eyetracking test have not yet been analysed. These two tests are not discussed further in this thesis.

Table 3-1. Cognitive tests in the Insight 46 cognitive battery

Name of Test	**Source**	**Cognitive domain**	**Brief description** (see subsequent chapters for further details)
Matrix Reasoning	Wechsler Abbreviated Scale of Intelligence (WASI) (Wechsler, 1999)	Fluid intelligence / non-verbal reasoning	A test of non-verbal reasoning. Participants are shown a matrix of geometric shapes with a piece missing and are required to select the missing piece from five options. (See Chapter 4.)
Mini-Mental State Examination	(Folstein, Folstein and McHugh, 1975)	Global cognition	A 30-point screening tool for cognitive impairment, which covers multiple cognitive domains including orientation to time and place, registration, recall, attention, calculation, visuospatial function, language, repetition, writing, reading, following a 3-stage command. (See Chapter 4.)
Logical Memory Immediate and Delayed Recall	Wechsler Memory Scale – Revised (WMS-R) (Wechsler, 1987)	Short-term verbal memory	A 25-point test of free recall. The participant is read a story and asked to recall it immediately and after a delay of ~20 minutes. (See Chapter 4.)
Digit-Symbol Substitution	Wechsler Adult Intelligence Scale – Revised (WAIS-R) (Wechsler, 1981)	Executive function / processing speed	Participants are given a code table of digits paired with symbols. On a worksheet with rows of digits, they are asked to fill in the corresponding symbols as quickly and accurately as possible. The score is the number of symbols completed correctly within 90 seconds. (See Chapter 4.)
Face-name Associative Memory Exam (FNAME-12A)	(Papp *et al.*, 2014) stimuli available from the authors on request	Verbal and visual memory (recall and recognition)	Participants are given two exposures to 12 faces paired with a name and an occupation. After each exposure and after delays of ~10 and ~35 minutes, they are asked to recall the name and occupation associated with each face. Facial recognition is also assessed at the 35-minute delay by asking participants to identify each learned face from two distractors. (See Chapter 4.)

Name of Test	**Source**	**Cognitive domain**	**Brief description** (see subsequent chapters for further details)
"What was where?"	Shortened version designed for Insight 46 by Dr Yoni Pertzov (Pertzov *et al.*, 2012)	Visual short-term memory binding	Participants are shown 1 or 3 objects on a screen and asked to remember the objects and their locations. After a delay of 1 or 4 seconds, they are required to identify the learned object from a distractor and place it in its remembered location. (See Chapter 5.)
Choice Reaction Time and Response Inhibition	Designed by Prof. Sebastian Crutch based on Aron *et al.* (2004)	Executive function / response inhibition	Part 1 assesses Choice Reaction Time – participants are required to respond to arrows or words indicating left or right. Part 2 presents word/arrow combinations which are congruent or incongruent – participants must respond according to the cue ('arrow' or 'word'). (See Chapters 6 and 7.)
Circle-tracing and Serial Subtraction	Shortened version designed for Insight 46 by Prof. Julie Stout's lab, based on Say *et al.* (2011)	Visuomotor integration; calculation; dual tasking	Participants are asked to trace round a circle on a tablet screen as quickly and accurately as possible, with either direct or indirect visual feedback. This task was performed with and without concurrent serial subtraction (single / dual task). (See Chapter 8.)
Instructionless Eyetracking *	Designed by Dr Silvia Primativo *et al.* (2017)	Various domains including memory and social cognition	Participants' eye movements are tracked while they are asked simply to watch a screen which presents images, sentences and moving patterns.
Irrelevant Distractor*	Shortened version designed for Insight 46 by Prof. Nilli Lavie's lab (Forster and Lavie, 2008)	Attention	This computerised letter search task requires participants to make a rapid decision as to which target letter has appeared in the search display. There are three different load conditions (the number of letters to be searched). On some trials a distractor appears which could be task-irrelevant (a cartoon character) or task-relevant (a letter).

** Instructionless Eyetracking was introduced in February 2017, replacing the Irrelevant Distractor task – see text for explanation.*

Table 3-2. Cognitive measures in the National Survey of Health and Development from age 8 to 69

Cognitive Domain		Age 8	Age 11	Age 15	Age 26	Age 43	Age 53	Age 60-64	Age 69	Insight 46 (ages 69-71 & 71-73)
PREMORBID IQ							National Adult Reading Test			
VERBAL REASONING and READING COMPREHENSION		Reading comprehension test	Verbal abilities test	Verbal section of Alice Heim Test; Watts-Vernon reading comprehension test.	Watts-Vernon reading comprehension test					
NON-VERBAL REASONING		Picture intelligence test	Non-verbal abilities test	Non-verbal section of Alice Heim Test						Matrix Reasoning
MEMORY	Short-term Verbal Memory					Word list	Word list	Word list	Word list	Logical Memory
	Short-term Visual Memory					Visual memory test				"What was where?"
	Short-term Associative Memory									FNAME-12
	Prospective Memory						Prospective memory test			
NUMERACY			Arithmetic test	Mathematics test						

LITERACY		Word reading test; Vocabulary test	Word reading test; Vocabulary test							
EXECUTIVE FUNCTION	Verbal Fluency						Category fluency		Phonemic fluency and category fluency (as part of the ACE-III)	
	Reaction Time							Simple RT; Choice RT		Choice RT
	Inhibition									Response Inhibition
PROCESSING SPEED and ATTENTION						Letter Search	Letter Search	Letter Search	Letter Search	Digit-Symbol Substitution; Irrelevant Distractor
VISUOMOTOR INTEGRATION						Timed pegboard				Circle-tracing
GENERAL / MULTIPLE DOMAINS									ACE-III	MMSE; Instruction-less Eyetracking

Details of the cognitive measures can be found in the following papers: (Gaysina *et al.*, 2014) *(ages 8, 11, 15, 60-64); (Pigeon, 1964)(ages 8 and 11);* (Ross and Simpson, 1971a, 1971b) *(ages, 8, 11, 15 and 18);* (Gale *et al.*, 2012) *(ages 11, 15, 53);* (Stewart, Hardy and Richards, 2015) *(ages 15 and 53);* (Albanese *et al.*, 2012) *(ages 26, 43, 53);* (Richards, Hardy and Wadsworth, 2005) *(ages 26, 43 and 53);* (Davis *et al.*, 2016) *(age 43);* (Richards *et al.*, 2005) *(ages 43 and 53);* (Davis *et al.*, 2017; Rawle *et al.*, 2018) *(ages 43, 53, 60-64 and 69);* (Richards *et al.*, 2014) *(ages 53, 60-64);* (Hurst *et al.*, 2013; Silverwood *et al.*, 2014; Masi *et al.*, 2018) *(age 60-64);* (Lane *et al.*, 2017; Lu *et al.*, 2019) *(Insight 46).*

ACE-III = Addenbrooke's Cognitive Examination; FNAME-12 = Face-Name Associative Memory Examination (12 item); MMSE = Mini Mental State Examination; RT = reaction time; WAIS-R = Wechsler Adult Intelligence Scale - Revised; WASI = Wechsler Abbreviated Scale of Intelligence; WMS-R = Wechsler Memory Scale - Revised

Table 3-3. Order and timings of the Insight 46 cognitive battery

Cognitive test	Mean duration (minutes)
Circle-tracing pre-exposure	8
Logical Memory Immediate Recall	2
Matrix Reasoning	10
Instructionless Eyetracking (or Irrelevant Distractor for participants tested before 08/02/17)	13
Logical Memory Delayed Recall	1
FNAME-12 trials 1 and 2	8
Choice Reaction Time and Response Inhibition	10
FNAME-12 trial 3	3
Digit-Symbol Substitution	3
"What was where?"	10
Circle-tracing and Serial Subtraction	10
FNAME-12 trial 4	4

Note: mean durations are based on a sample of 51 participants who all completed the Irrelevant Distractor test

3.2.2. Biomarker measures and genetic risk factors

Full details of biomarker measures collected in Insight 46 are provided in the protocol paper (Lane *et al.*, 2017). Details are provided below for only those variables which I have used in subsequent analyses.

β-amyloid PET and MRI data were collected simultaneously during a 60-minute scanning session on a Biograph m MR 3 T PET/MRI scanner (Siemens Healthcare, Erlangen), with intravenous injection of a β-amyloid PET ligand, 370 MBq florbetapir F18 (Amyvid). Aβ deposition was quantified using a global **Standard Uptake Volume Ratio (SUVR)**, using 10 minutes of static steady state Florbetapir data ~50 mins post-injection. The SUVR was calculated from a cortical grey matter composite (composed of frontal,

temporal, parietal, and cingulate regions), with a reference region of eroded subcortical white matter. Aβ-PET attenuation correction was performed using pseudo-CT correction. Due to technical issues, only console attenuation correction was available for 26 participants. For these participants a pseudo-CT corrected value was imputed based on their console value.

A cut-point for Aβ positivity was determined using a mixture model to define two Gaussians, and using the 99th percentile of the lower (Aβ negative) Gaussian, at SUVR > 0.6104. This gives a dichotomous variable of **amyloid status**: Aβ+ (elevated levels of β-amyloid) or Aβ- (normal levels of β-amyloid).

Volumetric T1-weighted and FLAIR (fluid attenuated inversion recovery) MR images underwent visual quality control, before being processed using automated pipelines (Lane *et al.*, 2017). **Whole brain volume** was generated from high resolution 3D T1-weighted MRI using automated segmentation with manual editing (Leung *et al.*, 2011). Total intracranial volume (TIV) was calculated using statistical parametric mapping (SPM) software (SPM12; http://www.fil.ion.ucl.ac.uk/spm) (Malone *et al.*, 2015). Global **white matter hyperintensity volume (WMHV)** was generated from T1-weighted and FLAIR MRI using Bayesian Model Selection (BaMoS), an automated segmentation algorithm based on a multivariate Gaussian mixture model (Sudre *et al.*, 2015), followed by visual quality control, generating a global WMHV including subcortical grey matter but excluding infratentorial regions.

APOE genotyping was conducted at LGC, Hoddesdon UK. For the analyses presented in this thesis, participants were classified into two categories based on the presence of the ***APOE*-ε4** allele: ε4-carriers and non-carriers.

3.2.3. Major neurological and psychiatric conditions

Participants were coded as having a major neurological or psychiatric condition according to the criteria in Table 3-4. Participants not meeting any of these criteria are hereafter referred to as cognitively normal and represent a sample who might be considered eligible for a clinical trial of cognitively healthy individuals, free from possible confounding comorbidities. This does not imply that all participants with a major neurological or psychiatric condition necessarily had a measurable cognitive impairment.

Table 3-4. Criteria for major neurological or psychiatric conditions

- Clinical evidence of dementia, Parkinson's disease and other neurodegenerative disorder
- Psychiatric disorder requiring anti-psychotic medication
- Depression requiring electroconvulsive shock therapy
- Epilepsy requiring active treatment
- Radiological evidence of traumatic brain injury or major neurosurgery
- Clinical diagnosis or radiological features of multiple sclerosis
- Clinical diagnosis of stroke, or radiological evidence of cortical ischaemia or haemorrhage consistent with previous cortical stroke
- Radiological evidence of possible brain malignancy
- Mild cognitive impairment (MCI) defined as follows, based on published criteria (Petersen *et al.*, 2013):
    - No clinical evidence of dementia
    - **AND** participant concern regarding their cognition (memory or cognitive difficulties more than other people the same age, or if they felt they would seek medical attention regarding their difficulties) and/or informant concern regarding the participant's cognition (AD8 score ≥ 2)
    - **AND** objective evidence of either an amnestic (Logical Memory delayed recall ≥ 1.5 *SD* below the mean) and/or non-amnestic deficit (Digit-Symbol Substitution score ≥ 1.5 *SD* below the mean). These cognitive tests were chosen for defining a cognitive deficit on the basis of their normal distribution across the entire cohort.

3.2.4. Life-course data

As discussed in sections 2.4.2 and 2.4.3, three key predictors of life-time cognition among NSHD participants are childhood cognitive ability, educational attainment and adult socioeconomic position. These variables are defined as follows, using definitions used in many previous analyses from the NHSD (e.g. Richards and Sacker (2003); Rawle *et al.* (2018)).

Childhood cognitive ability was measured at age 8 using four tests of verbal and non-verbal ability devised by the National Foundation for Education Research (Pigeon, 1964) (see Table 3-2). The sum of scores from these four tests was standardised into a z-score representing overall cognitive ability. If these data were missing, the standardised score

from the tests at age 11 was used (or if this was missing, the standardised score from the tests at age 15). Note that these scores were standardised for the full cohort (N=5362) rather than the Insight 46 sample.

Educational attainment was recorded at age 26. Highest educational or training qualification achieved was classified using the Burnham scale (Department of Education and Science, 1972) and grouped into five categories: no qualification, below O-levels (vocational), O-levels and equivalents, A-levels and equivalents, higher education (degree and equivalents).

Adult socioeconomic position (SEP) was derived from participants' own occupation at age 53, or earlier if this was missing. Occupations were coded according to the UK Registrar General's Standard's Occupational Classification, then classified into six categories: unskilled, partly skilled, skilled manual, skilled non-manual, intermediate, professional.

3.3. Completion rates and reasons for missing data

Completion rates for the biomarker measures and reasons for missing biomarker data are shown in Figure 3-1.

Figure 3-2 shows completion rates for cognitive assessments and Table 3-5 gives the reasons for missing cognitive data.

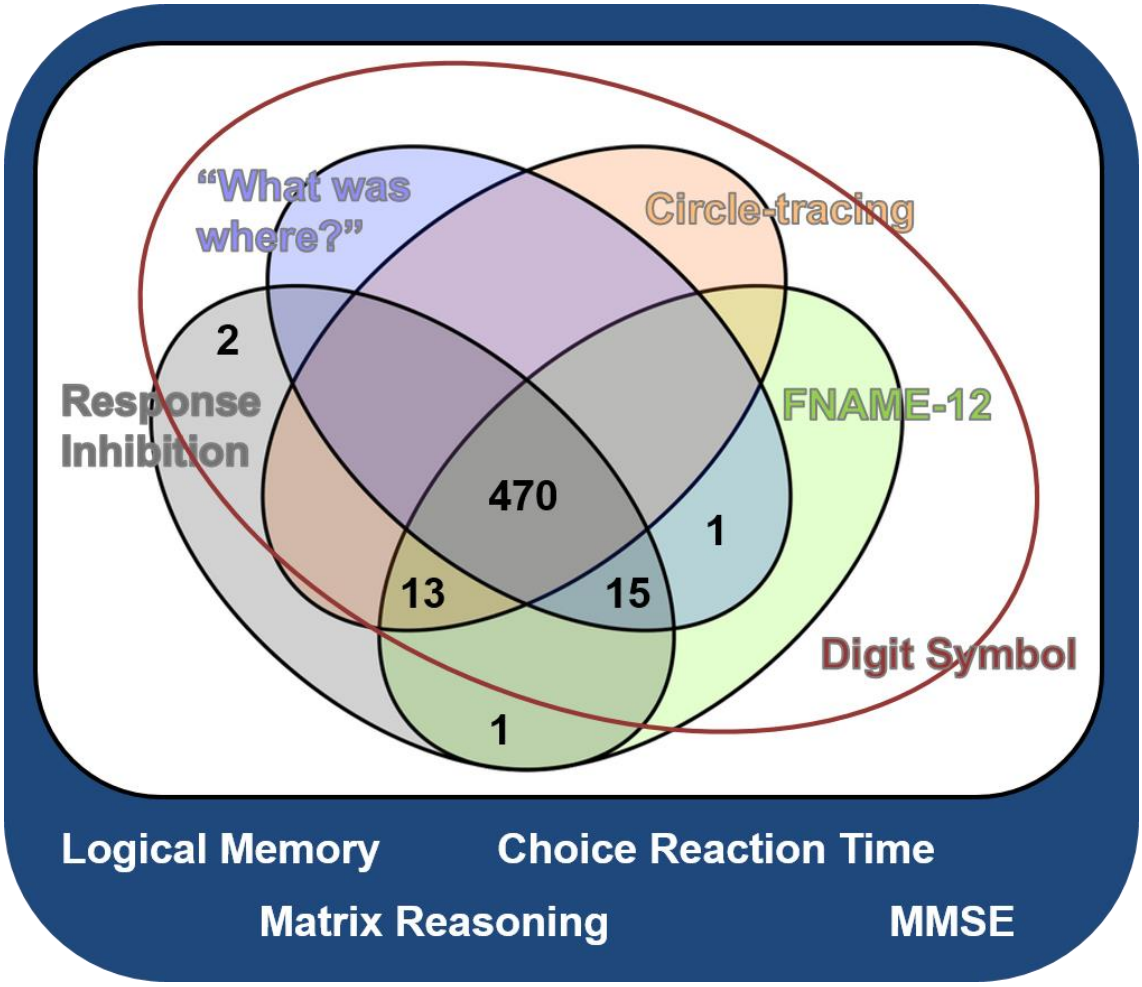


Figure 3-2. Venn diagram of completion rates for Insight 46 cognitive tests

As shown here, 470 out of 502 participants completed all of the cognitive tests. There were no missing data for the Logical Memory, MMSE, Matrix Reasoning and Choice Reaction Time tests, but all other tests had some missing data.

Table 3-5. Reasons for missing cognitive data

Cognitive test	Reason for missing data	Number of participants
Circle-tracing	Technical problems	11
	Participant unable due to tremor in hands	3
	Participant declined to complete task	4
	Lack of time*	1
Response Inhibition	Data file not saved	1
"What was where?"	Not administered due to technical problems	10
	Participant started the task but declined to complete it	1
	Participant had recently had hand surgery and could not operate the touchscreen	1
	Lack of time*	4
FNAME-12	Technical problems	1
	Data from this task are partially missing – the fourth trial was administered in place of the third trial, because the time elapsed since the second trial was already 30 minutes.*	1
Digit-Symbol	Participant declined to attempt this task	1

** Three participants took longer than usual to complete tasks due to neurological conditions affecting cognition. The reason for the fourth participant who ran out of time is unknown.*

3.4. Processing of cognitive data

Test scores collected on paper (MMSE, Matrix Reasoning, Logical Memory, Digit-Symbol, FNAME-12) were inputted into XNAT, a customised web-based database (www.xnat.org). XNAT was also used to record reasons for any missing data. After consultation with the various authors of the computerised cognitive tests, I wrote programmes in Stata v15 (StataCorp, 2017) to clean the data and generate the outcome variables of interest, described in the relevant chapters (4 to 9).

3.5. Data analysis

Chapters 4 to 8 report the results of the cognitive tests administered to Insight 46 participants. In each chapter, analyses are split into two parts. Part 1 addresses the first and second research questions posed in section 2.7, namely "What are the patterns of performance on each cognitive test?" and "What are the relationships between demographic and life-course predictors and performance on the cognitive tests?". Part 2 addresses the third research question posed in section 2.7, namely "What are the relationships between biomarkers and genetic risk factors for AD and performance on the cognitive tests?". My approach to these analyses is described below in sections 3.5.1 and 3.5.2, followed by an explanation of the statistical models for the computerised tests (section 3.5.3).

All analyses were conducted in Stata v15 (StataCorp, 2017). Results were considered statistically significant if the chance of a false positive finding was below 0.05.

3.5.1. Patterns and predictors of performance

Most of the computerised cognitive tests generate trial-by-trial outcome data (e.g. reaction time for each individual response). In order to explore patterns of performance in as much detail as possible, these trial-by-trial data were used rather than summary scores, and relationships between different task outcomes (e.g. speed and accuracy) were explored.

For analyses investigating the effects of demographic and life-course predictors on cognition, all participants were included. The rationale for including all participants (rather than excluding those with a major neurological or psychiatric condition) was to be able to describe the predictors of cognition in as representative a sample as possible. Neurological and psychiatric conditions are not uncommon among older people, and their inclusion takes advantage of a key strength of Insight 46: it is a population-based sample, likely to be more representative of the general population than most studies in ageing and AD research (see section 10.3.1 for a discussion of the representativeness of Insight 46 and comparison with other studies). However, as many neurological and psychiatric conditions affect cognition, it is important to account for these effects which may confound the effects of other predictors (for example, a lifelong condition affecting cognition is likely to be associated with lower childhood cognitive ability and educational attainment). Therefore, a dichotomous factor coding for the presence of a major neurological or psychiatric condition (yes *vs.* no) was included in all analyses. While the different neurological and psychiatric conditions are associated with different profiles of

cognitive impairment, it was not appropriate to compare one condition against another due to the very small numbers of participants with each condition (see section 3.6).

Multivariate models were used since I was interested in determining the independent effects of each predictor (e.g. the effect of educational attainment *accounting for childhood cognitive ability and all other predictors).* Predictors in the models were sex, age at assessment, childhood cognitive ability, education, adult socioeconomic position and presence of a neurological or psychiatric condition. (See section 3.2.4 for definitions of variables.) Education was treated as a continuous variable with values of 1 to 5 corresponding to the five categories defined in section 3.2.4. Similarly, adult socioeconomic position was treated as a continuous variable with values of 1 to 6.

3.5.2. Associations with biomarkers and APOE-ε4

The second analysis section of each chapter investigates associations between cognitive performance and the following biomarkers measures: Aβ pathology, whole brain volume, WMHV and *APOE*-ε4. (See sections 2.6 and 3.2.2 for background and definitions of these measures). In terms of Aβ pathology, I chose to focus primarily on amyloid status (Aβ+ *vs.* Aβ-), since it identifies those whose Aβ burden is abnormal and is integral to standard criteria for preclinical AD (see section 2.2). However, I also reran all analyses using SUVR, the continuous measure of Aβ burden, instead of dichotomous amyloid status, in recognition of the fact that the cut-point for Aβ+ is arbitrary and sub-threshold accumulation of Aβ may have an impact on cognition (see section 2.3.3). To check whether associations between SUVR and cognition were sensitive to the inclusion of the imputed SUVR values (see section 3.2.2), these analyses were additionally rerun excluding the 26 participants with imputed data.

Participants meeting criteria for major neurological or psychiatric conditions (see section 3.2.3) were excluded from these analyses, as the aim was to investigate whether the cognitive tests may be sensitive to pathology in cognitively-normal individuals. The cognitively-normal sub-sample represents a non-demented, non-MCI population free from known possible confounding comorbidities, who might be considered eligible for a clinical trial targeting preclinical AD.

Since I was interested in determining the relative contributions of each biomarker to the cognitive outcome measures, and I know that they are not necessarily independent, multivariate models were used so that the reported effects of each biomarker are adjusted for all the others. The models also included the factors tested in the first section

(sex, age at assessment, childhood cognitive ability, education and adult socioeconomic position).

3.5.2.1. Selection of summary outcome variables

As the computerised tests are complex with many trials and various different conditions and outcomes, I aimed to identify a few summary outcomes for each test that capture the key aspects of performance. Reducing the data to a small number of summary outcomes is desirable as it allows me to describe associations between biomarkers and participants' overall pattern of performance (as opposed to their trial-by-trial performance) and enables key outcomes to be compared across the different cognitive tests in the Insight 46 battery in a standardised manner. It is important to be able to compare the magnitude of subtle cognitive deficits associated with Aβ pathology across different cognitive tasks and domains, as this may deepen our understanding of the nature and timing of cognitive decline associated with accumulation of pathology – this is addressed in Chapter 9. Some of the computerised tests have obvious summary outcomes, such as mean RT, but for other tests the most meaningful summary outcome depends on the pattern of results across different conditions of the task, such as discrepancies in performance between easy and hard conditions. In choosing summary outcomes, I prioritised those which showed promise in sensitivity to Aβ pathology.

3.5.3. Statistical models

3.5.3.1. Computerised cognitive tests

I reviewed the statistical approaches of previous studies that have used the computerised cognitive tests in the Insight 46 battery and discussed the various models with Jennifer Nicholas (study statistician) in order to decide on the best approach for modelling the outcomes of each test. As each computerised task contains various different conditions, there are repeated measures on each participant, so models must account for the fact that repeated measures within a participant are likely to be correlated.

Some previous papers on the "What was where?" and Circle-tracing tasks chose to use repeated measures analysis of variance (ANOVA) where each participant was given a mean score for each condition and these mean scores were entered into the model (e.g. (Pertzov *et al.*, 2012, 2015; Vaportzis *et al.*, 2015b). Other analyses have used Generalised Estimating Equations (GEE), which are similar to standard regression but

enable repeated measures to be accounted for by specifying an appropriate correlation structure (see below), and using robust standard errors to account for the fact that the assumed correlation structure may not be entirely accurate (e.g. (Say *et al.*, 2011)). Disadvantages of the ANOVA approach are that information is lost by the reduction of data to mean scores, particularly information about within-participant variability, and it is heavily reliant on the assumption that the outcome is normally distributed. In contrast, GEE models allow analysis of the full trial-by-trial data (e.g. reaction times for every individual response) and are also more flexible, being able to accommodate different distributions of outcome variables, so I chose to adopt a GEE approach.

For continuous outcomes (e.g. reaction time), GEE linear models were used with an exchangeable correlation structure, which assumes that each response is equally likely to be correlated with each other response within-subject. For continuous outcomes with skewed distributions, appropriate transformations were first applied so that the data more closely approximated the normal distribution (e.g. log-transformation of reaction time data). For dichotomous outcomes (e.g. correct *vs.* incorrect response), GEE logistic models were used with an independent correlation structure. The GEE method accounts for the fact that the assumption of this correlation structure is likely to be violated (as repeated responses within-participant are likely to be correlated rather than totally independent) by estimating the within-participant correlations and using this to calculate adjusted estimates of the *p* values and confidence intervals.

3.5.3.2. Paper-and-pencil cognitive tests and summary scores

For analyses of the paper-and-pencil tests or the summary outcome variables from the computerised tests (see section 3.5.2.1), each participant had a single score, so repeated measures were not an issue. Where the distribution of these scores approximated to the normal distribution, linear regression was performed. Examination of residuals was performed to check model fits. For outcomes with skewed distributions, appropriate transformations were considered (e.g. log-transformation or square-root transformation). For outcomes where a transformation was not appropriate but skew was still a concern, bootstrapping was used to produce bias-corrected and accelerated 95% confidence intervals from 2000 replications. For outcomes that were derived from multiple binary responses (e.g. percentage of correct responses across a task where each response was either correct or incorrect), GEE logistic regression was performed as above.

3.5.3.3. Consideration of correction for multiple comparisons and multicollinearity of predictors

Although each model contained several predictors (e.g. sex, education, amyloid status), corrections for multiple comparisons were not applied. This was felt to be appropriate because my analyses were testing specific hypotheses about the effects of each predictor based on the literature, with the purpose of replicating and extending previous findings. The assumption underlying the practice of correcting for multiple comparisons is that the first explanation for non-null findings is chance, which may lead to errors of interpretation (Rothman, 1990). Correction for multiple comparisons is appropriate in scenarios where the result will be used to justify a decision with significant impact (e.g. in a clinical trial with multiple primary outcome measures where a drug may be licensed based on a positive effect on any one outcome measure) or when a large exploratory analysis is conducted without prior hypotheses (e.g. a genome-wide association study where thousands of genes are compared and it is statistically likely that numerous false positives would be detected). Where significant results are detected in this thesis, they are interpreted cautiously with reference to previous literature.

Models were not checked for multicollinearity through variance inflation factors, but this was not felt to be a concern as correlations between the predictors were modest at most. For example, Spearman's correlations between childhood cognitive ability, education and adult socioeconomic position are as follows: childhood cognitive ability and education $\rho = 0.50$, $p < 0.0001$; education and adult socioeconomic position $\rho = 0.51$, $p < 0.0001$; childhood cognitive ability and adult socioeconomic position $\rho = 0.29$, $p < 0.0001$.

3.6. Participant Characteristics

Forty-nine out of 502 participants met criteria for a neurological or psychiatric condition (see section 3.2.3), leaving 453 participants classified as cognitively-normal. Prevalence of each neurological or psychiatric condition is shown in Table 3-6.

Table 3-6. Numbers of Insight 46 participants with different neurological and psychiatric conditions

Disorder	N
Dementia	3
Parkinson's disease or other neurodegenerative disorder	5
Psychiatric disorder requiring anti-psychotic medication	2
Depression requiring electroconvulsive shock therapy	2
Epilepsy requiring active treatment	7
Traumatic brain injury or major neurosurgery	2
Multiple sclerosis	3
Cortical stroke	18
Brain malignancy	1
Mild cognitive impairment (MCI)	11

Note: although 49 participants met criteria for neurological or psychiatric conditions, the numbers in this table add up to 54 because some participants had more than one condition.

As detailed in Figure 3-1, 445 participants had complete biomarker data. Of these, 406 were classified as cognitively-normal. Of these, 18.3% were classified as Aβ+, which is around the expected prevalence for this age (Jansen *et al.*, 2015).

Table 3-7 shows the participant characteristics for the full sample (n = 502, used for the first section of analyses in Chapters 4 to 8) and the sub-sample of cognitively-normal participants with complete biomarker data, stratified by amyloid status (n = 406, used for the second section of analyses in Chapters 4 to 8).

Chi-square, t-tests and rank-sum tests were used as appropriate to test whether the Aβ+ and Aβ- groups differed in their characteristics. As expected, Aβ+ were more likely to be *APOE*-ε4 carriers (χ^2 = 41.6, p < 0.0001) but there was no evidence of statistically significant differences in any other characteristics (sex: χ^2 = 0.52, p = 0.47; age at assessment: t = 0.33, p = 0.74; education: z = 0.73, p = 0.47; childhood cognitive ability: t = -0.35, p = 0.72; adult socioeconomic position: z = -1.31, p = 0.19; WMHV: z = -0.68, p = 0.50, whole brain volume: t = -0.61, p = 0.54).

Table 3-7. Participant characteristics

	All participants	Cognitively-normal [a] participants with complete biomarker data	
		Aβ+	Aβ-
N	502	74	332
Sex: % female	49	46	51
Age at assessment: mean, *SD*, (range)	70.7, *0.68* (69.2 to 71.8)	70.6, *0.66* (69.4 to 71.8)	70.6, *0.70* (69.2 to 71.8)
Educational attainment: %			
None	15.5	17.6	15.4
Below O-levels (vocational)	5.2	6.8	4.2
O-levels or equivalent	24.9	25.7	26.2
A-levels or equivalent	35.7	32.4	35.2
Degree or equivalent	18.7	17.6	19.0
Childhood cognitive ability (z-score) [b]: mean, *SD*, (range)	0.39, *0.74* (-1.60 to 2.50)	0.44, *0.74* (-1.37, 2.50)	0.41, *0.74* (-1.59 to 2.47)
Adult SEP: %			
Unskilled	1.0	1.4	0.6
Partly skilled	4.8	2.7	5.4
Skilled manual	9.4	9.5	9.3
Skilled non-manual	21.3	16.2	22.0
Intermediate	52.2	55.4	51.8
Professional	11.4	14.9	10.8
SUVR: median, *IQR*, (range)	0.55, *0.51 to 0.58* (0.45 to 0.87) [c]	0.67, *0.64 to 0.71* (0.61 to 0.87)	0.53, *0.51 to 0.56* (0.47 to 0.61)
WMHV (ml): median, *IQR*, (range)	3.1, *1.6 to 6.8* (0.3 to 33.7) [d]	3.3, *1.8 to 6.8* (0.3 to 33.7)	2.9, *1.5 to 6.4* (0.3 to 32.8)
Whole brain volume (ml): mean, *SD*, (range)	1100, *99* (819 to 1494) [e]	1118, *103* (819 to 1326)	1098, *97* (860 to 1494)
***APOE* genotype**: % [f g]			
ε4-carriers	29.6	60.8	22.9
non-carriers	70.4	39.1	77.1

[a] Defined as the absence of major neurological or psychiatric conditions (see section 3.2.3); [b] Z-scores for childhood cognitive ability were based on the full NSHD cohort of N=5362, so the mean for Insight 46 participants indicates that they had higher childhood cognitive ability on average than their peers not recruited to this sub-study. [c] n=462 due to missing data; [d] n=455 due to missing data; [e] n=468 due to missing data; [f] n=500 due to missing data. [g] The full breakdown of APOE genotypes was as follows: 14% ε2/ε3, 2% ε2/ε4, 56% ε3/ε3, 25% ε3/ε4, 3% ε4/ε4.

Aβ = β-amyloid; IQR = interquartile range; SD = standard deviation; SEP = socioeconomic position; SUVR = standard uptake volume ratio. WMHV = white matter hyperintensity volume.

4. STANDARD COGNITIVE TESTS

This chapter focuses on the paper-and-pencil cognitive tests within the Insight 46 cognitive battery. A paper based on this chapter has been accepted for publication in *Neurology* (Lu *et al.*, 2019).

4.1. Introduction

Emerging evidence that subtle cognitive decline begins in the preclinical phase of AD has led to efforts to develop sensitive cognitive measures to detect and track this decline. One such measure is the Preclinical Alzheimer Cognitive Composite (PACC), composed of four cognitive tests which measure global cognition, episodic memory and executive function (Donohue *et al.*, 2014). It is the primary outcome measure in the A4 trial (Donohue *et al.*, 2014), the first clinical trial in Aβ+ cognitively-normal older adults. Several variations of the PACC have been tested (Sperling *et al.*, 2014; Lim, Snyder, *et al.*, 2016; Donohue, Sperling, *et al.*, 2017; Donohue, Sun, *et al.*, 2017; Mormino *et al.*, 2017; Papp *et al.*, 2017; Merluzzi *et al.*, 2019) and a revised version is under development (Hassenstab *et al.*, 2017). The standard cognitive tests included in the Insight 46 battery (see section 3.2.1) were chosen to allow computation of the PACC, complemented by a test of non-verbal reasoning (WASI Matrix Reasoning) that was chosen for its similarity to aspects of the cognitive tests completed in childhood.

Cognitively-normal Aβ+ older adults have shown faster decline on the PACC than Aβ- individuals, over intervals of around 3 years (Donohue, Sperling, *et al.*, 2017; Mormino *et al.*, 2017). However there are mixed results when comparing Aβ groups at baseline, with some studies finding lower PACC scores in Aβ+ participants (Donohue, Sperling, *et al.*, 2017) but some finding no difference (Donohue *et al.*, 2014; Burnham *et al.*, 2016; Soldan *et al.*, 2016; Buckley *et al.*, 2017; Mormino *et al.*, 2017; Rabin *et al.*, 2018). One reason for this could be that cross-sectional analyses struggle to account for the wide variation that exists between individuals in terms of their overall cognitive abilities (beyond adjusting for age, sex and educational attainment), whereas longitudinal analyses account for this variation by effectively using each participant as their own control. The better we can account for predictable variation between individuals at baseline, the greater the likelihood that outcomes such as the PACC will be sensitive enough to detect differences in cognition associated with preclinical disease pathology, if such differences exist.

Insight 46 is uniquely placed to address this, since data are available on participants' cognition since childhood. While it is well-established that higher educational attainment is associated with better performance on most cognitive tests in older age (Weintraub *et al.*, 2009; Shirk *et al.*, 2011; Gaertner *et al.*, 2018; Jansen *et al.*, 2018), the life-course determinants of performance on the PACC are unknown, and the effects of childhood cognitive ability and adult socioeconomic position on PACC score have never been investigated before.

There is also a need to understand more about possible sex differences on the PACC. As discussed in section 2.4.1, women are at greater risk of AD (although this may be explained by greater longevity) but the relevance of sex differences to the detection of subtle cognitive decline in preclinical AD is unclear. There is consistent evidence that women tend to perform slightly better than men on the Digit-Symbol Substitution test of processing speed included within the PACC (Royer, 1978; Roivainen, 2011; Gaertner *et al.*, 2018) as well as on tests of verbal memory (Andreano and Cahill, 2009). Normative data for over 3000 cognitively-normal older adults reported an advantage for women on three of the four sub-tests included within the PACC (Digit-Symbol Substitution, Mini Mental State Examination, and Logical Memory story recall – see section 4.2.1 below for description of sub-tests), although the authors concluded that the differences are small and may not have clinical relevance (Weintraub *et al.*, 2009).

Consistent with the approach outlined in section 3.5, I aimed first to characterise the performance of Insight 46 participants on the paper-and-pencil cognitive tests including the PACC with respect to sex, childhood cognitive ability, education and adult socioeconomic position. I then explored whether cognitive performance was influenced by amyloid status, whole brain volume, white matter hyperintensity volume (WMHV), and genetic risk for AD (*APOE*-ε4).

4.2. Methods

4.2.1. Stimuli and Procedure

The original PACC (Donohue *et al.*, 2014) is composed of four cognitive tests – the Mini Mental State Examination (MMSE) (Folstein, Folstein and McHugh, 1975), Logical Memory IIa from the Wechsler Memory Scale Logical Memory (Wechsler, 1987), Digit-Symbol Substitution test from the Wechsler Adult Intelligence Scale-Revised (Wechsler, 1981), and the Free and Cued Selective Reminding Test (FCSRT) (Grober *et al.*, 2008).

We replaced the FCSRT with the 12-item Face-Name test (FNAME-12) (Papp *et al.*, 2014), to avoid potential overlap with a similar word-learning memory test administered to the NSHD cohort at multiple time-points throughout adulthood (Rawle *et al.*, 2018). FNAME-12 is similar to FCSRT in terms of being an episodic memory test of immediate and delayed recall, is moderately correlated with FCSRT free recall scores (Papp *et al.*, 2014) and is also relatively challenging for cognitively-normal populations. Two previous studies have reported that FNAME is sensitive to Aβ deposition (Rentz *et al.*, 2011; Sanabria *et al.*, 2018).

Participants also completed the Matrix Reasoning test from the Wechsler Abbreviated Scale of Intelligence (Wechsler, 1999) – a measure that was chosen for its similarity to aspects of the cognitive tests completed in childhood.

The **MMSE** is a 30-point composite screening tool for cognitive impairment which is widely used within clinical practice.

Digit-Symbol Substitution is an index of executive function and psychomotor speed. The score is the number of items completed correctly within 90 seconds.

Logical Memory IIa assesses free recall of a short story, which the participant is asked to recall immediately and after a delay of approximately 20 minutes. The exact delay duration is recorded so that it can be accounted for in analyses, as it may affect performance (Montgomery *et al.*, 2017).

The **Face-Name test (FNAME-12)** assesses associative memory for face-name and face-occupation pairs. Two versions exist: FNAME-12A and FNAME12-B. This study used FNAME-12A. Participants are shown 12 unfamiliar face-name and face-occupation pairs (e.g. “Sarah, Reporter”), with 8 seconds to study each one. They are then presented with each face and asked to recall the associated name and occupation. This process is repeated with a second learning phase and a second recall test. After a ~10-minute delay they are again shown each face and asked to recall the names and occupations (the third recall test). After a ~30-minute delay participants are shown 12 sets of three faces and asked to identify each previously learned face from the two distractors (facial recognition) and to recall the name and occupation (the fourth recall test). If they cannot recall the name and/or occupation, they are asked to select the correct answer from three options comprising: the correct answer, a distractor (a name/occupation that belongs with a different face in the set), and a name/occupation that did not feature in the set. The summary outcomes are FN-N (total names recalled, max. 48), FN-O (total occupations recalled, max. 48) and FNAME-total (FN-N + FN-O, max. 96) – these outcomes are based on the four recall tests. Precise administration

times were recorded for a sample of 50 participants to check that the delay times conformed to expectations: the mean delay times were 10.0 mins and 35.5 mins.

The **Matrix Reasoning** test assesses non-verbal reasoning, an aspect of fluid intelligence. Participants are shown a matrix of geometric shapes and are required to select the missing piece from five options. There are 32 matrices, graded in difficulty, and the test is discontinued when participants make four consecutive errors (or four errors within five consecutive items), as specified in the manual (Wechsler, 1999).

The four components of our version of the PACC were: MMSE total score, Logical Memory delayed recall score, Digit-Symbol Substitution score and FNAME-total. Following the method described in previous studies (Lim, Snyder, *et al.*, 2016; Buckley *et al.*, 2017; Mormino *et al.*, 2017; Papp *et al.*, 2017; Rabin *et al.*, 2018), the four components were converted into z-scores based on the full Insight 46 sample, and then averaged. A higher PACC score indicates better performance. Two participants did not complete the FNAME test and one did not complete the Digit-Symbol Substitution test (see section 3.3). For these three participants, their PACC score was the average of the z-scores for the three tests they completed. This is consistent with a previous study which required at least 2 out of the 4 components to be present (Soldan *et al.*, 2016). Excluding these three people did not change any of the results.

4.2.2. Participants

Out of 502 participants, 499 completed all cognitive tests and three were missing one test score as mentioned above (see section 3.3 for reasons for missing data). Participant characteristics are reported in section 3.6.

4.3. Patterns and predictors of performance

4.3.1. Statistical Analyses

To investigate the relationship between cognitive outcomes and demographic factors, all participants were included (n=502), as explained in section 3.5. Raw scores from each cognitive test were standardised to *z*-scores based on the full Insight 46 sample to allow

comparison of effect sizes across different cognitive tests. Multivariable linear regression models were run where the outcome was the z-score on a particular cognitive test and the predictors were sex, age at assessment, childhood cognitive ability, education, adult socioeconomic position and presence of a neurological or psychiatric condition (including MCI). For outcomes with skewed distributions (MMSE and Matrix Reasoning), bootstrapping was used to produce bias-corrected and accelerated 95% CIs from 2000 replications. For the Logical Memory delayed recall score, the model contained an additional factor of delay duration (time elapsed between the immediate and delayed recall). Mean delay duration was 24.6 minutes (*SD* = 4.66) and there was no evidence that this was associated with performance (regression coefficient = -0.006, *95% CIs* -0.024 to 0.012, *p* = 0.53), but it was included in the models as per standard practice.

4.3.2. Results

Descriptive statistics for each test are given in Table 4-1.

On average, Insight 46 participants performed at the expected level for their age on the MMSE, Digit Symbol and Logical Memory tests, according to normative data (Shirk *et al.*, 2011). On the Matrix Reasoning test, their performance (mean = 24) was above the expected level based on normative data (sample mean for 70- to 74-year-olds is 16 (Wechsler, 1999)) but comparable to a sample of healthy older adults recruited by Washington University (mean = 24 (Emery, Hale and Myerson, 2008)). To date, only two studies have published FNAME-12 data (Papp *et al.*, 2014; Kormas *et al.*, 2018) and Insight 46 means are higher than these.

Table 4-1. Descriptive statistics for the standard cognitive tests

	mean	*SD*	median	range
MMSE (max. 30)	29.3	1.0	30	22, 30
Matrix Reasoning (max. 32)	24.0	4.9	25	4, 32
Digit-Symbol [a] (max. 93)	47.6	10.4	48	19, 82
Logical Memory: Immediate (max. 25)	12.8	3.5	13	4, 22
Delayed (max. 25)	11.5	3.7	12	0, 23
FNAME-12 [b]**: FN-N** (max. 48)	27.0	11.7	28	0, 47
FN-O (max. 48)	38.2	8.0	40	1, 48
Total (max. 96)	65.3	18.3	67	3, 95
Facial Recognition (max. 12)	12.0	0.2	12	9, 12
Names recognition (max. 12)	10.3	1.8	11	3, 12
Occupations recognition (max. 12)	11.6	0.9	12	4, 12
PACC (mean of z-scores)	-0.00	0.73	0.07	-3.49, 1.72

[a] n=501 due to missing data. [b] n=500 due to missing data.

FN-N = names recalled; FN-O = occupations recalled; PACC = Preclinical Alzheimer Cognitive Composite

Results of the multivariable regression models exploring associations with demographic and life-course predictors are reported in Table 4-2. On average, participants with neurological or psychiatric conditions (including MCI) scored significantly lower on all tests (Table 4-2). The analyses were rerun excluding the participants with MCI to check that these differences could not be explained by circularity in the definition of MCI (since low scores on the Logical Memory or Digit-Symbol Substitution tests formed part of the MCI criteria); the results were unchanged except that the differences were no longer statistically significant on MMSE and Matrix Reasoning.

Females scored significantly higher than males on all measures except Matrix Reasoning (Table 4-2); the greatest difference was on the FNAME-12, particularly in recalling names.

As expected across this narrow age range (2.6 years – reflecting the time it took to collect the data, since participants were all born in the same week) there was no evidence of age effects on cognition, except on the Matrix Reasoning test where older age was associated with slightly poorer performance (Table 4-2).

Higher childhood cognitive ability was associated with better performance on every cognitive outcome (Table 4-2, Figure 4-1). Higher educational attainment and higher adult socioeconomic position were independently positively associated with the majority of cognitive outcome measures, including the PACC. Higher educational attainment showed a notable positive association with the Matrix Reasoning task.

All these effects were maintained when excluding participants with neurological or psychiatric conditions, except that the following two associations were directionally but no longer statistically significant: Logical Memory Delayed and adult socioeconomic position (p=0.073); FNAME FN-O and education (p=0.12).

Table 4-2. Associations between demographic and life-course predictors and performance on the standard cognitive tests (n = 502)

Cognitive test	Coefficients and 95% confidence intervals for each predictor variable						R^2
	Sex (female as reference)	**Age at assessment** (per year)	**Education** (per category) [a]	**Adult SEP** (per category) [a]	**Childhood cognitive ability** (per z-score)	**Neurological or psychiatric condition** [b]	
MMSE	**-0.19** -0.35, -0.01	-0.11 -0.25, 0.01	**0.15*** 0.07, 0.24	0.02 -0.08, 0.11	**0.17** 0.05, 0.31	**-0.50** -1.15, -0.12	0.12
Matrix Reasoning	-0.14 -0.31, 0.03	**-0.17*** -0.28, -0.05	**0.23*** 0.15, 0.31	**0.12*** 0.03, 0.21	**0.14*** 0.03, 0.26	**-0.32** -0.65, -0.05	0.21
Digit-Symbol [c]	**-0.35*** -0.51, -0.19	-0.07 -0.19, 0.04	**0.16*** 0.09, 0.24	0.05 -0.05, 0.13	**0.24*** 0.11, 0.36	**-0.70*** -0.97, -0.43	0.20
Logical Memory Immediate	**-0.42*** -0.59, -0.26	-0.01 -0.13, 0.11	**0.09** 0.01, 0.17	0.06 -0.03, 0.15	**0.29*** 0.16, 0.41	**-0.40** -0.67, -0.12	0.16
Logical Memory Delayed	**-0.47*** -0.64, -0.31	0.02 -0.09, 0.14	0.05 -0.03, 0.13	**0.09** 0.00, 0.18	**0.31*** 0.18, 0.44	**-0.59*** -0.86, -0.32	0.19
FNAME-12 FN-N [d]	**-0.55*** -0.71, -0.40	-0.01 -0.12, 0.10	**0.08** 0.00, 0.15	**0.12*** 0.04, 0.21	**0.35*** 0.23, 0.47	**-0.47*** -0.74, -0.21	0.24
FNAME-12 FN-O [d]	**-0.42*** -0.58 -0.26	-0.08 -0.19, 0.03	**0.09** 0.01, 0.16	**0.19*** 0.10, 0.28	**0.27*** 0.15, 0.40	**-0.52*** -0.79, -0.26	0.22
FNAME-12 Total [d]	**-0.54*** -0.69, -0.398	-0.04 -0.15, 0.07	**0.09** 0.01, 0.16	**0.16*** 0.08, 0.25	**0.34*** 0.22, 0.46	**-0.53*** -0.79, -0.27	0.26
PACC	**-0.39*** -0.50, -0.28	-0.05 -0.13, 0.02	**0.11*** 0.06, 0.16	**0.08*** 0.02, 0.14	**0.26*** 0.18, 0.35	**-0.60*** -0.78, -0.43	0.34

Units are in z-scores. Coefficients in bold are significant at p < 0.05 and asterisks indicate significance at p < 0.01. Multivariable regression models were used so each association is independent of all others. R-squared gives the proportion of variance in each cognitive outcome that is explained by the combined predictors. Logical Memory Delayed score was additionally adjusted for time elapsed between the immediate and delayed recall. [a] see section 3.2.4 for definitions of categories; [b] cognitively-normal as reference category (see section 3.2.3 for definitions). [c] n=501 due to missing data. [d] n=500 due to missing data. FN-N = names recalled. FN-O = occupations recalled. MMSE = Mini Mental State Examination. PACC = Preclinical Alzheimer's Cognitive Composite. SEP = socioeconomic position

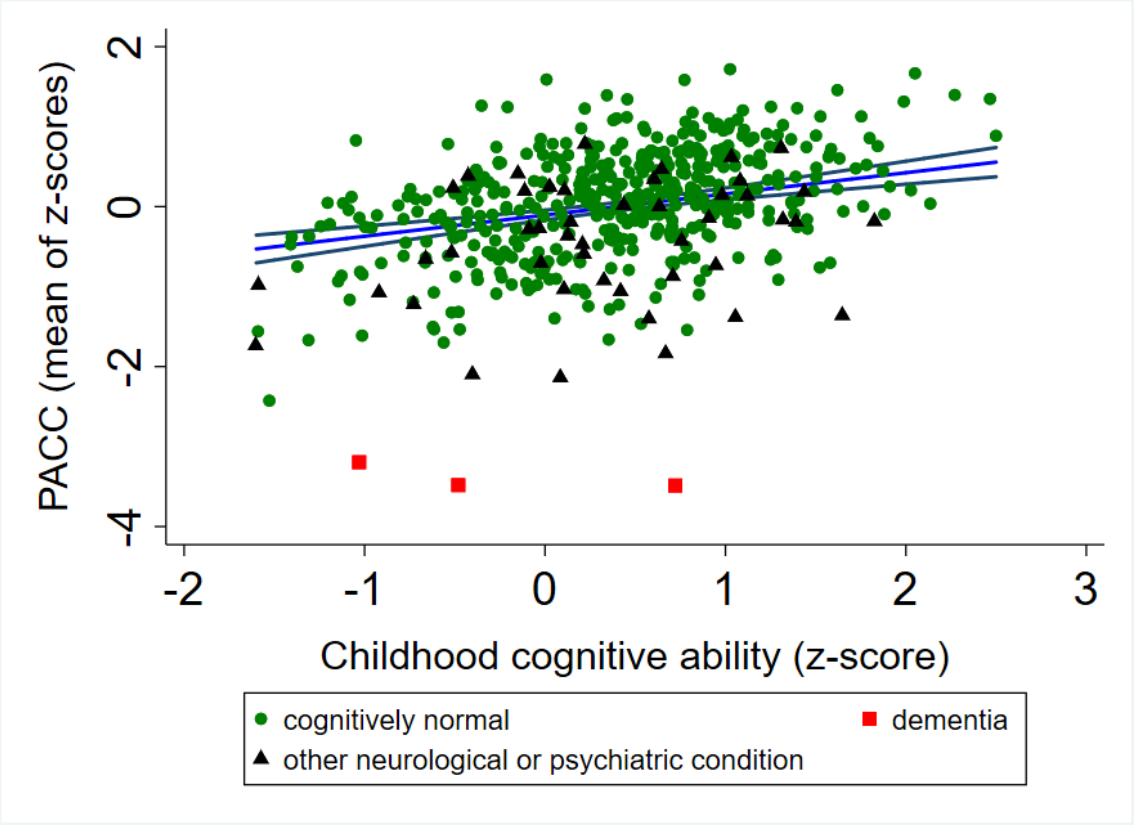


Figure 4-1. Preclinical Alzheimer Cognitive Composite (PACC) score against childhood cognitive ability

Scatter plot shows the raw data, colour-coded by clinical group. Alzheimer's disease dementia is distinguished from other neurological or psychiatric conditions for interest. The blue line is the line of best fit from the multivariable regression model (adjusted for sex, age at assessment, education, adult socioeconomic position and presence of neurological or psychiatric conditions), and the navy lines are its 95% confidence intervals. For an explanation of the childhood cognitive ability variable, see section 3.2.4.

4.4. Associations with biomarkers and APOE-ε4

4.4.1. Statistical Analyses

Following the format laid out in 3.5, the second part of this chapter aims to investigate associations between performance on the standard cognitive tests and biomarkers of AD in cognitively-normal participants for whom complete biomarker data are available. The number of participants meeting these criteria was 406 (see section 3.3).

The z-score on a particular cognitive test was the outcome and amyloid status, whole brain volume, WMHV and *APOE* genotype were included as predictors in multivariable regression models to examine the effects of each biomarker adjusted for all the others.

To adjust for the correlation between whole brain volume and head size, total intracranial volume was included in all models, as were the demographic and life-course factors investigated in the first analysis (sex, age at assessment, childhood cognitive ability, education and adult socioeconomic position). Interactions were investigated between amyloid status and brain volume, and amyloid status and WMHV.

The models were additionally rerun replacing dichotomised amyloid status with a continuous measure of Aβ (SUVR) to test whether increasing Aβ deposition was associated with differences in performance. To check whether associations between SUVR and cognition were sensitive to the inclusion of the imputed SUVR values (see section 3.2.2), the analyses were rerun excluding the 26 participants with imputed data.

4.4.2. Results

Results of the multivariable regression models are reported in Table 4-3. Results for the demographic and life-course factors (sex, age at assessment, childhood cognitive ability, education and adult socioeconomic position) are not reported as they are essentially unchanged from the first analysis section (4.3.2).

On average, Aβ+ participants scored lower than Aβ- participants on every cognitive measure (Table 4-3, Figure 4-2). The unadjusted differences were only statistically significant for the MMSE and Matrix Reasoning (Figure 4-2A), but in the multivariable model adjusting for demographic, life-course and biomarker factors, the differences were also statistically significant for Logical Memory immediate recall and PACC (Figure 4-2B, Table 4-3).

Replacing dichotomised amyloid status with the continuous SUVR revealed weak associations between higher SUVR and poorer performance on MMSE (regression coefficient -1.21, *95% CIs* -2.39 to -0.10), Logical Memory immediate recall (regression coefficient -1.87, *95% CIs* -3.22 to -0.52) and PACC (regression coefficient -1.09, *95% CIs* -1.90 to -0.29). Similar trends were observed on the other tests but did not reach statistical significance. These results were unchanged when the analyses were rerun excluding the 26 participants with imputed SUVR data.

The only outcome which showed an association with whole brain volume was Digit-Symbol Substitution, where larger whole brain volume was associated with better performance (Table 4-3).

Digit-Symbol Substitution and PACC showed associations with WMHV, where higher WMHV was associated with poorer performance (Table 4-3).

On average, *APOE*-ε4 carriers performed better than non-carriers on Logical Memory immediate recall (after adjustment for the detrimental effect of Aβ and all other confounders). There was a trend in the same direction on Logical Memory Delayed recall ($p = 0.06$) (Table 4-3).

There was no evidence of interactions between amyloid status and whole brain volume, or amyloid status and WMHV.

Table 4-3. Associations between biomarkers and performance on the standard cognitive tests in cognitively-normal participants (n=406)

Cognitive Test	Coefficients and 95% confidence intervals for each predictor variable				R^2
	Amyloid status (Aβ- as reference)	**WMHV** (per 10 cm^3)	**Whole Brain Volume** (per 10 cm^3)	***APOE*-ε4** (non-carrier as reference)	
MMSE	**-0.24** -0.46, -0.05	-0.00 -0.21, 0.14	-0.01 -0.03, 0.01	0.10 -0.06, 0.29	0.14
Digit-Symbol	-0.17 -0.41, 0.06	**-0.21*** -0.37, -0.06	**0.05*** 0.03, 0.07	-0.01 -0.21, 0.19	0.21
Logical Memory Immediate	**-0.30** -0.54, -0.05	-0.10 -0.26, 0.06	-0.01 -0.03, 0.01	**0.22** 0.02, 0.43	0.16
Logical Memory Delayed	-0.19 -0.43, 0.05	-0.12 -0.28, 0.04	-0.01 -0.03, 0.01	0.19 -0.01, 0.40	0.17
FNAME FN-N [a]	-0.05 -0.28, 0.19	-0.06 -0.22, 0.09	0.01 -0.01, 0.03	-0.06 -0.25, 0.14	0.21
FNAME FN-O [a]	-0.16 -0.38, 0.07	-0.10 -0.25, 0.05	0.00 -0.01, 0.02	0.13 -0.06, 0.32	0.21
FNAME-total [a]	-0.10 -0.33, 0.13	-0.09 -0.24, 0.07	0.01 -0.01, 0.02	0.02 -0.17, 0.21	0.23
PACC	**-0.17** -0.32, -0.03	**-0.10** -0.20, -0.01	0.01 -0.01, 0.02	0.08 -0.05, 0.20	0.34
Matrix Reasoning	**-0.40*** -0.69, -0.15	-0.02 -0.19, 0.12	0.01 -0.01, 0.03	0.13 -0.09, 0.35	0.23

Multivariable regression models were used so each association is independent of all others. In addition to the predictors listed, models also included total intracranial volume, sex, age at assessment, childhood cognitive ability, education and adult socioeconomic position. Logical Memory Delayed score was additionally adjusted for time elapsed between the immediate and delayed recall.

All units are in z-scores. Coefficients in bold are significant at $p < 0.05$ and asterisks indicate significance at $p < 0.01$. R-squared gives the proportion of variance in each cognitive outcome that is explained by the combined predictors. [a] N=405 due to missing data. Aβ- = β-amyloid negative; FN-N = names recalled; FN-O = occupations recalled; MMSE = Mini Mental State Examination; PACC = Preclinical Alzheimer Cognitive Composite; WMHV = white matter hyperintensity volume.

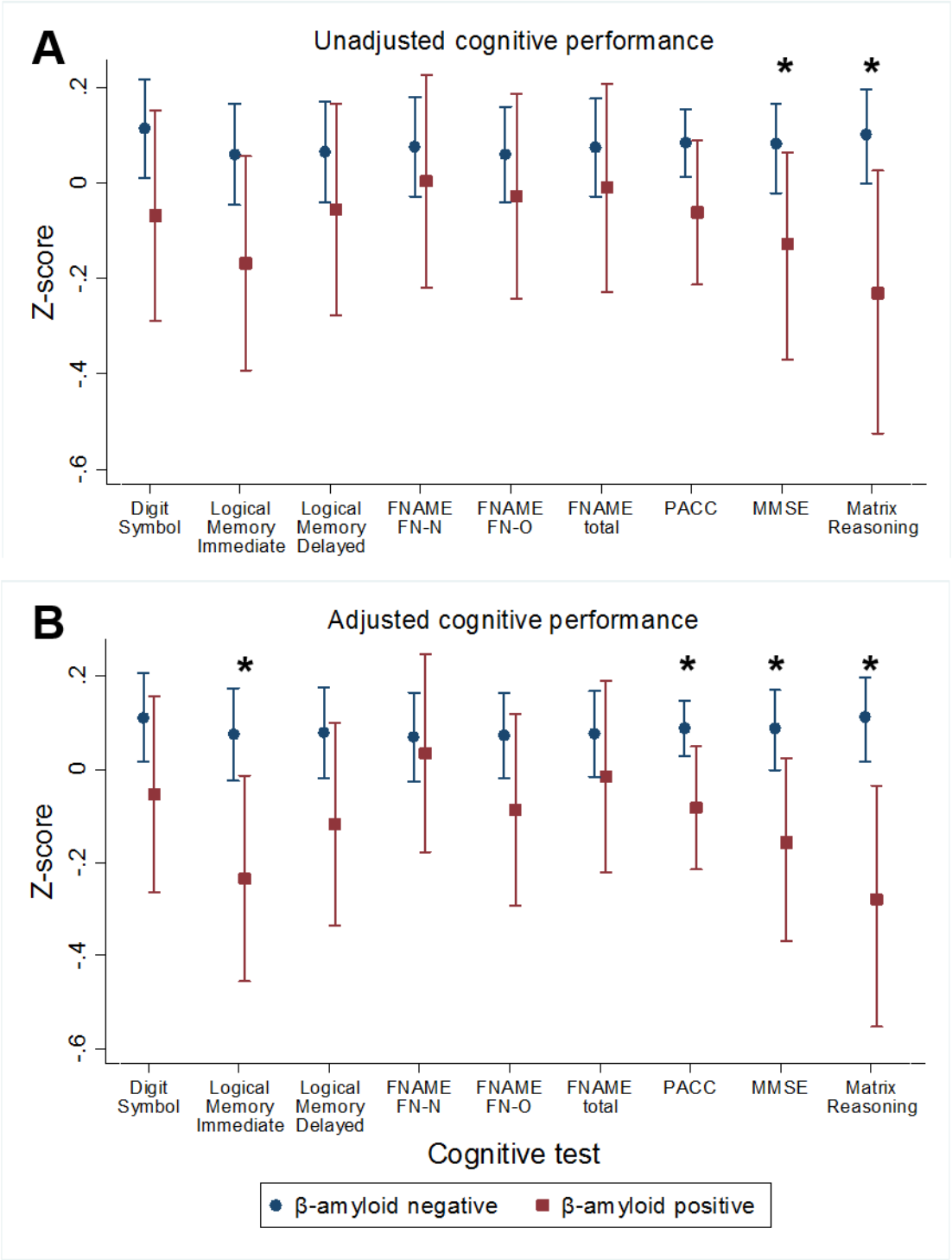


Figure 4-2. Performance of β-amyloid positive and β-amyloid negative individuals on the standard cognitive tests: means and 95% confidence intervals

A = unadjusted means; B = adjusted means predicted from the multivariable regression models (adjusted for age at assessment, sex, childhood cognitive ability, education, adult socioeconomic position, whole brain volume, total intracranial volume, white matter hyperintensity volume and APOE genotype). Asterisks indicate statistically significant differences ($p < 0.05$).

4.5. Discussion

4.5.1. Summary

In this large population-based sample of older adults of approximately the same age, I investigated predictors of performance on a range of cognitive measures including the Preclinical Alzheimer Cognitive Composite (PACC). The key findings are that childhood cognitive ability was strongly associated with all cognitive scores, significant sex differences in cognition were observed, and Aβ positivity and white matter hyperintensity volumes (WMHV) were associated with lower PACC scores among cognitively-normal participants. These results are discussed in the following sub-sections. For a discussion of strengths and limitations that apply to all the analyses presented in this thesis, such as considerations relating to the generalisability of the sample, see Chapter 10.

4.5.2. Demographic and life-course predictors

4.5.2.1. Associations with childhood cognitive ability, education and adult socioeconomic position

Childhood cognitive ability was consistently an important predictor with a notable effect on every cognitive outcome. My finding that educational attainment and adult socioeconomic position were associated with many cognitive outcomes, *independent of childhood cognition,* is consistent with previous NSHD analyses which have shown that these factors are only moderately correlated and all have direct and indirect influences on cognition across the life-course ((Richards and Sacker, 2003; Richards *et al.*, 2019) – see section 2.4.2). It is also consistent with evidence that education and occupational attainment may have protective effects on later life cognition (Stern, 2012). This is discussed in more detail in Chapter 10, as similar effects were found on other cognitive tests that feature in later chapters.

4.5.2.2. Sex differences

The finding of an advantage for women on the sub-tests of the PACC is consistent with previous studies (Weintraub *et al.*, 2009; Shirk *et al.*, 2011; Gaertner *et al.*, 2018; Jansen *et al.*, 2018). The effect size of sex on PACC score was large enough to be potentially clinically meaningful (0.4 *SD*), suggesting that accounting for sex differences on the

PACC may be important. As the FNAME test was the component where females had the greatest advantage, versions of the PACC which include a different memory test may be less susceptible to sex differences. However, sex differences have also been reported on the FCSRT (the test used in the original PACC, which we replaced with FNAME) (Mura *et al.*, 2017). Sex differences are discussed further in Chapter 10.

4.5.2.3. Associations with age

The interpretation of the association observed between older age at assessment and lower Matrix Reasoning score is unclear. While scores on this test are known to decline with age (Wechsler, 1999; Bugg *et al.*, 2006; Emery, Hale and Myerson, 2008), the effect size of the association in Insight 46 (-0.17 z-score units) equates to -0.83 points on the test per year, which is incompatible with the much lower rate of decline across adulthood reported by others (Wechsler, 1999; Emery, Hale and Myerson, 2008). I considered the possibility that associations between age and cognition in Insight 46 could be explained by recruitment bias – this is discussed in Chapter 10.

4.5.3. Associations with biomarkers and APOE-ε4

In cognitively-normal participants (i.e. excluding those who fulfilled dementia or MCI criteria and those with another neurological or psychiatric condition), Aβ positivity was associated with poorer performance on the PACC. Statistically significant differences were also observed on several individual tests assessing a range of cognitive domains: memory (Logical Memory Immediate), non-verbal reasoning (Matrix Reasoning), and a global measure of cognitive state (MMSE). Previous studies have tended not to find a difference in PACC score between Aβ+ and Aβ- individuals at baseline, with differences only emerging after longitudinal assessment (Donohue *et al.*, 2014; Burnham *et al.*, 2016; Soldan *et al.*, 2016; Buckley *et al.*, 2017; Mormino *et al.*, 2017; Rabin *et al.*, 2018), although a difference at baseline has been reported before in a sample where the Aβ+ group were slightly older and less educated than the Aβ- group (Donohue, Sperling, *et al.*, 2017). These results add to accumulating evidence for subtle cognitive differences associated with Aβ deposition, even at an age when those who are destined to develop dementia are still likely to be many years from symptoms (Prince *et al.*, 2014).

In cognitively-normal participants with a generally low burden of white matter disease, I also found an independent association between WMHV and PACC score which, to my knowledge, has not been reported before. This suggests that the PACC may be a

sensitive, rather than a specific, marker of cerebral pathology – an important consideration for clinical trials.

Controlling for childhood cognitive ability, education and adult socioeconomic position, as well as other brain pathologies and *APOE* genotype, enabled detection of a difference in PACC score between Aβ+ and Aβ- participants, whereas the unadjusted group difference was not statistically significant. This may be partially explained by negative confounding effects, whereby one or more factors that predicted higher PACC score also had weak positive associations with Aβ positivity. This was indeed the case for childhood cognitive ability and adult socioeconomic position, which were slightly higher in Aβ+ individuals (although differences were not statistically significant). Such differences may well be due to chance but can suppress the association between Aβ and cognition when not adjusted for. Another factor may have been that adjustment for these variables reduced the unexplained residual variance in PACC score, thus increasing the ability to detect smaller differences between the groups. Combined together, the demographic, life-course and biomarker factors accounted for one third of the variance in PACC score among cognitively-normal participants.

Fluid intelligence measures themselves are not usually considered candidates for detecting subtle cognitive decline in preclinical AD, so my finding that Aβ positivity was associated with poorer performance on the Matrix Reasoning test, to a greater degree than the PACC, (accounting for childhood cognitive ability etc.) is interesting. It is consistent with evidence that non-verbal IQ declines early in presymptomatic FAD mutation carriers (Fox *et al.*, 1998). As a high-level test involving multiple domains (including visuoperceptual, working memory, and executive function), Matrix Reasoning is rather different to the tests comprising the PACC, and its potential as a marker of cognitive decline merits further investigation.

The Digit-Symbol test was the single test most sensitive to overall brain health, showing associations with WMHV and whole brain volume in cognitively-normal participants, and being the task on which participants with neurological or psychiatric conditions were most disadvantaged. Negative effects of WMHV on processing speed are well established, consistent with subcortical damage (Gunning-Dixon and Raz, 2000; Oosterman *et al.*, 2004; Prins *et al.*, 2005; van Dijk *et al.*, 2008; Prins and Scheltens, 2015). The Digit-Symbol task may be particularly sensitive to brain pathologies because good performance on this task requires multiple cognitive functions, including visuomotor skills, executive functioning, working memory and attention, hence people with an impairment in any one of these areas might perform badly (Jaeger, 2018). The fact that the Digit-Symbol task is timed may also contribute to its sensitivity at detecting small differences in performance.

The finding that *APOE*-ε4 carriers showed evidence of better short-term memory on the Logical Memory immediate recall test – a measure on which Aβ positivity was associated with poorer performance – is consistent with the model of antagonistic pleiotropy discussed in section 2.6.1, whereby *APOE*-ε4 may have both beneficial and detrimental effects. This is discussed in greater detail in Chapter 10.

4.5.4. Comments on the Face-Name Associative Memory Exam (FNAME-12)

Few studies have published results of the FNAME test. Two previous studies found a sex difference on FNAME which was reduced in older adults (Alegret *et al.*, 2015) and attenuated in postmenopausal women (Rentz *et al.*, 2017). Here, I found a significant sex difference in 70-year-olds. It has been argued that one potential benefit of the FNAME test is, in contrast to many other memory tests, its reported lack of association with education (Kormas *et al.*, 2018), although this has been contradicted in one study (Papp *et al.*, 2014). In the current study which benefits from prospective collection over the life course, I found that childhood cognitive ability, education and adult socioeconomic position were all significant predictors of FNAME scores.

A recent consensus statement on recommended outcomes in preclinical AD from the European Prevention of Alzheimer's Dementia project (EPAD) raised concerns about a ceiling effect on this task in healthy populations, recommending the use of the "Favorites test", a refinement of FNAME designed to reduce the ceiling effect by pairing faces with unrelated words (Ritchie *et al.*, 2017). These results from Insight 46 suggest that FNAME-12 is sufficiently challenging for 70-year-olds, despite scores on the occupations sub-scale being somewhat skewed towards the top end. Indeed, for some participants who particularly struggled with recalling names and required encouragement, the 16-item FNAME might have been stressful.

Two previous studies reported an association between Aβ deposition and FNAME performance, specifically on the FN-N outcome (recall for names) (Rentz *et al.*, 2011; Sanabria *et al.*, 2018). While my results followed this trend, differences between Aβ+ and Aβ- participants did not reach statistical significance, and the FN-N outcome did not appear more sensitive than FN-O (recall for occupations) or FNAME-total.

4.5.5. Conclusions

In summary, these data show that childhood cognitive ability, education and adult socioeconomic position all independently influence cognitive performance at age 70, which has implications both for the interpretation and analysis of cognitive data measured in later life. These results provide evidence that the PACC can be used to detect subtle cross-sectional differences in cognition associated with Aβ deposition and white matter disease in cognitively-normal older adults at an age where dementia prevalence is very low.

5. "WHAT WAS WHERE?" VISUAL SHORT-TERM MEMORY BINDING

5.1. Introduction

Visual memory binding describes the ability to integrate multiple features of an object in memory, such as colour and shape, or object and location. This vital aspect of both short- and long-term memory is reported to be impaired in AD at an early stage (Parra *et al.*, 2009, 2010, 2011, 2017; Fernández *et al.*, 2018). Furthermore, evidence from studies of presymptomatic carriers of mutations causing familial Alzheimer's disease (FAD) suggests that subtle deficits in visual memory binding may be one of the earliest detectable changes in cognition in the preclinical stage of AD (Parra *et al.*, 2010, 2011, 2015; Liang *et al.*, 2016). A recent review concluded that although memory generally declines with age, a specific deficit in binding is not seen in healthy ageing once this general decline is accounted for (Schneegans and Bays, 2019), which strengthens the case for investigating the potential of memory binding tasks as specific markers for early change associated with AD pathology.

The study by Liang *et al.* (2016) used the "What was where?" task, a computerised visual short-term memory binding task that requires participants to recall the identity and location of objects (Pertzov *et al.*, 2012, 2013, 2015). One notable feature of this task is that it incorporates memory for locations as a continuous analogue measure (the distance between the location reported by the participant and the true location), in contrast to most memory measures which are made up of binary responses (correct vs. incorrect). This approach is based on a recognition that failure to recall an item correctly does not mean that its representation in memory has been entirely lost, and this may make the localisation measure particularly sensitive to small differences in performance (Pertzov *et al.*, 2013; Ma, Husain and Bays, 2014). In the "What was where?" task, a 'binding error' is captured when a participant correctly recalls the identity of an object but mislocalises it to the location of a different object held in memory. Liang *et al.* (2016) reported that presymptomatic FAD mutation carriers tended to make more binding errors than controls, despite having unimpaired memory for objects and locations.

While it is well-established that the hippocampus plays a key role in associative learning and memory (Mayes, Montaldi and Migo, 2007), the role of medial temporal lobe structures in visual short-term memory binding is still subject to debate (Parra *et al.*, 2009, 2010; Koen *et al.*, 2017; Liang *et al.*, 2017; Parra, 2017; Schneegans and Bays, 2019). However, three separate studies using the "What was where?" task have reported results suggestive of hippocampal involvement in visual short-term memory binding.

Liang *et al.* found that smaller hippocampal volume was associated with increased binding errors in presymptomatic mutation carriers, but hippocampal volume was not associated with object or location recall (Liang *et al.*, 2016). Two other studies of patients with medial temporal lobe damage due to voltage-gated potassium channel complex antibody-associated limbic encephalitis (Pertzov *et al.*, 2013), and patients who had undergone anterior temporal lobectomy for treatment of pharmacoresistant epilepsy (Zokaei, Nour, *et al.*, 2019) both reported that the patients had intact memory for object identity and location but a deficit in binding.

The "What was where?" task has also provided evidence of interesting effects of the *APOE*-ε4 allele on short-term memory. In two studies comparing the performance of ε4-carriers and non-carriers – one in middle-aged adults (mean age 45.7) and one in older adults (mean age 68.8) – an apparent *advantage* for *APOE*-ε4 carriers was observed in terms of more accurate recall for object locations after delays of a few seconds (Zokaei *et al.*, 2017; Zokaei, Čepukaitytė, *et al.*, 2019). In the older adults, this was in contrast to a long-term memory task where *APOE*-ε4 carriers had poorer recall for locations after a delay of about 20 minutes compared to non-carriers (Zokaei, Čepukaitytė, *et al.*, 2019), a finding which is suggestive of subtle memory decline associated with preclinical AD, although no biomarker data were available in this study. The authors proposed two possible explanations for the apparent beneficial effect of *APOE*-ε4 on short-term localisation memory: (1) that it could reflect compensatory mechanisms whereby there may be increased activation in frontal and parietal regions that are not directly affected by prodromal AD pathology but are implicated in attention and short-term memory; (2) that it could reflect phenotypical effects of the *APOE*-ε4 allele whereby the ε4 allele may have some beneficial effects in earlier life which could explain its survival in the population (the principle of antagonistic pleiotropy, discussed in section 2.6.1). As the Insight 46 protocol includes both *APOE* genotyping and amyloid-PET imaging, this provides an opportunity to investigate potential independent effects of *APOE*-ε4 and β-amyloid pathology on the "What was where?" task in more detail.

In terms of factors that may predict between-subject differences in performance on the "What was where?" task, one study (n = 66) tested for sex differences and found none (Zokaei, Čepukaitytė, *et al.*, 2019), although another (n = 60) reported that the beneficial effect of *APOE*-ε4 described above was specific to males (Zokaei *et al.*, 2017). No studies so far have examined the effects of childhood cognitive ability, educational attainment, or socioeconomic position. Accounting for potential effects of these predictors may increase the sensitivity of "What was where?" outcomes to subtle effects of *APOE*-ε4 and β-amyloid pathology.

Following the structure of the previous chapter, the aims of this study were firstly to understand patterns of performance on the “What was where?” task and characterise associations between task performance and demographic and life-course predictors, and secondly to investigate associations between performance and biomarkers of brain pathologies among cognitively-normal participants.

5.2. Methods

5.2.1. Stimuli and Procedure

The stimuli and procedure of the ‘What was where?’ task have been described in detail in previous papers (Pertzov *et al.*, 2012, 2013, 2015; Liang *et al.*, 2016). The participant was seated in front of a DELL Optiplex 9030 all-in-one touchscreen computer. The dimensions of the screen were 51.2 x 28.7 cm and the approximate distance from the subject’s eyes to the centre of the screen was 58 cm.

The procedure for the “What was where?” task is presented in Figure 5-1. In each trial, one or three fractals were displayed on the screen in random locations, presented on a black background. Participants were asked to look at the fractals and to try to remember their identities and locations.

1-fractal trials are referred to as ‘low load’ and 3-fractal trials are referred to as ‘high load’. The low load trials were displayed for 1 second whereas the high load trials were displayed for 3 seconds to allow time for encoding. This was followed by a blank screen for either a short or long delay (1 or 4 seconds), and then a test array appeared in which two fractals were displayed along the vertical meridian. One of these fractals had appeared in the memory array on the previous screen (the target) and the other was a foil or distractor. Participants were instructed to touch the fractal that they remembered seeing and drag it to the location where they think it was originally presented (Figure 5-1). There was no time-limit for reporting the location – the tester pressed the space bar to initiate the next trial when the participant was ready.

Previous studies using the “What was where?” task have administered at least 100 trials (Pertzov *et al.*, 2012, 2013, 2015; Liang *et al.*, 2016; Zokaei *et al.*, 2017; Zokaei, Čepukaitytė, *et al.*, 2019), but for Insight 46 a shortened version was used containing 24 trials: 4 low load and 20 high load. Within each load condition, trials were equally likely to have a short or long delay. Therefore, there were four possible combinations of load

and delay (2 x low load with short delay; 2 x low load with long delay; 10 x high load with short delay; 10 x high load with long delay). The experiment was preceded by 4 practice trials – one of each of the load x delay combinations, and the tester ensured that the participant understood the task before continuing.

All fractals including the foils were drawn from a pool of 60 fractals that were used across the experiment (rendered using http://sprott.physics.wisc.edu/fractals.htm, see Figure A1 in the Appendix).

The locations of the fractals were generated in a pseudo-randomised manner by a MATLAB script (MathWorks, Inc). The script imposed the following restrictions which are necessary to allow analysis of localisation error which is a key outcome of this task: fractals were always at least 9^{o} away from each other to avoid crowding and to ensure that there was a clear zone of 4.5^{o} around each fractal which is necessary for the calculation of swap errors (see below), and fractals were at least 6.5^{o} from the centre of the screen and 3.9^{o} from the edges. The 24 trials were the same for all participants (i.e. the same fractals were presented in the same locations) but the trials were presented in a random order for each participant. The reason for using a random order is to avoid the results being confounded by practice effects on the one hand (familiarity with the procedure could cause performance to improve throughout the task) and by interference effects on the other hand (as fractals appear more than once during the task, the foil in the test array could be recognised from a previous trial, which could increase the likelihood of errors in object identification throughout the task).

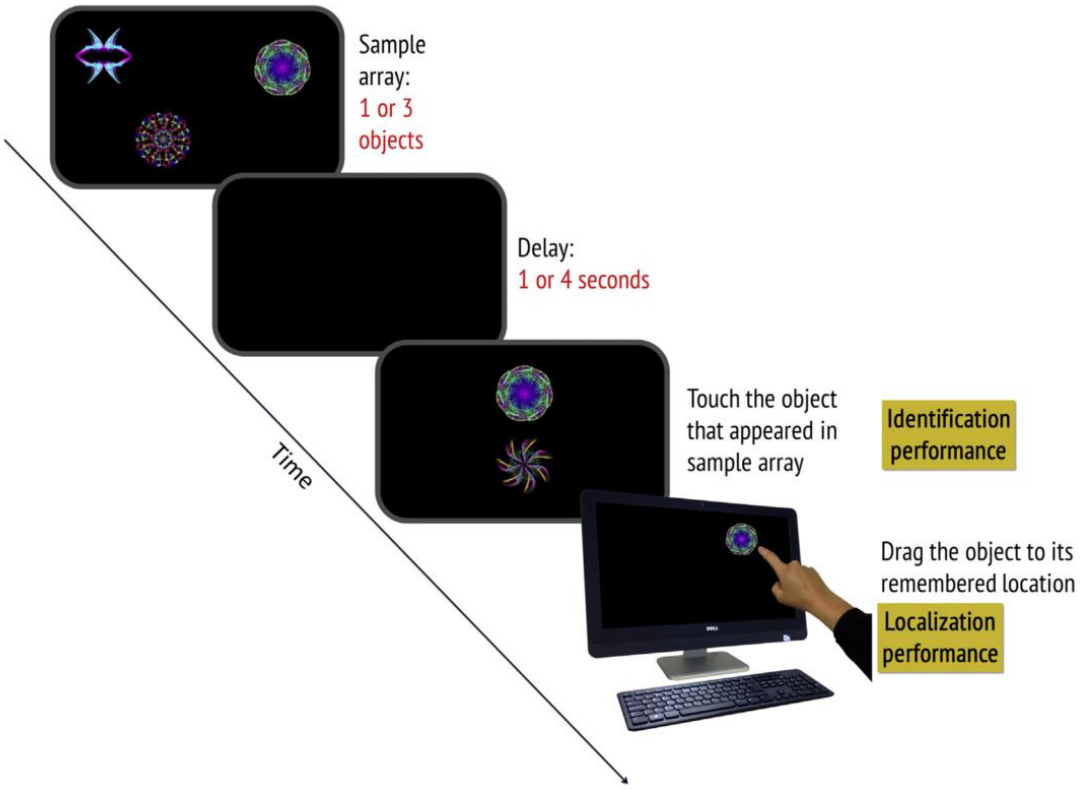


Figure 5-1. The "What was where?" task

Figure reprinted from Liang et al. (2016) under the terms of the Creative Commons Attribution License (CC BY).

5.2.2. Outcome Variables

Memory for object **identification** was defined as the percentage of trials in which the participant chose the correct fractal from the test array (Figure 5-2).

Memory for object location was defined in two different ways. The first, **gross localisation error**, is the distance between the centre of the target in the location reported by the participant and its true location in the original memory array, measured in degrees of visual angle (Figure 5-2B). The second definition takes account of the fact that, in high load trials, participants may mislocalise the target to the location of a different (unprobed) object from the memory array (i.e. they make a swap error – see definition below). In this situation the gross localisation error could be very large. The **nearest item control** measure of localisation error accounts for this by subtracting out the effect of swap errors: it is defined as the distance between the location reported by the participant and the closest of the three original locations from the memory array (Figure 5-2C).

Localisation error was only analysed in trials where the participant identified the correct fractal.

A **swap error** occurs when a participant correctly identifies the target fractal but they swap its location of the target with the location of another object (Figure 5-2C). If the target is positioned within 4.5º of the location of a different object from the memory array, this is counted as a swap error. 4.5º was used as the threshold to ensure that a location could not be attributed to more than one object, as objects were always at least 9º apart. Note that in the low load condition it is not possible to make a swap error as there is only one fractal in the memory array. For each participant, the percentage of swap errors (out of the number of trials in which they identified the correct fractal) was calculated. It is possible that swap errors could occur by chance, a possibility which is discussed and tested in section 5.3.2.3.

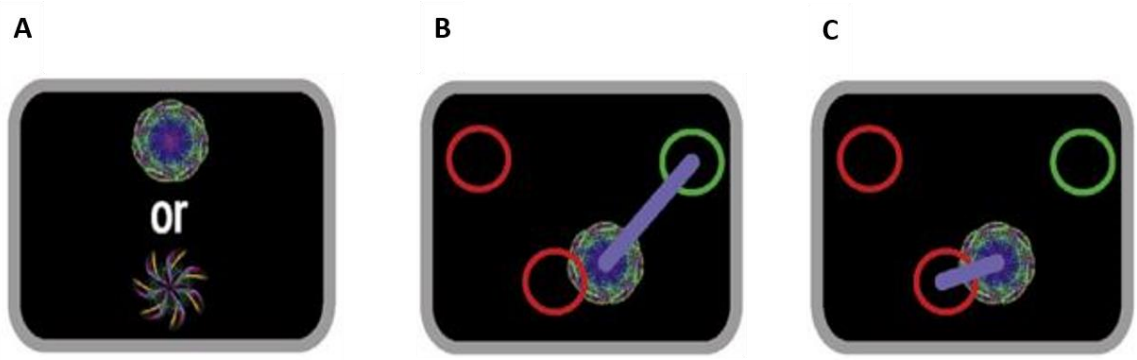


Figure 5-2. Outcome measures on the "What was where?" task

Figure reprinted from Liang et al. (2016) under the terms of the Creative Commons Attribution License (CC BY). Green circles indicate the original location of the target fractal; red circles indicate the original locations of non-target fractals; blue line indicates measured localisation error. (A) Object identification: the participant is required to select the fractal that they remember seeing. (B) Gross localisation error is measured from the location reported by the participant to the original location of the target fractal. (C) 'Nearest item control' localisation error is measured from the location reported by the participant to the location of the closest fractal. If the reported location is within 4.5º of the location of a non-target fractal, this is considered to be a swap error.

5.2.3. Hypotheses

Based on the literature discussed in section 4.1, I hypothesised that cognitively-normal Aβ+ participants would make a greater number of swap errors than Aβ- participants, indicating a subtle deficit in memory binding.

I also aimed to test the hypotheses that *APOE*-ε4 carriers would perform better than non-carriers on localisation memory, and that smaller hippocampal volume would be associated with an increased number of swap errors.

5.2.4. Participants

486 participants completed the "What was where?" experiment (see section 3.3). Participant characteristics are reported in section 3.6.

5.2.5. Data Processing

Each participant's raw data file was processed using a MATLAB script which generated a score for each outcome variable in each of the four conditions (low load with short delay; low load with long delay; high load with short delay; high load with long delay).

Six participants had one trial where the software did not record whether they selected the correct or incorrect fractal. I corresponded with Yoni Pertzov, the creator of the test, and he concluded that this was likely to be caused by the participant touching the screen exactly midway between the two fractals. These six individual trials were excluded and the mean scores for the relevant conditions were recalculated.

One previous study excluded participants who had an object identification accuracy of less than 70%, to ensure that interpretations were not made based on performance at chance level (Liang *et al.*, 2016). Thirty-six participants had an overall identification rate of less than 70%, of whom one performed below the 50% chance level with a score of 46%. This participant was classified as cognitively normal (see section 3.2.3 for criteria). I decided not to exclude any participants from analysis based on their identification rate because other studies in healthy participants have not employed such exclusion criteria (Pertzov *et al.*, 2012, 2015) whereas the study by Liang *et al.* included patients with symptomatic FAD, so adopting a 70% threshold was a way of excluding individuals with significant memory impairment whose data were not been pertinent to the research question (Liang *et al.*, 2016).

5.3. Patterns and predictors of performance

5.3.1. Statistical Analyses

Analysis of **identification** used a GEE logistic regression model for the odds of correctly identifying the fractal, with an independent correlation structure and robust standard errors to allow for the correlation between repeated measures of the same participant. The outcome was number of correct identifications, which was treated as a proportion of the total number of trials in each condition (2 x low load with short delay; 2 x low load with long delay; 10 x high load with short delay; 10 x high load with long delay). For the 6 participants who had a trial excluded during the data cleaning process (see 5.2.5), the total number of trials in each condition was reduced accordingly. Results are expressed as odds ratios for ease of interpretation

Regression models were fitted for the **localisation error** variables (gross error and 'nearest item control') using GEE assuming a normal distribution for the dependent variable and an identity link (as with standard linear regression), but including an exchangeable correlation structure and robust standard errors. Localisation errors were first log-transformed as the distributions were positively skewed.

Analysis of **swap errors** was carried out using the absolute measure rather than the corrected measures (see sections 5.2.2 and 5.3.2.3). The corrected measures are strict as they are based on the upper limit of swap errors that could be explained by chance, so they are more suitable for checking the validity of the absolute measures by confirming that the results cannot be explained by chance (see section 5.3.2.3). The absolute measure is more appropriate for cases in which performance of different participants is compared. A GEE logistic regression model was used for the odds of making a swap error, with an independent correlation structure and robust standard errors. The outcome was number of swap errors, which was treated as a proportion of the number of trials in which the participant identified the correct fractal in each condition (high load with short delay; high load with long delay). Results are expressed as odds ratios for ease of interpretation.

Factors in the regression models were load (low *vs.* high), delay (short *vs.* long), age at assessment, sex, childhood cognitive ability, education, adult socioeconomic position and presence of a neurological or psychiatric condition (yes *vs.* no).

As previous studies have found an interaction between load and delay whereby the detrimental effect of long delay was disproportionately greater when the load was high

(Pertzov *et al.*, 2012, 2015), I tested for such an interaction on the identification and localisation outcomes. (This is not applicable to swap errors since swap errors cannot be made in the low load condition.)

Previous studies have reported interactions between load and between-subjects predictors (such as clinical group, age at assessment, sex or *APOE* genotype). For example, Pertzov *et al.* found that patients with medial temporal lobe damage had disproportionately poorer localisation in the high load condition compared to controls (2013), and older adults had disproportionately poorer identification and localisation in the high load condition compared to younger adults (2015). Similarly, interactions between delay and between-subjects predictors have been reported; for example, Zokaei *et al.* (2017) reported that the detrimental effect of a longer delay on localisation performance was reduced in *APOE*-ε4 carriers compared to non-carriers. To investigate whether similar interaction effects were present among Insight 46 participants, I tested for interactions between load (low *vs.* high) and between-subject predictors (age, sex, childhood cognitive ability, adult socioeconomic position and presence of a neurological or psychiatric condition), and between delay (short *vs.* long) and between-subject predictors.

5.3.2. Results

Descriptive statistics for the three outcome measures are shown in Table 5-1. Results of the multivariable regression models for the three primary outcomes are given in Table 5-2. In addition, results of statistically significant interactions are reported in the text.

Table 5-1. Descriptive statistics for the "What was where?" task

		Low load		High load	
		Short delay	**Long delay**	**Short delay**	**Long delay**
Identification (% correct)	Median	100	100	80	80
	IQR	100 - 100	100 - 100	70 - 90	70 - 90
	Range	50 – 100	0 – 100	40 – 100	30 – 100
Gross Localisation Error (degrees of visual angle)	Median	2.05	2.38	7.65	7.36
	IQR	1.42 – 2.73	1.73 – 3.14	5.47 – 9.84	5.90 – 9.54
	Range	0.12 – 20.84	0.28 – 14.36	0.79 – 20.68	2.33 – 20.57
"Nearest Item Control" Localisation Error (degrees of visual angle) *	Median	2.05	2.38	3.09	3.43
	IQR	1.42 – 2.73	1.73 – 3.14	2.54 – 3.86	2.88 – 4.09
	Range	0.12 – 20.84	0.28 – 14.36	0.79 – 8.41	1.54 – 7.96
Swap errors (%)	Median	N/A		17	17
	IQR			11 - 29	11 - 29
	Range			0 – 100	0 – 75

** In the low load condition, Nearest Item Control is the same as Gross Localisation Error because there is only one fractal so no swap errors can be made. IQR = interquartile range.*

Table 5-2. Associations between demographic and life-course predictors and "What was where?" outcomes (n = 486)

Predictor	**Identification:** odds ratio and 95% confidence intervals	**Localisation error** (degrees of visual angle, log-transformed): coefficient and 95% CIs		**Swap errors:** odds ratio and 95% confidence intervals
		Gross error	**Nearest Item Control**	
High load (low load as reference)	**0.23*** (0.18, 0.30)	**1.23*** (1.19, 1.27)	**0.44*** (0.40, 0.47)	N/A
Long delay (short delay as reference)	0.96 (0.90, 1.05)	**0.09*** (0.06, 0.13)	**0.13*** (0.11, 0.16)	1.00 (0.90, 1.12)
Sex (female as reference)	**0.90** (0.80, 1.00)	**-0.12*** (-0.17, -0.06)	**-0.12*** (-0.17, -0.07)	1.03 (0.91, 1.17)
Age at assessment (per year)	0.95 (0.86, 1.04)	-0.01 (-0.04, 0.01)	-0.01 (-0.05, 0.03)	1.01 (0.92, 1.11)
Childhood cognitive ability (per z-score)	**1.20*** (1.11, 1.31)	-0.02 (-0.06, 0.03)	-0.00 (-0.05, 0.04)	**0.90** (0.81, 1.00)
Education (per category) [a]	0.98 (0.93, 1.03)	-0.01 (-0.04, 0.01)	-0.02 (-0.04, 0.01)	0.99 (0.94, 1.06)
Adult SEP (per category) [a]	1.04 (0.98, 1.11)	-0.02 (-0.04, 0.01)	-0.01 (-0.04, 0.02)	1.01 (0.94, 1.07)
Neurological or psychiatric condition [b] (cognitively-normal as reference)	0.95 (0.80, 1.13)	**0.10** (0.01, 0.20)	0.08 (-0.01, 0.17)	1.12 (0.90, 1.38)

Coefficients in bold are significant at p < 0.05 and asterisks indicate significance at p < 0.01. Multivariable regression models were used so each association is independent of all others.
[a] See section 3.2.4 for definition of categories; [b] See section 3.2.3 for definitions.

CI = confidence interval; SEP = socioeconomic position

5.3.2.1. Load and delay

5.3.2.1.1. Identification

As expected, identification performance was poorer in the high load condition compared to the low load condition (Table 5-1, Table 5-2, Figure 5-3). In contrast to previous studies that have reported poorer identification after a long delay than a short delay (Pertzov *et al.*, 2012, 2015; Zokaei *et al.*, 2017), there was no statistically significant effect of delay on identification performance (Table 5-2, Figure 5-3). However, there was an interaction between load and delay, whereby the long delay had a more detrimental effect on identification performance in the low load condition, compared to the high load condition (*OR* = 7.27, *95% CIs* 4.22 to 12.51, $p < 0.0001$, Figure 5-3). This is contrary to previous studies which reported that the long delay had a disproportionately greater effect when there were more items to be remembered (Pertzov *et al.*, 2015; Zokaei *et al.*, 2017).

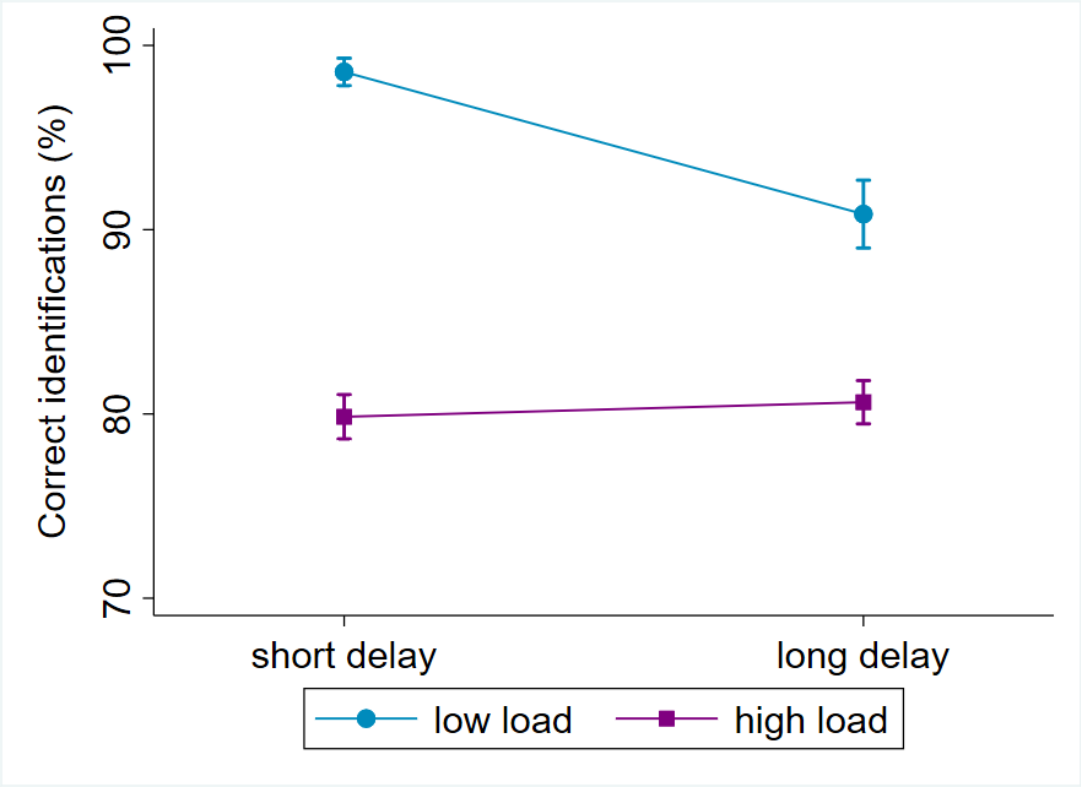


Figure 5-3. Identification performance on the "What was where?" task, by load and delay: means and 95% confidence intervals

Markers show unadjusted means and error bars show 95% confidence intervals. This figure illustrates that, although delay did not have a statistically significant effect overall, there was an interaction between load and delay.

5.3.2.1.2. Localisation error

As expected, localisation error was greater in the high load condition than the low load condition (Table 5-1, Table 5-2, Figure 5-4). This applied to the 'nearest item control' measure as well as the gross error measure, showing that the increased localisation error in the high load condition cannot be explained by swap errors. Also, as expected, localisation error was greater after a long delay compared to a short delay, both on the gross error measure and on the 'nearest item control' measure (Table 5-2, Figure 5-4).

In contrast to previous studies that have reported that the detrimental effect of a longer delay was exaggerated when there were more items to be remembered (Pertzov *et al.*, 2012, 2015; Zokaei *et al.*, 2017), there was a significant interaction between load and delay in the opposite direction, such that the detrimental effect of a long delay was slightly reduced in the high load condition compared to the low load condition, on both the gross localisation error (regression *coefficient* = -0.15, *95% CIs* -0.23 to -0.08, $p < 0.0001$) and 'nearest item control' measure (regression *coefficient* = -0.07, *95% CIs* -0.14 to -0.01, $p = 0.018$). However, the interaction effect appears minimal on visual inspection, as the lines are almost parallel (Figure 5-4).

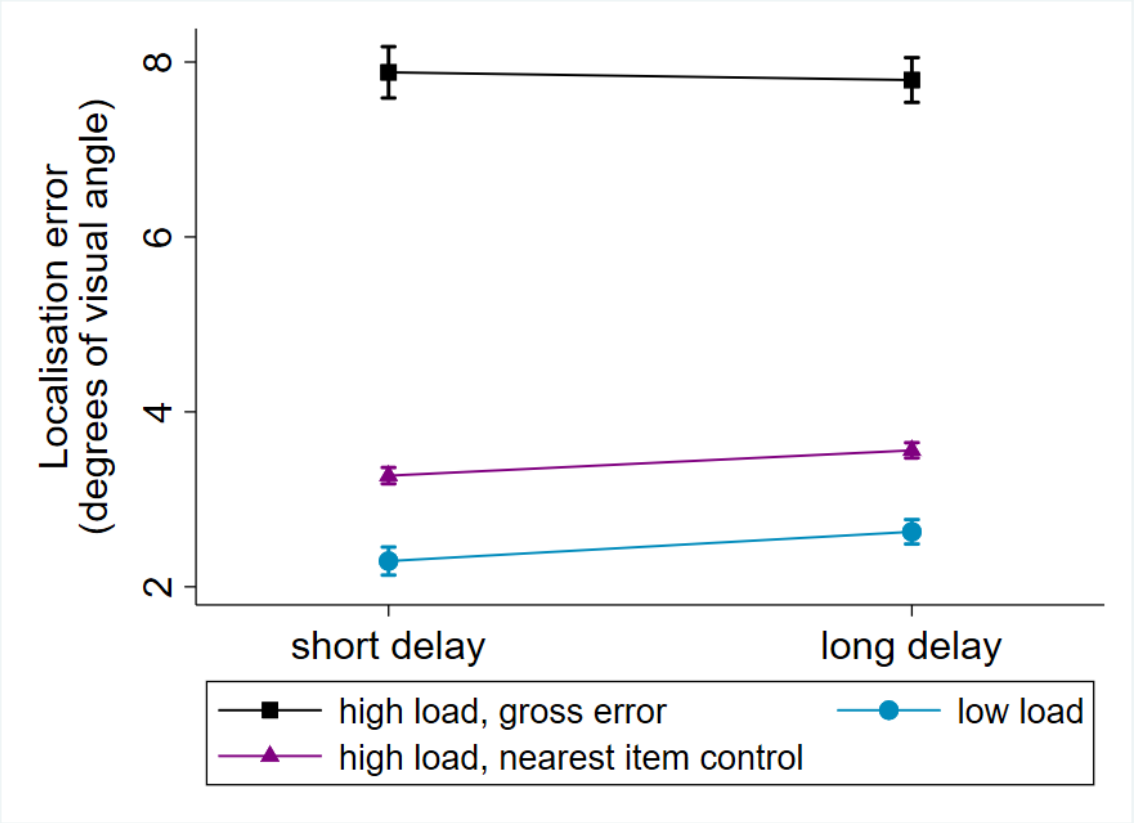


Figure 5-4. Localisation error on the "What was where?" task, by load and delay: means and 95% confidence intervals

Markers show unadjusted means and error bars show 95% confidence intervals. 'Nearest item control' accounts for the impact of swap errors – see section 5.2.2 for full definitions. (Note that swap errors cannot be made in the low load condition as there is only one fractal, so no 'nearest item control' is necessary.)

5.3.2.1.3. Swap errors

Delay had no significant effect on the proportion of swap errors (Table 5-2), in contrast to a previous study where healthy participants made more swap errors in the long delay condition (Pertzov *et al.*, 2012).

5.3.2.2. Demographic and life-course predictors

5.3.2.2.1. Childhood cognitive ability, education and adult socioeconomic position

Higher childhood cognitive ability was associated with better identification performance (Table 5-2, Figure 5-5). There was evidence that this effect was exaggerated in the long delay condition compared to the short delay (*OR* = 1.14, *95% CIs* 1.01 to 1.30, *p* = 0.037). There was no evidence of a statistically significant association between childhood cognitive ability and localisation error, either in the gross error measure nor the 'nearest item control' measure (Table 5-2). Participants with higher childhood cognitive ability tended to make fewer swap errors (Table 5-2, Figure 5-6).

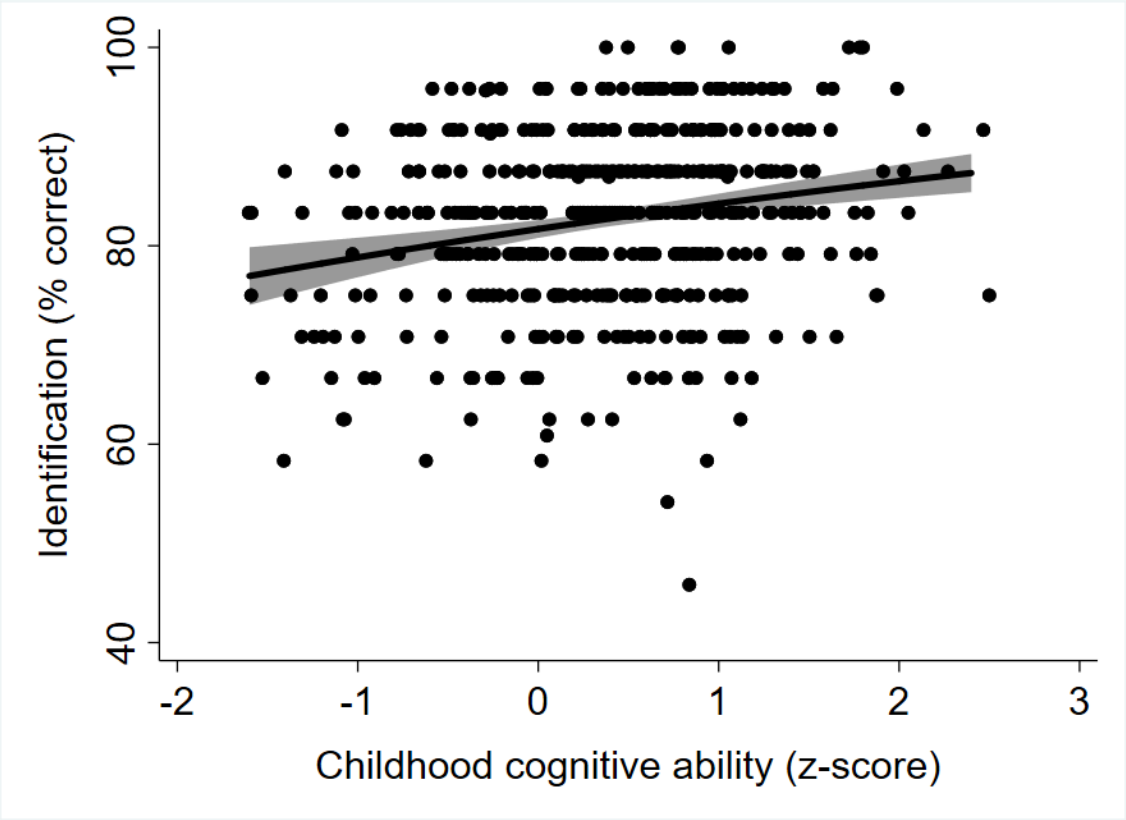


Figure 5-5. Association between childhood cognitive ability and identification memory on the "What was where?" task

Solid line represents prediction from the multivariate regression model, adjusted for sex, age at assessment, education, adult socioeconomic position and presence of a neurological or psychiatric condition. Shaded area represents 95% confidence intervals. Markers show each participant's identification rate across the experiment as a whole (combined across the different conditions of load and delay). For an explanation of the childhood cognitive ability variable, see section 3.2.4.

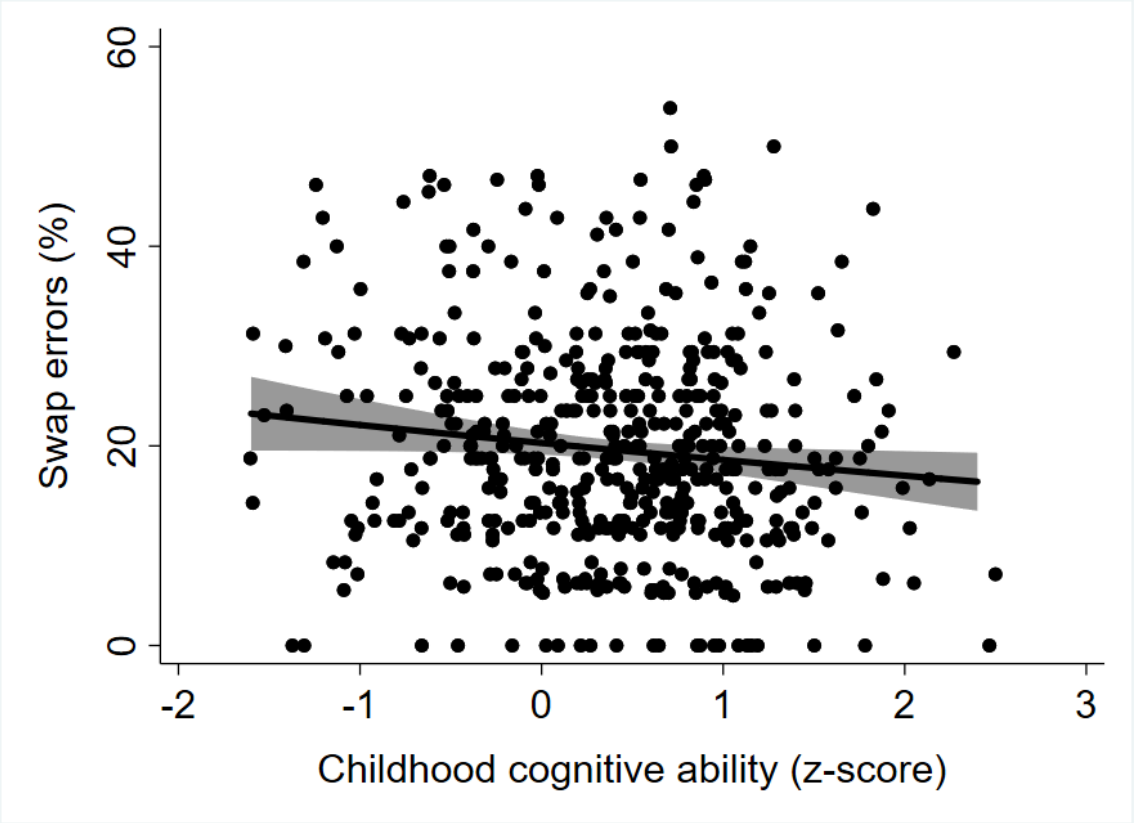


Figure 5-6. Association between childhood cognitive ability and swap errors on the "What was where?" task

Solid line represents prediction from the multivariate regression model, adjusted for sex, age at assessment, education, adult socioeconomic position and presence of a neurological or psychiatric condition. Shaded area represents 95% confidence intervals. Markers show each participant's swap error rate across the experiment as a whole (as a percentage of trials on which they identified the correct fractal). For an explanation of the childhood cognitive ability variable, see section 3.2.4.

Educational attainment and adult socioeconomic position did not show evidence of associations with performance on any outcome measure (Table 5-2).

5.3.2.2.2. Sex differences

Males had slightly poorer identification memory than females on average, having 10% lower odds of identify the correct fractal (Table 5-2). However, males made smaller localisation errors than females on average, both in the gross measure (unadjusted means: 4.93º *vs.* 5.39º) and 'nearest item control' (unadjusted means: 2.78º *vs.* 3.10º) (Table 5-2). There was evidence that this sex difference was reduced in the high load condition compared to the low load condition (gross error: *interaction coefficient* = 0.09, 95% CIs 0.01 to 0.18, $p = 0.034$; nearest item control: *interaction coefficient* = 0.09, 95% CIs 0.02 to 0.16, $p = 0.012$) and in the long delay condition compared the short delay

condition (gross error: *interaction coefficient* = 0.13, 95% CIs 0.05 to 0.20, *p* = 0.13; nearest item control: *interaction coefficient* = 0.09, 95% CIs 0.03 to 0.15, *p* = 0.004). The untransformed results for males and females are illustrated in Figure 5-7 for the 'nearest item control' measure, as an aid to the interpretation of these interaction effects. Overall, these results indicate that males tended to report the location of objects more accurately than females when the memory demands were easier, whereas sex differences were reduced when the memory demands were harder.

There was no sex difference in the proportion of swap errors (Table 5-2).

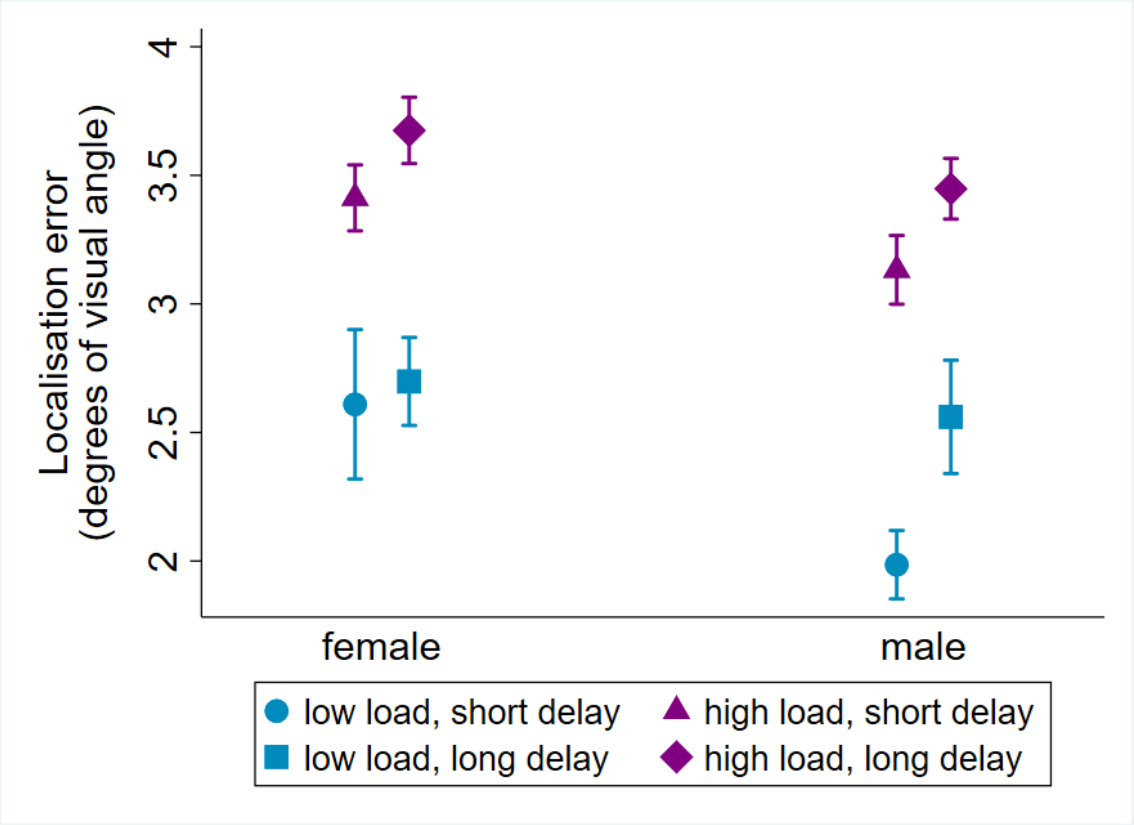


Figure 5-7. 'Nearest Item Control' localisation error on the "What was where?" task for males and females, by load and delay

Markers show unadjusted means and error bars show 95% confidence intervals.

5.3.2.2.3. Age at assessment

Age did not show evidence of associations with performance on any outcome measure (Table 5-2).

5.3.2.2.4. Neurological and psychiatric conditions

Participants with neurological and psychiatric conditions made slightly larger gross localisation errors than cognitively-normal participants, although this did not reach statistical significance in the 'nearest item control' measure ($p = 0.067$) (Table 5-2). There was no evidence that participants with neurological and psychiatric conditions differed from cognitively-normal participants in terms of identification performance or swap errors (Table 5-2).

5.3.2.3. Exploration of swap errors

It is possible that swap errors could occur by chance – firstly the participant could guess the correct object by chance and secondly they could localise it to the location of a non-target fractal by chance. To account for this, three measures were calculated to estimate the proportion of swap errors that may be expected due to chance for each participant, using the methods set out in previous studies (Pertzov *et al.*, 2012, 2013). As there was no evidence of an effect of delay on the proportion of swap errors (see section 5.3.2.1.3), this was done based on each participant's overall proportion of swap errors across the short and long delay conditions.

Swap errors due to chance identification. If a participant could not remember the identity of the object and guessed it correctly by chance, but their memory for the locations of the objects was intact, they would be expected to localise the object around one of the three remembered locations. Therefore, in one third of cases they would be expected to localise it to the target location and in two thirds of cases they would be expected to localise it to one of the two non-target locations, which would be recorded as a swap error. Therefore, the upper limit of the number of swap errors that may be expected due to guessing the identity of the object correctly by chance can be calculated by multiplying each participant's number of identification errors by two thirds. This gives a number of "swap errors due to chance identification".

Swap errors due to chance localisation. For every trial on which a swap error was observed, the probability that the swap error was obtained by chance despite the locations having been forgotten can be estimated by simulating all possible locations (using steps of 1^{o}) that the target could have been placed with the same amplitude of error, and then calculating the proportion of these locations that would be classified as a

swap error (Figure 5-8). Combining these probabilities across all the trials gives a number of “swap errors due to chance localisation” for each participant.

Swap errors due to chance identification and localisation can be calculated by combining the two measures described above. This represents the strictest upper limit of swap errors that could be expected due to chance.

Figure 5-8. Swap errors on the “What was where?” task

Dashed green circles indicate the original location of the target fractal, dashed red circles indicate the original locations of non-target fractals. The left-hand image shows the location reported by the participant in this example. The right-hand image shows all potential locations that the fractal could be placed by chance, assuming the same amplitude. Blue lines are drawn to locations that would not be classified as a swap error, and orange lines are drawn to locations that would be classified as a swap error (i.e. within 4.5^{o} of a non-target location). Therefore, the probability of obtaining this swap error by chance equals the number of orange lines divided by the total number of orange and blue lines. Reprinted from Pertzov et al. (2013) under the terms of the Creative Commons Attribution License (CC BY-NC 3.0).

The mean and *SD* for the various swap error measures were as follows: observed swap errors = 20 ± 11%; swap errors expected due to chance identification = 9 ± 6%; swap errors expected due to chance localisation = 2 ± 2%; swap errors expected due to chance identification and localisation = 11 ± 6%.

Paired sample t-tests were used to test the hypothesis that the observed proportion of swap errors was significantly greater than the proportion of swap errors expected due to chance. This hypothesis was supported (swap errors due to chance identification: $t(485) = 26.5$, $p < 0.0001$; swap errors due to chance localisation $t(485) = 36.2$, $p < 0.0001$; swap errors due to chance identification and localisation $t(485) = 24.4$, $p < 0.0001$). These results confirm that when participants did not remember the correct location of the target fractal, they did not report a random location but tended to mislocalise it around the locations of the other two fractals.

5.4. Associations with biomarkers and *APOE*-ε4

Following the format laid out in section 3.5, the second part of this chapter aims to investigate associations between performance on the "What was where?" task and biomarkers of AD in cognitively-normal participants for whom complete biomarker data are available. The number of participants meeting these criteria who also had usable data from the "What was where?" task was 398 (see section 3.3).

As explained in section 3.5.2.1, I wanted to derive some summary scores that capture the key aspects of performance on each task, to use for comparing results across the different cognitive tests in the Insight 46 battery (see Chapter 9). For the "What was where?" task I calculated the following three summary outcomes for each participant:

i) **Identification**, defined as the percentage of trials on which the correct fractal was selected (out of 24 trials). This is a measure of memory for the identity of objects.
ii) **Mean 'nearest item control' localisation error**, defined as above (see section 5.2.2). The scores were log-transformed to more closely approximate the normal distribution. I chose the 'nearest item control' measure over the gross error measure (see section 5.2.2) because the gross error measure is heavily influenced by swap errors, whereas 'nearest item control' is a purer measure of memory for the location of objects.
iii) **Swap errors**, defined as the percentage of trials on which a swap error was made (see definition in section 5.2.2). This is a measure of 'misbinding' of objects and locations in memory.

Combining outcomes across the different conditions of load (low *vs.* high) and delay (short *vs.* long) was justified because my hypotheses about associations between AD pathology and "What was where?" outcome measures (see section 5.2.3) were general rather than specific to certain conditions.

For each of the three summary outcomes, I tested for associations with the same biomarkers as in the previous chapter (see section 3.5.2). I also tested for associations between "What was where?" outcomes and hippocampal volume, in order to test the hypothesis that smaller hippocampal volume would be associated with increased swap errors. **Hippocampal volume** was generated using the Similarity and Truth Estimation for Propagated Segmentations (STEPS) automated segmentation method with appropriate manual editing (Cardoso *et al.*, 2013).

5.4.1. Statistical Analyses

As before, **identification** was analysed using a GEE logistic regression model for the odds of correctly identifying the fractal with an independent correlation structure and robust standard errors. The outcome was number of correct identifications, which was treated as a proportion of the total number of trials (24 for most people, but 23 for the 6 participants who had a trial excluded during the data cleaning process – see 5.2.5). Results are expressed as odds ratios for ease of interpretation.

Mean 'nearest item control' localisation error was analysed using a linear regression model.

As before, a GEE logistic regression model was used for the odds of making a **swap error**, with an independent correlation structure and robust standard errors. The outcome was number of swap errors, which was treated as a proportion of the number of trials in which the participant identified the correct fractal. Results are expressed as odds ratios for ease of interpretation.

All models included predictors of amyloid status (positive *vs.* negative), whole brain volume, WMHV and *APOE* genotype (ε4-carrier *vs.* non-carrier). To adjust for the correlation between whole brain volume and head size, total intracranial volume (TIV) was included in all models, as were the demographic factors investigated in section 6.3 (sex, age at assessment, childhood cognitive ability, education and adult socioeconomic position).

The models were additionally rerun replacing dichotomised amyloid status with SUVR to test whether increasing Aβ deposition was associated with differences in performance. To check whether associations between SUVR and cognition were sensitive to the inclusion of the imputed SUVR values (see section 3.2.2), the analyses were rerun excluding the 26 participants with imputed SUVR data.

To test for associations between "What was where?" outcomes and hippocampal volume (total volume of left plus right hippocampus), I reran the above models with covariates of hippocampal volume, sex, age at assessment, childhood cognitive ability, education, adult socioeconomic position and total intracranial volume.

5.4.2. Results

Results of the regression models for the three outcomes are reported in Table 5-3. Results for the demographic and life-course factors (sex, age at assessment, childhood cognitive ability, education and adult socioeconomic position) are not reported as they are essentially unchanged from the first analysis section (5.3.2).

Table 5-3. Associations between biomarkers and "What was where?" outcomes in cognitively-normal participants (n = 398)

Predictor	**Identification**: odds ratio and 95% CIs	**Mean "Nearest Item Control" localisation error** (degrees of visual angle, log-transformed): coefficient and 95% CIs	**Swap errors**: odds ratio and 95% CIs
β-amyloid status (negative as reference)	**0.84** (0.72, 0.97)	0.03 (-0.04, 0.09)	0.93 (0.77, 1.14)
WMHV (per 10 ml)	1.01 (0.92, 1.11)	-0.00 (-0.04, 0.04)	0.97 (0.88, 1.08)
Whole brain volume (per 10 ml)	1.01 (0.99, 1.02)	-0.00 (-0.01, 0.00)	1.00 (0.98, 1.02)
***APOE*-ε4** (non-carriers as reference)	**1.15** (1.01, 1.31)	**-0.07** (-0.13, -0.02)	1.01 (0.87, 1.17)

Multivariable regression models were used so each association is independent of all others. In addition to the predictors listed, models also included sex, age at assessment, childhood cognitive ability, adult socioeconomic position, and total intracranial volume.

Coefficients in bold are significant at p < 0.05 and asterisks indicate significance at p < 0.01.

CI = confidence interval; WMHV = white matter hyperintensity volume

On average, Aβ+ participants had slightly poorer identification memory than Aβ- participants (adjusted means from the multivariable regression model: 80% *vs.* 83% correct) (Table 5-3). There was no evidence of statistically significant differences between the amyloid groups in terms of localisation error or swap errors (Table 5-3). As an aid to interpretation, the unadjusted untransformed means are quoted as follows: 'nearest item control' localisation error: Aβ+ = 3.19º; Aβ- = 3.21º; swap errors: Aβ+ = 18.9%; Aβ- = 20.0%.

When rerunning the models replacing dichotomous amyloid status with continuous SUVR, there were no statistically significant associations between SUVR and any of the

"What was where?" outcome variables (Identification: *Odds ratio* = 0.55, *95% CIs* 0.23 to 1.32, *p* = 0.18; Localisation error: *regression coefficient* = 0.08, *95% CIs* -0.28 to 0.43, *p* = 0.67; Swap errors: *Odds ratio* = 0.73, *95% CIs* 0.22 to 2.36, *p* = 0.59). These results were unchanged in a sensitivity analysis excluding the individuals with imputed SUVR values (see section 3.2.2).

Although the raw unadjusted scores for object identification were slightly lower in *APOE*-ε4-carriers compared to non-carriers (ε4-carriers = 82% correct; non-carriers = 84% correct), ε4-carriers were found to perform better than non-carriers on this measure after adjustment for amyloid status and all other covariates (Table 5-3). As the effect of *APOE*-ε4 worked in the opposite direction to the effect of Aβ, this means that the best-performing group were Aβ- ε4-carriers, and the worst-performing group were Aβ+ non-carriers (Figure 5-9).

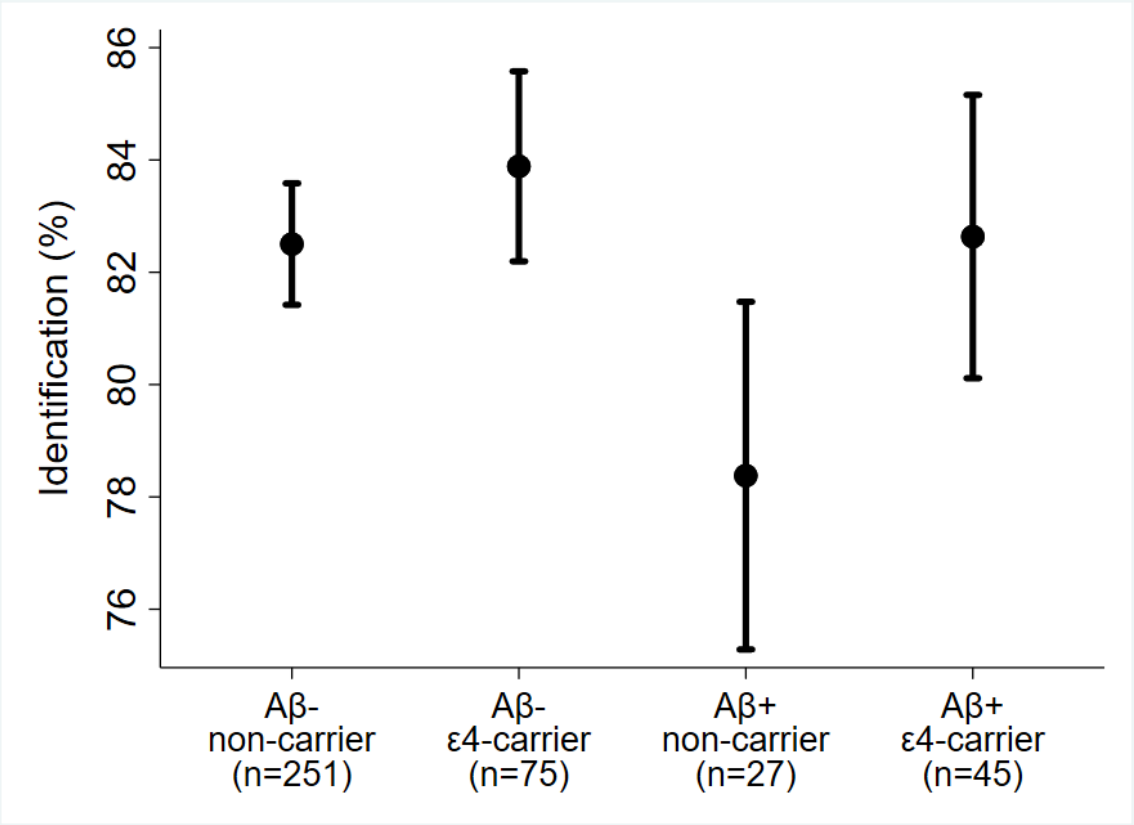


Figure 5-9. Identification performance on the "What was where?" task by amyloid status and *APOE* genotype: means and 95% confidence intervals

Markers show unadjusted means and error bars show 95% confidence intervals.

APOE-ε4 carriers also had better localisation memory than non-carriers on average (Table 5-3), with a regression coefficient equivalent to -0.25 *SD*. A t-test using the raw unadjusted localisation error scores confirmed that this group difference was statistically

significant without adjusting for any other covariates (ε4-carriers = 3.0° error; non-carriers = 3.3° error, $t(396) = 2.50$, $p = 0.013$). As a previous study using the "What was where?" task reported that the beneficial effect of *APOE*-ε4 on localisation memory was only present in males (Zokaei *et al.*, 2017), I tested for an interaction between sex and *APOE* genotype: there was no evidence of such an interaction (interaction coefficient = 0.02, *95% CIs* -0.08 to 0.13, $p = 0.67$).

Whole brain volume and WMHV did not have any statistically significant effects on any of the "What was where?" outcomes (Table 5-3).

There was no evidence of associations between hippocampal volume and any "What was where?" outcomes (Identification: *OR* = 1.00 per additional ml, *95% CIs* 0.91 to 1.11, $p = 0.93$; Localisation error: -0.02, *95% CIs* -0.06 to 0.03, $p = 0.45$; Swap errors: *OR* = 1.06, *95% CIs* 0.94 to 1.19, p = 0.38).

5.5. Discussion

5.5.1. Summary

The "What was where?" task measured memory for the identity and location of objects, and binding of these two features in memory. Subtle sex differences were observed for the first time on this task, with females performing better on the identification measure but males performing better on the localisation measure. Higher childhood cognitive ability was associated with better memory for the identity of objects and fewer binding errors. The results did not support the hypothesis that cognitively-normal Aβ+ participants would show subtle binding deficits, but there was evidence that Aβ+ participants had slightly poorer memory for object identity. As predicted, *APOE*-ε4 appeared to have a beneficial effect on memory for locations, and a beneficial effect on memory for object identity was also observed. There was no evidence to support the hypothesis of an association between hippocampal volume and memory binding.

5.5.2. Patterns of performance

Performance on the three outcomes (identification, localisation and binding) was broadly similar to previous studies that have used this task in healthy adults (Pertzov *et al.*, 2012, 2015; Zokaei *et al.*, 2017; Zokaei, Čepukaitytė, *et al.*, 2019). Neither a ceiling nor a floor

effect was observed in terms of identification rate (mean = 83%; *SD* = 8.9) and the percentage of swap errors (mean = 20%) was in line with previous studies (Pertzov *et al.*, 2012, 2015).

Unexpectedly, the length of the delay before participants were asked to report the identity and location of the fractal had little impact on their performance. There was no evidence of a statistically significant effect of delay duration on identification, although localisation performance was a little poorer after the longer delay as expected. Furthermore, this study did not observe the expected interactions between load and delay, whereby the detrimental effects of a long delay on identification and localisation were expected to be disproportionately greater when there were more items to be remembered (high load) compared to the low load condition (Pertzov *et al.*, 2012, 2015; Zokaei *et al.*, 2017). One possible explanation for this could derive from the way that this shortened version of the task was adapted from the original version. The 24 trials in this version were randomly selected from the 100 trials in the original task, and it is possible that the 'short delay' and 'long delay' trials could have ended up unbalanced in terms of difficulty. Anecdotally the difficulty of individual trials can vary depending on the features of the fractals such as their relative salience, similarity and positions (for example, if the target and foil fractals in the test array happen to contain similar colours, it may be more difficult to choose between them). This variability in the appearance of the fractals is inherent to the task and is impossible to fully control for. Also, as 24 trials is a small number compared to many neuropsychology tests, the signal-to-noise ratio may not have been good enough to measure these effects.

5.5.3. Demographic and life-course predictors

5.5.3.1. Associations with childhood cognitive ability, education and adult socioeconomic position

Higher childhood cognitive ability was associated with better memory for the identity of objects and fewer 'misbinding' errors. It is interesting that localisation memory did not show evidence of an association with childhood cognitive ability, especially given that the continuous nature of this measure (as opposed to a discrete correct *vs.* incorrect recall measure) makes it sensitive to small differences in performance (Pertzov *et al.*, 2013). I would speculate that the adoption of encoding strategies may possibly be relevant here; anecdotally, I observed that participants sometimes commented on trying out different encoding strategies for remembering the identity of the objects, for example *"I tried to give the shapes a name in my head"* or *"I tried to remember the colours",* and

sometimes remarked on whether or not such strategies had been successful. I do not recall any participants commenting on strategies for remembering the locations and I would argue that the range of possible strategies for remembering precise locations on a blank screen may be more limited. I would speculate that the greater potential for adopting a range of strategies to encode object identities may be of relevance in explaining why the identification memory measure appears to be more related to general cognitive ability, as it introduces an element of problem-solving into the task, creating potential for participants to generate and adapt strategies throughout the task. Although fractal stimuli were used to reduce the likelihood of verbal encoding, it is interesting to note that such strategies are effectively an attempt to introduce a verbal memory component into the task (i.e. encoding the names of colours or shapes), allowing the object identities to be more easily rehearsed during the delay interval.

Given that previous studies have reported the 'swap error' measure to be a sensitive and specific measure of memory binding impairment in individuals with presymptomatic FAD (Liang *et al.*, 2016) and patients with medial temporal lobe damage (Pertzov *et al.*, 2013; Zokaei, Nour, *et al.*, 2019), the novel association reported here between higher childhood cognitive ability and fewer swap errors has implications for such studies. While most cohorts do not have access to data on childhood cognitive ability, interpretation of results should consider the fact that individual differences in memory binding ability may be partially driven by general cognitive ability as well as by pathology.

5.5.3.2. Sex differences

The finding that males had better memory for the location of objects is interesting in the context of a previous study using the "What was where?" task which reported an interaction between sex and *APOE* genotype on the localisation memory measure: male *APOE*-ε4 carriers were found to be at an advantage compared to male non-carriers in terms of localisation memory, but no such effect of *APOE*-ε4 was observed in females (Zokaei *et al.*, 2017). As that study found no statistically significant overall difference between males and females on the localisation measure (although they did report a significant sex difference in terms of swap errors), the authors' interpretation focussed on reasons why *APOE*-ε4 may have differential effects on men and women, rather than on reasons why there may be sex differences in localisation memory. However, the direction of their data appears to be consistent with the effect I report here, and their hypotheses about differential effects of *APOE*-ε4 were not supported by their later study which reported the same effect of *APOE*-ε4 in both males and females (Zokaei, Čepukaitytė, *et al.*, 2019). Perhaps the lack of a statistically significant sex difference in localisation memory in these two studies could be explained by the fact that it is a very

subtle effect and those studies had much smaller sample sizes (n=60 and n=66 respectively) and contained wider age ranges.

As the male advantage for localisation was primarily driven by the conditions with easier memory demands (1 object to remember and 1-second delay), this suggests that the difference may lie in the precision of encoding and/or reporting locations, rather than in the rate of decay of representations in memory (in which case the opposite result would be expected, i.e. a greater sex difference after 4 seconds delay). One previous study conducted a detailed examination of sex differences in reporting of point locations within a blank circle and within photographs of complex natural scenes, and concluded that the sex differences may arise during encoding as the differences were present even when there was no retention interval (i.e. participants were asked to reproduce a location from an image on a screen onto a response sheet while the screen remained visible) (Holden, Duff-Canning and Hampson, 2015). The nature of the errors in this task revealed an interesting pattern from which the authors concluded that women were more likely to rely on categorical cues whereas men tended to rely on metric information. Within a blank circle, women were more likely to localise the point closer to the centre of one of the four quadrants that would be formed if the horizontal and vertical axes of the circle were drawn in. Similarly in the photographic scenes, they were more likely to localise the point closer to the centre of the section that the stimulus location was in, where the sections were defined by an algorithm using edge detection and shared perceptual features (e.g. the shadow on the side of a mountain). Both cases suggest an increasing tendency to categorise the image into regions of interest and to adopt a strategy of remembering which region the point is in. Anecdotally, this fits with my perception of watching participants complete the “What was where?” task, where I got the impression that some participants appeared to be recalling the locations using strategies such as “over by the right-hand edge” or “up towards the top-left corner”, as they sometimes seemed to make stereotyped errors such as locating the fractal as far towards the edge or corner as possible.

There is a well-established literature on sex differences in spatial abilities, with males reported to outperform females on a range of tasks such as spatial navigation and mental rotation, in rats as well as humans (for meta-analyses see Voyer, Voyer and Bryden, 1995; Jonasson, 2005; Voyer, Voyer and Saint-Aubin, 2017). Various explanations for sex differences in spatial abilities have been advanced, including hormonal (see Jonasson for a review in rats) and evolutionary factors (e.g. Ecuyer-Dab and Robert, 2004). However, there is also evidence that sex differences in spatial abilities in humans are dependent on cultural factors. In a recent analysis of over 2.5 million people from all 195 nation states who played a mobile-app-based game designed to measure spatial

navigation ability, there was a consistent sex difference across nations with males performing better, but the size of a nation's sex difference was positively correlated with its level of gender inequality, as defined by the World Economic Forum's Gender Gap Index (Coutrot *et al.*, 2018). Consistent with the evidence described above of a sex difference in preference for cues versus metric information, studies of spatial navigation have reported sex differences in navigation strategies, with females tending to place more emphasis on allocentric strategies (based on a perception of the positions of landmarks and their relationships to other landmarks) and males tending to rely more on egocentric strategies (based on an internal representation of spatial information from one's own viewpoint) (Saucier *et al.*, 2002; Jonasson, 2005).

With respect to sex differences on object location memory tasks specifically, two meta-analyses have concluded that women outperform men, but these meta-analyses focussed on specific types of task which require recall of categorical locations. The first meta-analysis included variations of a task where participants are asked to study an array of objects and then are presented with a new array in which some objects have moved and others are in the same places as before – the task is to identify which objects have switched places (Voyer *et al.*, 2007). The second meta-analysis focussed on a task where participants are asked to memorize the locations of pairs of coloured dots hidden underneath flaps (similar to the 'Pairs' card game) (Voyer, Voyer and Saint-Aubin, 2017). It has been argued that the apparent female advantage on these sorts of location memory tasks may be dependent on non-spatial aspects of these tasks (Andreano and Cahill, 2009; Holden, Duff-Canning and Hampson, 2015), which may explain the discrepancy with tasks based on recall of absolute metric locations such as "What was where?".

This argument may also be relevant to my finding that females had slightly better memory for the identity of objects, as good performance on this measure may be aided by verbalisation strategies (as alluded to in section 5.5.3.1) rather than relying purely on visual memory. This result is in line with the results from the other memory tests in the Insight 46 battery (Logical Memory and FNAME – see Chapter 4) where females had higher scores on average.

In summary, the finding of sex differences on the "What was where?" task is novel, but is consistent with a body of literature on sex differences in verbal and visuospatial abilities. This finding is noteworthy for the fact that both male and female advantages were demonstrated in a single task.

5.5.3.3. Associations with age

Age effects have previously been observed on the "What was where?" task, whereby older adults had poorer memory for object identity and location compared to younger adults, although there was no age effect on memory binding (Pertzov *et al.*, 2015). It is not surprising that no such evidence of associations between age and performance were seen among Insight 46 participants, given that the age range was so small (2.6 years- reflecting the time it took to collect the data since all participants were born in the same week).

5.5.4. Associations with biomarkers and APOE-ε4

Previous studies have suggested that visual memory binding may be one of the first aspects of memory to decline as AD pathology accumulates, as subtle binding deficits have been detected in presymptomatic FAD mutation carriers, at a stage when memory for the individual features was unaffected (Parra *et al.*, 2010, 2011, 2015; Liang *et al.*, 2016). The results of this study contradict this as there was no evidence of a difference between cognitively-normal Aβ+ and Aβ- participants in terms of object-location binding, but Aβ+ participants did have slightly poorer memory for the identity of objects (as well as evidence of slightly lower scores on the Logical Memory Immediate recall task reported in Chapter 4). To my knowledge, associations between Aβ deposition and visual short-term memory binding have not been investigated before in cognitively-normal older adults, but binding deficits in MCI (Pietto *et al.*, 2016; Parra *et al.*, 2017) and AD are well-established (Parra *et al.*, 2009, 2010, 2011; Fernández *et al.*, 2018), so further studies are needed to determine at what point such deficits may emerge. As the presymptomatic mutation carriers in the study by Liang *et al.* showed evidence of hippocampal atrophy which correlated with greater binding deficits, and Aβ deposition is understood to precede hippocampal atrophy by some years (Bateman *et al.*, 2012; Jack *et al.*, 2013) it is currently unclear whether binding deficits may be detectable before measurable neurodegeneration occurs.

This is now the third independent cohort to report a beneficial effect of *APOE*-ε4 on memory for locations in the "What was where?" task (Zokaei *et al.*, 2017; Zokaei, Čepukaitytė, *et al.*, 2019). This is consistent with a recent review of the association between *APOE*-ε4 and cognition in younger adulthood, which highlighted spatial memory as one area of cognition in which *APOE*-ε4 carriers appear to be at an advantage, along with advantages in many aspects of physical fitness (Smith, Ashford and Perfetti, 2019).

The authors of this review acknowledged the issue of "transition from ε4-associated cognitive advantage to cognitive deficit" across the life span, but it is important to remember that beneficial effects of *APOE*-ε4 on cognition may not cease to operate in older age, even if outweighed on average by the detrimental effects of Aβ-pathology – as highlighted by the novel finding that *APOE*-ε4 carriers in Insight 46 had better memory for object identity on the "What was where?" task, *once the detrimental effect of Aβ had been accounted for*. Many studies comparing *APOE*-ε4 carriers and non-carriers in older age do not have access to biomarker data on Aβ levels, so it is possible that subtle independent effects of *APOE*-ε4 are being masked by Aβ. This could potentially explain why Zokaei *et al.* observed no effect of *APOE*-ε4 on memory for object identity, if this measure is more sensitive to Aβ-pathology as the results reported here suggest. It could also explain why the *APOE*-ε4 carriers in their study showed evidence of *poorer* memory on a delayed recall task, whereas in Insight 46 *APOE*-ε4 carriers performed better on a delayed recall task (Logical Memory – see Chapter 4) after the detrimental effect of Aβ was accounted for. This finding on the Logical Memory task further supports the hypothesis that the purported beneficial effect of *APOE*-ε4 is not specific to location memory. Furthermore, in light of evidence from studies of spatial navigation that have observed poorer performance in *APOE*-ε4 carriers, both in younger and older adults (Kunz *et al.*, 2015; Coughlan *et al.*, 2019), it is possible that the superior performance of *APOE*-ε4 carriers on the "What was where?" localisation measure could be primarily due to its demands on working memory and attention, rather than spatial abilities *per se*. This is discussed further in Chapter 10.

5.5.5. Strengths and limitations

Although anecdotally this was one of the tasks on which participants were most likely to report a perception of poor performance, it worked well with no floor or ceiling effects, and only one participant declining to complete the full task (see section 3.3). The results suggest that the adapted short form of the "What was where?" task used in Insight 46 is sufficient to measure the outcomes of interest, and suitable for detecting subtle differences in performance in this age group. This is encouraging as this version (24 trials) takes approximately 8-10 minutes to complete, whereas the original version was much longer with 100 trials. For participants who did report finding this task particularly difficult, a much longer version could have been demoralising. However, more trials may be necessary for studies which aim to examine the effects of load and delay, as the small

number of trials may have affected our ability to detect differences between the conditions, especially since the low load condition only had 4 trials.

The software had a limitation that resulted in occasional instances of the opposite fractal being selected to the one the participant intended. This happened when a participant selected a fractal and dragged it to their chosen location, passing through the unselected fractal on the way. This could generally be avoided by instructing the participant to tap their chosen fractal first and then tap the location, rather than doing both in one movement. It was not possible to record which trials this occurred on, but my estimate is that it affected 5-10% of trials for 1-5% of participants i.e. no more than 0.5% of trials in total. This means that a few participants may have received lower identification scores than they should have. Localisation error and swap error scores are unlikely to have been affected since they are calculated based on correct identifications only. I am confident that this has not affected the results and conclusions, but would recommend for future studies that the software is altered to require the identification step to be logged before moving onto the localisation step.

Strengths and limitations that apply to all the analyses presented in this thesis, such as considerations relating to the generalisability of the sample, are discussed in Chapter 10.

5.5.6. Conclusions

In summary, this sensitive computerised task allows precise measurement of multiple aspects of visual short-term memory, revealing an interesting pattern of sex differences and a novel dissociation between the effects of Aβ pathology and *APOE*-ε4, whereby Aβ was associated with a subtle deficit in object recognition memory whereas *APOE*-ε4 was associated with superior performance.

6. CHOICE REACTION TIME

6.1. Introduction

Processing speed is an aspect of cognition in which age-related decline is particularly notable (Salthouse, 2000; Harrington *et al.*, 2018), and also an area which becomes impaired early in the AD process. Evidence that decline in processing speed can be detected in the preclinical phase of AD comes primarily from studies that have reported a difference between cognitively-normal individuals with and without amyloid-β and tau pathology, on paper-and-pencil tasks such as the WAIS-Digit Symbol and Trail Making Test A (e.g. Baker *et al.*, 2017; Ho and Nation, 2018). An aspect of processing speed which merits further investigation as a potential marker of subtle cognitive decline in preclinical AD is reaction time (RT) because it underpins many other cognitive processes and is of constant relevance to daily functioning. Simple RT tasks can be conceptualised as requiring sustained attention and the energisation and execution of a pre-planned motor or other response (Stuss *et al.*, 2005). Choice RT tasks (which require selection of the correct response from two or more options) can be considered to additionally require monitoring a larger set of stimuli and inhibiting responses to non-target stimuli (Stuss *et al.*, 2005). Choice RT may be a particularly informative measure because – compared to simple RT – it is more sensitive to age and cerebral dysfunction (Benton, 1986; Der and Deary, 2006). As well as response speed itself, another interesting aspect of performance on such tasks is intra-individual variability (or inconsistency) in RT.

It is well-established that Choice RT slows with age. Most evidence comes from cross-sectional studies comparing younger and older adults (Deary and Der, 2005; Bugg *et al.*, 2006; Der and Deary, 2006; Garrett, Macdonald and Craik, 2012; Vincent *et al.*, 2012; Nissan, Liewald and Deary, 2013; Woods *et al.*, 2015), but longitudinal decline has also been reported (Ritchie *et al.*, 2016). While there is a component of motor slowing, most of the age-related increase in RT is understood to be due to slower cognitive processing, which includes the specific steps of response selection and response generation (Salthouse, 1996; Godefroy *et al.*, 2010; Woods *et al.*, 2015), as well as to a decline in attention (Godefroy *et al.*, 2010). Intra-individual variability in RT is also widely-reported to increase with age (Anstey *et al.*, 2005; Deary and Der, 2005; Williams *et al.*, 2005; Dykiert *et al.*, 2012a; Nissan, Liewald and Deary, 2013). As intra-individual variability in RT is a function of overall response speed (i.e. variability increases as responses get slower), experimental studies often use normalised measures of intra-individual variability that account for each individual's average RT. One such index is the "coefficient of variation", defined as *standard deviation / mean* (e.g. (Der and Deary, 2006) or *interquartile range / median* (e.g. (Phillips *et al.*, 2013)).

Impaired RT and increased intra-individual variability in RT are common across a variety of neurological conditions including the dementias, traumatic brain injury, schizophrenia and attention deficit hyperactivity disorder (Benton, 1986; MacDonald, Nyberg and Bäckman, 2006). A number of diverse associations suggest that RT might be a general indicator of the integrity of the central nervous system (MacDonald, Nyberg and Bäckman, 2006), for example, RT correlates with a wide range of physical health measures such as grip strength, forced expiratory volume, corrected visual acuity and concentration of macular pigment (Anstey *et al.*, 2005; Feeney *et al.*, 2013). Insight 46 participants have previously completed a Choice RT task at age 60-64, as part of the wider MRC National Survey of Health and Development (NSHD), and the results provided evidence that slower RT is associated with greater adiposity (waist circumference) (Masi *et al.*, 2018) and impaired kidney function, partially accounted for by socioeconomic factors (Silverwood *et al.*, 2014).

Multiple studies have reported that patients with AD have slower and more variable RTs than controls (Hultsch *et al.*, 2000; Burton *et al.*, 2006; Phillips *et al.*, 2013; Christ, Combrinck and Thomas, 2018). There is evidence that decline in RT can be detected before the onset of AD dementia, as patients with MCI are reported to have slower RTs than controls (Christensen *et al.*, 2005; Ballesteros, Mayas and Reales, 2013; Phillips *et al.*, 2013; Haworth *et al.*, 2016). One study found that RT discriminated the MCI and control groups better than the commonly-used Trail-Making Test, another measure of information processing speed that arguably relies more on multiple high-level functions such as working memory, cognitive flexibility and set-shifting (Haworth *et al.*, 2016).

Intra-individual variability in RT has received relatively little attention as a potential marker of early cognitive decline in preclinical AD, but some studies have investigated intra-individual variability in RT in patients with MCI and reported mixed results. One study reported that patients with MCI were more variable than controls (Christensen *et al.*, 2005), whereas another reported that patients with MCI had larger standard deviations of RT than controls but this was accounted for by their slower overall response speed (i.e. there was no difference in the coefficient of variation) (Phillips *et al.*, 2013). A third study reported that MCI patients did not differ from controls in terms of intra-individual variability on a simple RT task, but greater variability within the MCI group was predictive of developing dementia within a 2.5 year follow-up (Tales *et al.*, 2012). To my knowledge, no studies have investigated associations between Choice RT (response speed and intra-individual variability) and biomarkers of AD pathology in cognitively-normal older adults.

If Choice RT and/or intra-individual variability in RT are to be useful measures in detecting and tracking subtle cognitive decline in the preclinical phase of AD, it is

important to disentangle disease-related decline from ageing effects, and also to account for other factors which may contribute to individual differences in RT, such as sex and prior cognitive ability. Factors associated with individual differences in RT are briefly reviewed below.

The relationship between RT and intelligence has been widely investigated, with the aim of understanding how differences in cognitive ability may be underpinned by differences in the simple cognitive processes that support more complex information processing. Associations between higher intelligence and faster responses in Choice RT tasks have been reported across adulthood (e.g. Deary, Der and Ford, 2001; Sheppard and Vernon, 2008; Deary, Johnson and Starr, 2010; Nissan, Liewald and Deary, 2013). Proposed explanations for this association include (1) intelligence is built up over time by the processing of information and RT is a measure of information processing efficiency (Vernon, 1983); (2) both are at least partially dependent on attention (Doebler and Scheffler, 2016); and (3) they share genetic predictive factors (Sheppard and Vernon, 2008). Evidence from the Lothian Birth Cohort has shown that childhood cognitive ability is predictive of Choice RT at age 70 (Deary, Johnson and Starr, 2010).

Higher intelligence has also been associated with lower intra-individual variability in RT, as reported in a recent meta-analysis (Doebler and Scheffler, 2016) – however this meta-analysis did not adjust the variability measure for its dependence on overall response speed, which may account for most of the association with intelligence (Der and Deary, 2017). Having access to data on Insight 46 participants' childhood cognitive ability and educational attainment provides an (imperfect) way to account for differences in general mental ability, which may improve the sensitivity of the Choice RT task to potential effects of brain pathologies. It also provides a rare opportunity to examine whether childhood cognitive ability and educational attainment have independent effects on RT and intra-individual variability in RT in older age.

Some studies have found no sex differences on Choice RT tasks (e.g. (Woods *et al.*, 2015)), while others have reported that males tend to have faster and less variable RTs (Deary and Der, 2005; Der and Deary, 2006; Dykiert *et al.*, 2012b; Vincent *et al.*, 2012; Phillips *et al.*, 2013; Haworth *et al.*, 2016). Deary and Der suggest a possible hormonal explanation (Deary and Der, 2005), and proposed behavioural explanations include that females may show stronger learning effects and greater post-error slowing (Reimers and Maylor, 2006; Thakkar *et al.*, 2014; Fischer *et al.*, 2016). This is in contrast to perceptual processing speed tests (such as Digit-Symbol Substitution) where females tend to be faster (see Chapter 4).

Another factor that needs to be considered when interpreting individual differences in RT is speed-accuracy trade-offs. Standard practice in Choice RT studies is to analyse error rate separately to RT – an approach which allows comparisons of speed and accuracy between groups, but does not allow a direct analysis of the trade-offs between speed and accuracy on an individual level. Correlations between slower RT and greater accuracy on Choice RT tasks have been observed (Nissan, Liewald and Deary, 2013), but the opposite result has also been reported (i.e. slower responses being associated with poorer accuracy (Vaportzis, Georgiou-Karistianis and Stout, 2013), or no correlation between RT and accuracy (Bugg *et al.*, 2006).

Following the structure of the previous chapter, the aims of this study were firstly to understand patterns of performance on the Choice RT task and to characterise associations between task performance and demographic and life-course predictors, and secondly to investigate associations between performance and biomarkers of brain pathologies among cognitively-normal participants.

6.2. Methods

6.2.1. Stimuli and Procedure

Two related experiments were combined into one task. This chapter concerns the first experiment, a Choice RT task. The second experiment is described in Chapter 7.

The Choice RT task was presented on a DELL Optiplex 9030 all-in-one touchscreen computer, using SuperLab software. A response box with two buttons side by side was connected to the computer and placed on the table in front of the participant. The experiment had 2 blocks of 12 trials each. In block 1 the stimulus was an arrow pointing left or right, and in block 2 the stimulus was a word – 'LEFT' or 'RIGHT' (see Figure 6-1). All stimuli were displayed in the centre of the screen. At the beginning of each trial, there was an interval of 1000ms throughout which a cue was displayed ('Arrow' in Block 1, and 'Word' in Block 2), before the stimulus appeared below it. The purpose of displaying the cue before the stimulus was to pace the experiment and to maintain a consistent format with the second experiment (see Chapter 7) where cues formed a necessary part of the instructions.

Six practice trials preceded each block. The practice trials were followed by a gap of about 5 seconds before the main block started automatically. The order of trials was the same for all participants: LRRRLRLRLLRL in Block 1, and LLLLRRRLRRL in Block 2.

However, the software output indicated that 3 participants were presented with the trials in a different order (RLLLRLRLLRRR in Block 1, and RRLRRLLLLRRL in Block 2), which appeared to be an unexplained consequence of the tester accidentally omitting to complete the *Session ID* field when launching the experiment.

Participants were instructed to press the correct button on the response box (left or right) as quickly as possible, according to the stimulus on the screen. They were asked to use the index and middle fingers of their dominant hand, with one finger on each button. If their response was correct, the next trial was initiated; if their response was incorrect, the stimulus remained on the screen and an error tone sounded to signal that they should respond again as quickly as possible. Regardless of whether their second attempt was correct or incorrect, no feedback was given and the next trial was initiated.

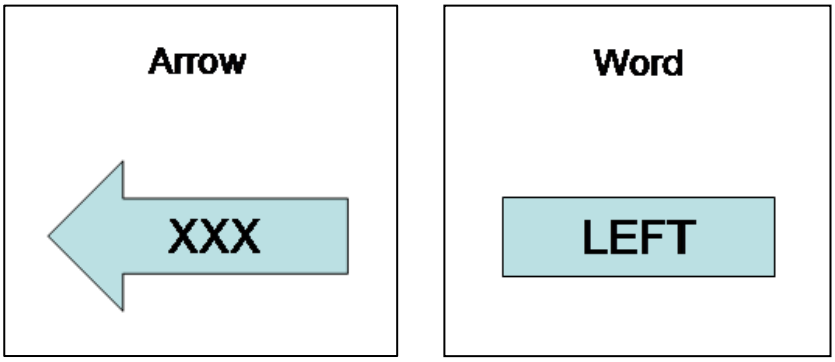


Figure 6-1. The Choice Reaction Time task.

The cue ("Arrow" or "Word") appeared on its own for 1000ms before the stimulus appeared underneath. The XXX inside the arrow is not relevant for the response, but was included so that the appearance and visual complexity of both blocks were as similar as possible.

6.2.2. Outcome variables

Two variables were recorded for each response: **RT (ms)** and **accuracy (correct or incorrect)**. All analyses were based on the initial response to each stimulus; second attempts were not analysed as their purpose was to reorient participants to the task.

Standard practice across many different tasks measuring RT is to analyse RT for correct responses only (e.g. Aron *et al.*, 2004; Forster and Lavie, 2008; Silverwood *et al.*, 2014; Stoet, 2017). This is because the cognitive processes involved in making a correct response are assumed to be different to those involved in making an incorrect response:

reasons for errors include anticipatory responses, attentional slips, and failures in understanding how to reach the correct answer. Therefore, restricting RT analyses to correct responses is a fairer way to compare individuals or groups (notwithstanding the fact that correct responses could also be achieved by anticipatory responses or chance). This approach was adopted here.

The third outcome of interest was **intra-individual variability in RT**. Variability in RT is a function of overall response speed, therefore mean RT needs to be accounted for when investigating individual differences in intra-individual variability. This analysis adopts the commonly-used *coefficient of variation*, calculated by dividing each participant's standard deviation by their mean RT (e.g. Hultsch *et al.*, 2000; Der and Deary, 2006). As before, only correct responses are considered because the RT of incorrect responses is likely to vary depending on what led to the error (e.g. inattention, anticipation, difficultly working out the answer), whereas the variability in RT of correct responses is a purer measure of an individual's variability in repeating a particular response process.

6.2.3. Hypotheses

Based on the literature, I hypothesised that higher childhood cognitive ability, higher educational attainment and male sex would be associated with faster and less variable RTs.

I aimed to test the hypothesis that cognitively-normal Aβ+ participants would have slower and more variable RTs than Aβ- participants.

6.2.4. Participants

All 502 participants completed the Choice RT task (see section 3.3). Participant characteristics are reported in section 3.6.

6.2.5. Data processing

The first trial in each block was not included in any analyses because participants did not always realise that the block had started and this affected RT. This limitation in the experimental design in discussed later (see section 6.5.5).

In line with the policy of keeping the sample as representative as possible, the only reason to exclude participants with outlying performance was deemed to be a clear indication that they deviated from the protocol e.g. a fundamental misunderstanding of the instructions. One participant had a lifelong condition affecting cognition and was noted to have difficulty in understanding the task, in telling left from right, and in confusing the meaning of 'right' with 'correct'. This participant was excluded from the analyses, leaving a sample of 501.

If any participants were to have an error rate of >50%, i.e. suggestive of guessing at chance level, this would be grounds to consider excluding them from the analyses because it would indicate that they might not have understood the instructions, paid attention to the task or known their left from their right. Some Choice RT tasks adopt stricter exclusion criteria (e.g. 90% accuracy (Deary and Der, 2005) or 80% accuracy (Der and Deary, 2006). The majority of participants (75%) made no errors and the highest error rate was 27%, therefore no participants were excluded from analyses based on their error rate.

Although only correct responses were considered for most of the analyses of RT (see section 6.3.1.1), RTs of incorrect responses were considered as part of the analysis of speed-accuracy trade-offs (see section 6.3.1.3). Therefore, for the purpose of checking for outlying RTs, I have considered all responses together, both correct and incorrect. The distribution of each participant's mean RT shows that there were few participants with outlying slow RTs (Figure 6-2). One of these participants had a tremor but this was not grounds to exclude them from analyses because they performed the task according to the instructions and they were not fundamentally different to some other participants with motor symptoms whose performance was not outlying but was affected by their symptoms nonetheless.

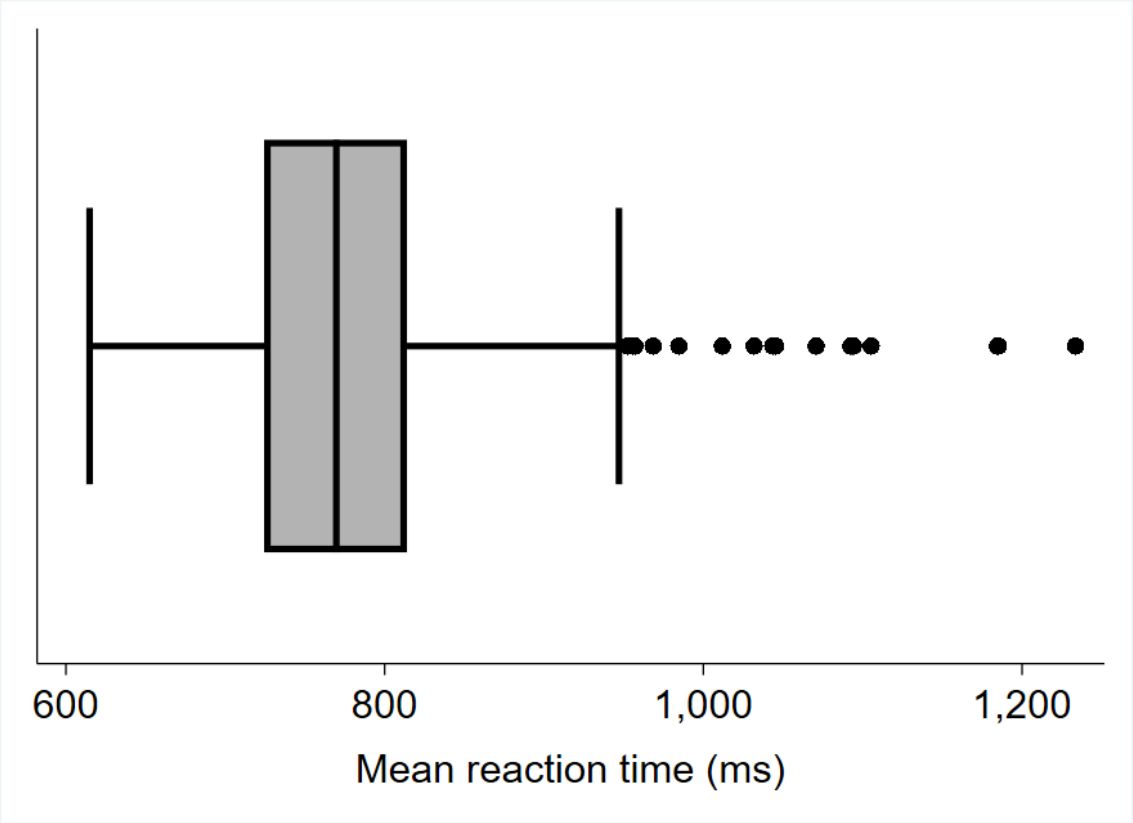


Figure 6-2. Box and whisker plot of mean reaction times (correct and incorrect responses) on the Choice Reaction Time task, before data cleaning

Reaction times to individual trials were checked for outliers. Extremely fast RTs can be assumed to be invalid as there is a minimum amount of time required to process the stimulus. In a previous paper using a similar task, a threshold of 300ms was used as the fastest possible time for a genuine response (Aron *et al.*, 2004) so I adopted the same criterion, although no trials had RTs of less than 300ms. There were a few outlying slow responses (Figure 6-3). Unusually slow responses are unlikely to reflect the participant's ability – for example I observed instances of participants stopping to cough or drink water – so I considered three approaches to choosing a threshold for excluding the slowest trials:

i) Excluding trials where the RT is more than 3 *SD* above the sample mean (i.e. > 1208 ms).
ii) Excluding trials where the RT is more than 3 *SD* above each participant's own mean, as in (Anstey *et al.*, 2005).
iii) Excluding trials where the RT is above an arbitrary cut-off such as 4000 ms as in (Aron *et al.*, 2004) or 1000 ms in (Anstey *et al.*, 2005).

Approach (ii) was adopted as it is most likely to filter out trials that are not representative of an individual's ability. As the range of mean RTs was wide, approach (i) could consistently exclude trials from participants who found the task difficult – possibly the very participants who may provide evidence to confirm the hypothesis that poor performance on this task is associated with brain pathologies. Approach (iii) is too arbitrary and would mainly exclude trials belonging to a few low-performing participants – again possibly the participants of particular interest to the hypothesis.

Therefore, I excluded trials more than 3 *SD* above each participant's own mean, which applied to 40 trials (0.4%) from 40 participants (1 trial each). This reduced the maximum RT to 1637 ms.

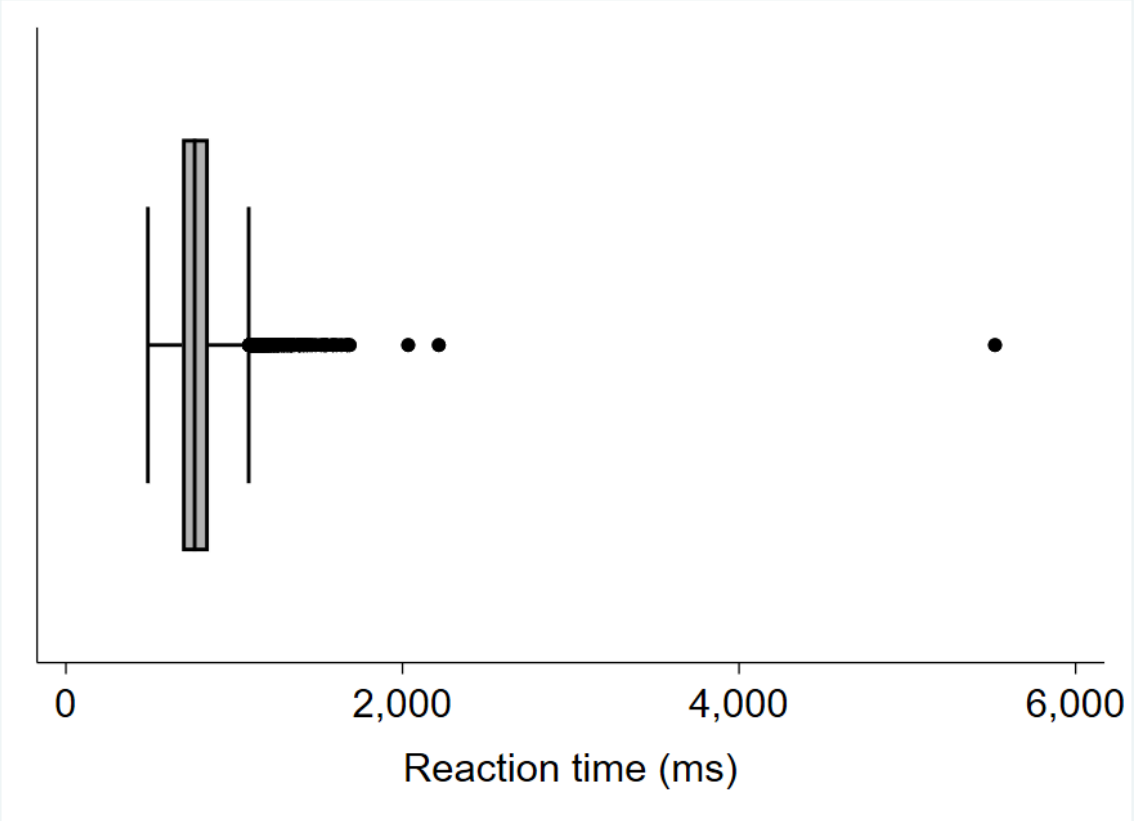


Figure 6-3. Box and whisker plot of individual reaction times (correct and incorrect responses) on the Choice Reaction Time task, before data cleaning

6.3. Patterns and predictors of performance

Following the format laid out in section 3.5, in the first part of this chapter I aimed to describe patterns of performance across the various outcomes and conditions of the task, and to investigate the effects of demographic and life-course predictors on performance in the full Insight 46 sample. Specifically I aimed to investigate whether RT and error rate differ for word and arrow stimuli, explore speed-accuracy trade-offs, and describe intra-individual variability in RT. The demographic and life-course predictors (sex, age at assessment, childhood cognitive ability, educational attainment, adult socioeconomic position and presence of a neurological or psychiatric condition) are defined in sections 3.2.4 and 3.2.3 respectively.

6.3.1. Statistical Analyses

6.3.1.1. RT and error rate

Rather than using summary scores for each participant (e.g. mean RT for arrows and mean RT for words), trial-by-trial responses to each individual stimulus were analysed to avoid losing information. RTs were first log-transformed so that the distribution more closely approximated the normal distribution.

RT was analysed using a GEE model, assuming a normal distribution for the dependent variable and an identity link (as with standard linear regression), but including an exchangeable correlation structure and robust standard errors to allow for the correlation between repeated measures of the same participant.

Response accuracy (correct *vs.* incorrect) was analysed using a GEE logistic regression model with an independent correlation structure and robust standard errors. Results are expressed as odds ratios for ease of interpretation.

Predictors in the models were stimulus type (arrow *vs.* word), sex, age at assessment, childhood cognitive ability, education, adult socioeconomic position and presence of a neurological or psychiatric condition (yes *vs.* no).

6.3.1.2. Practice effects

As the arrow block always preceded the word block, comparison between them could be confounded by practice effects. While this cannot be tested explicitly, exploring practice effects with the two blocks separately could give an indication of whether practice effects are generally observed on this test. Practice effects on RT and error rate were investigated by rerunning the models (see 6.3.1.1) with an additional factor of trial number (1 to 11).

6.3.1.3. Speed-accuracy trade-offs

Speed-accuracy trade-offs can be investigated both between-subject or within-subject. Between-subject analyses address questions such as: "Which participants are most likely to trade speed for accuracy?" or "Which participants tend to perform well on both speed and accuracy?" Within-subjects analyses address questions such as "Are participants more likely to make an error when they respond more quickly, and which experimental conditions promote this response pattern?" The between-subjects questions are my main focus in line with my research questions about whether this task can identify participants whose performance may be indicative of subtle cognitive decline, but the within-subject questions are also of interest in understanding more about general patterns of performance on this task.

To address the between-subject questions, each participant's mean RT for correct responses was compared to their error rate (percentage of incorrect responses), combined across arrow and word stimuli. This approach has been adopted in other studies, for example a study which investigated speed-accuracy trade-offs on a Choice RT experiment in participants with Huntington's disease (Vaportzis *et al.*, 2015a).

The distribution of mean RTs had a positive skew. Two transformations were considered: log-transforming the mean RTs or calculating the mean of the log-transformed raw RTs. As neither of these removed the skew, the untransformed data were used for ease of interpretation. The distribution of error rates also had a strong positive skew due to a ceiling effect on accuracy. The relationship between mean RT and error rate was analysed using Spearman's rank correlation. This nonparametric test was appropriate due to the non-linear relationship between the variables (see section 6.3.2.3), and because the main purpose of this section is a simple investigation of whether those who rank high in terms of speed tend to rank low in terms of accuracy.

To address the within-subject trade-offs, trial-by-trial responses to each individual stimulus were analysed to investigate whether the speed of a response predicted

whether that response would be correct or incorrect. This was done by rerunning the GEE model for the odds of making errors (see 6.3.1.1), with RT included as an additional factor. RT was not log-transformed this time, to aid interpretability of its effects as a predictor (it was previously transformed to comply with the assumption of a normal distribution for linear regression when it was being modelled as an outcome).

6.3.1.4. Intra-individual variability in RT

As intra-individual variability in RT was normally distributed, a linear regression model was used to investigate its associations with sex, age at assessment, childhood cognitive ability, education, adult socioeconomic position and presence of a neurological or psychiatric condition (yes *vs.* no).

Some studies have reported that variability in RT for correct responses is a function of error rate (Hultsch *et al.*, 2000; Der and Deary, 2006), probably because the response following an error is typically slower. To investigate the extent to which intra-individual variability in RT (for correct responses only) may be predicted from error rate on this task, I reran the regression model including error rate as an additional factor.

6.3.2. Results

6.3.2.1. RT and error rate

Descriptive statistics are presented in Table 6-1. Results of the multivariable regression models for RT and error rate are given in Table 6-2.

Table 6-1. Descriptive statistics for the Choice Reaction Time task

		Arrow	Word	Arrow and Word combined
RT for correct responses (ms)	Median	741	789	768
	IQR	668 - 828	723 - 870	695 - 850
	Range	489 - 1586	503 - 1637	489 - 1637
Error rate (%)		3.0	2.2	2.6

Statistics are based on 501 participants who each completed 22 trials. IQR = interquartile range.

Table 6-2. Associations between demographic and life-course predictors and Choice RT outcomes: RT and error rate (n = 501)

Predictor	RT [a] coefficient and 95% CIs	Odds Ratio for making an error and 95% CIs
Arrow stimulus (word stimulus as reference)	**1.07*** (1.06, 1.07)	**0.72*** (0.58, 0.89)
Sex (female as reference)	0.99 (0.97, 1.01)	1.12 (0.85, 1.47)
Age at assessment (per year)	**1.03*** (1.02, 1.04)	**0.81** (0.68, 0.98)
Childhood cognitive ability (per z-score)	**0.99** (0.97, 1.00)	**0.72*** (0.59, 0.89)
Education (per category) [b]	1.00 (0.99, 1.00)	1.06 (0.93, 1.21)
Adult socioeconomic position (per category) [b]	1.00 (0.99, 1.01)	1.04 (0.91, 1.19)
Neurological or psychiatric condition (cognitively-normal as reference) [c]	1.01 (0.98, 1.04)	**1.54*** (1.11, 2.12)

Coefficients in bold are significant at p < 0.05 and asterisks indicate significance at p < 0.01. [a] For the RT outcome, as the data were log-transformed, the coefficients are quoted in exponentiated form for ease of interpretation; for example, a coefficient of 1.05 would mean that the factor was associated with 5% longer response time. Multivariable regression models were used so each association is independent of all others. [b] See section 3.2.4 for definition of categories. [c] See section 3.2.3 for definitions.

CI = confidence interval; RT = reaction time.

On average, responses were 7% slower to words than arrows (Figure 6-4), and the odds of making an error were 28% lower for words than arrows. Taken together, this implies a speed-accuracy trade-off, which is investigated later (see 6.3.2.3).

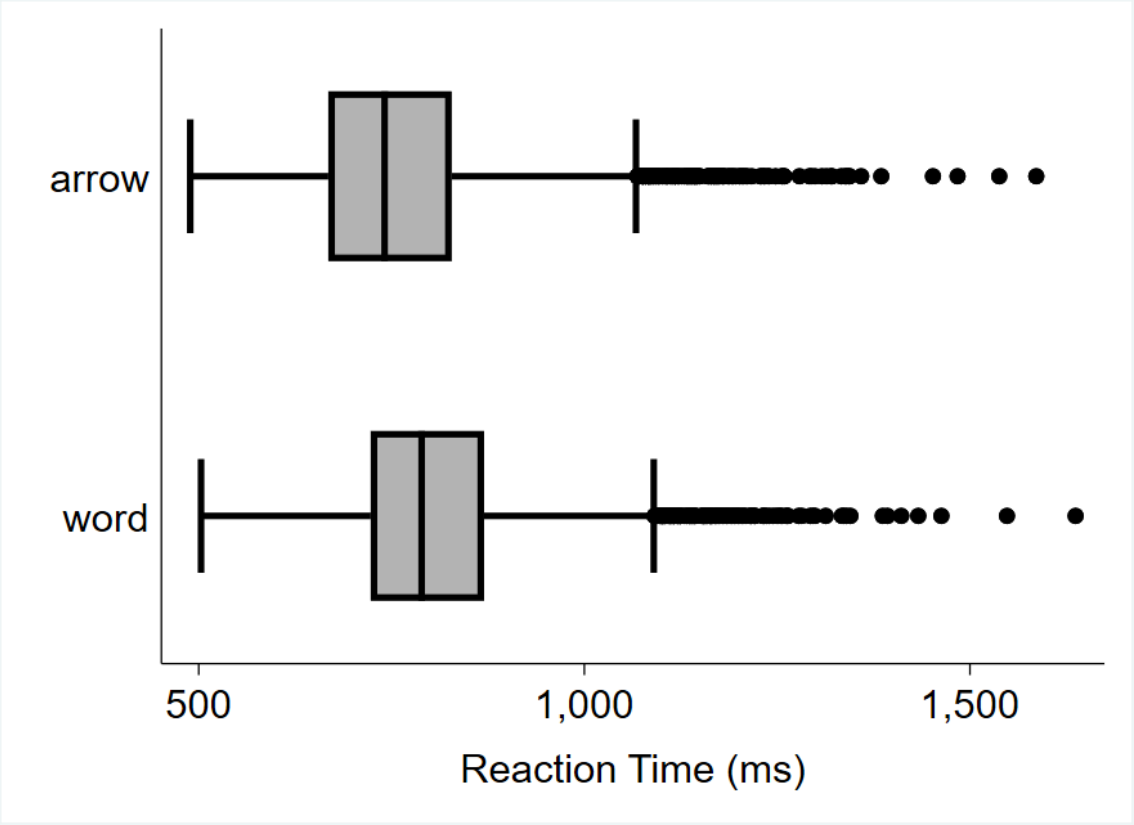


Figure 6-4. Box plot of reaction time for arrows and words on the Choice Reaction Time task

On average, males responded 1% faster than females (adjusted means 767 *vs.* 777 ms) and were 14% more likely to make errors (adjusted mean error rates 2.7% vs 2.4%), but these differences were not statistically significant (Table 6-2).

Older age at assessment was associated with slightly slower RT and fewer errors (3% slower RT, 19% less likely to make errors, per year of older age) (Table 6-2).

Higher childhood cognitive ability was associated with slightly faster responses and substantially lower odds of making errors (Table 6-2).

Education and adult socioeconomic position were not associated with RT or error rate (Table 6-2).

Participants with neurological and psychiatric conditions did not differ from cognitively-normal participants in terms of their RTs, but their odds of making an error were 54% higher (adjusted mean error rates 3.7% vs 2.5%) (Table 6-2).

6.3.2.2. Practice effects

The comparison of RT and error rate for word and arrow stimuli reported above could potentially be confounded by practice effects, as the word block was always administered after the arrow block. This could partially explain the observation of lower error rates for word stimuli, or could mean that the finding of slower RTs for word trials underestimates the true effect (see 6.3.2.1). Practice effects were explored in the two blocks separately to give an indication of whether they are a feature of performance on this test.

RT slightly decreased during the arrow block (regression coefficient = 0.996 per successive trial, *95% CIs* 0.995 to 0.997, $p < 0.0001$) suggesting a practice effect, but slightly increased during the word block (regression coefficient = 1.002, *95% CIs* 1.001 to 1.003, $p < 0.0001$). This difference could be due to the fact that the arrow block came first so perhaps participants were still getting used to the task. There was no evidence of statistically significant practice effects in error rate (Arrow block: *OR* = 1.004, *95% CIs* 0.998 to 1.094, $p = 0.063$; Word block: *OR* = 0.987, *95% CIs* 0.932 to 1.045, $p = 0.656$). Overall it does not appear that performance was strongly influenced by practice effects.

6.3.2.3. Speed-accuracy trade-offs

Slower mean RT was associated with lower error rate i.e. a speed-accuracy trade-off was observed (Spearman's $\rho = -0.28$, $p < 0.0001$) (Figure 6-5). The correlation is weak as there is a ceiling effect with many participants having an error rate of zero regardless of their mean RT. However, it is striking that all of the participants with the highest error rates had relatively fast mean RTs, which suggests that those participants were trading accuracy for speed. Also, all of the slowest responding participants had very low error rates, which suggests that they were trading speed for accuracy.

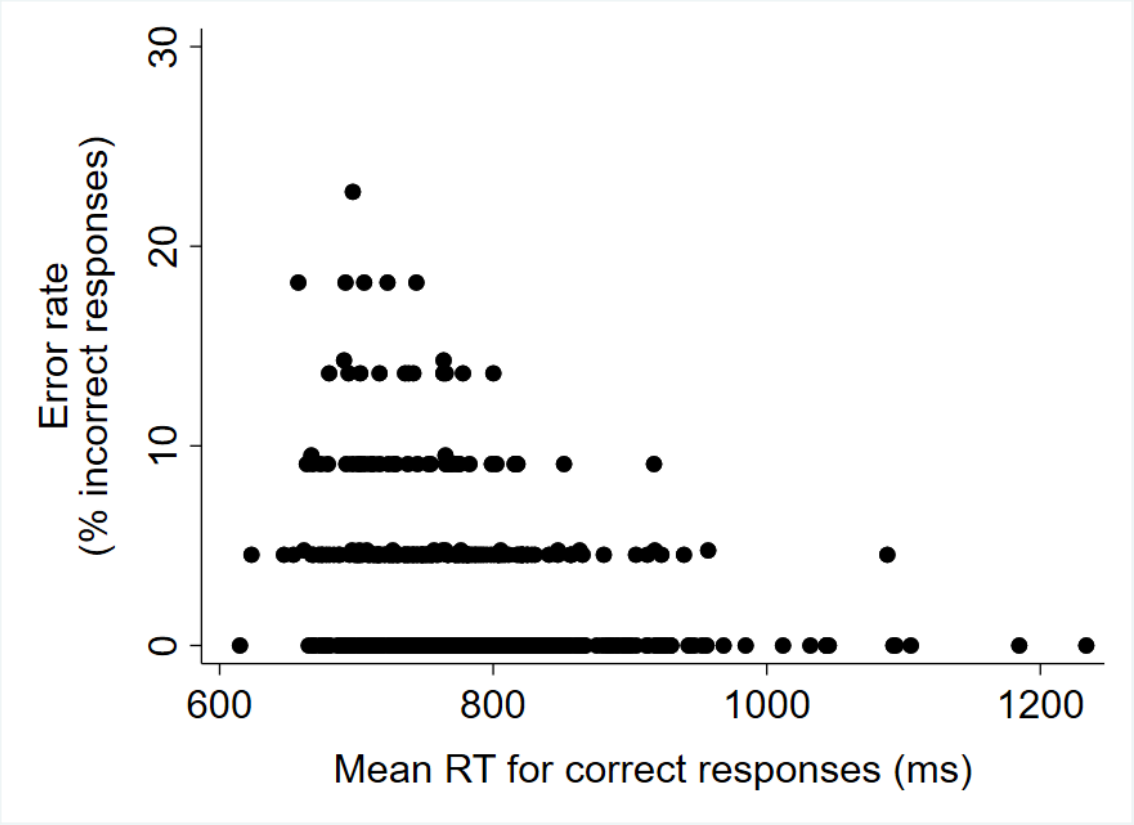


Figure 6-5. Error rate against mean reaction time on the Choice Reaction Time task

Incorrect responses were 21% faster than correct responses on average (621ms *vs.* 783 ms). The within-subjects analysis found that errors were less likely to occur with increasing RT, with a 2% reduction in the odds of making an error per additional millisecond (*OR* = 0.98, *95% CIs* 0.98 to 0.98, $p < 0.0001$). With RT included in the model, the difference in error rate between word and arrow stimuli was reversed such that word stimuli were associated with greater odds of an error (adjusted error rates: words 3.9% *vs* arrows 2.0%, *OR* = 2.16, *95% CIs* 1.68 to 2.78, $p < 0.0001$). This suggests that the earlier result of a higher error rate for arrows can be fully accounted for by speed-accuracy trade-offs, since responses to arrows were faster (see section 6.3.2.1).

6.3.2.4. Intra-individual variability in RT (for correct responses only)

Intra-individual variability (IIV) scores were approximately normally distributed across Insight 46 participants (mean = 0.126, *SD* = 0.034). Figure 6-6 illustrates how the IIV scores are derived. Results of the multivariable regression models for RT and error rate are given in Table 6-3.

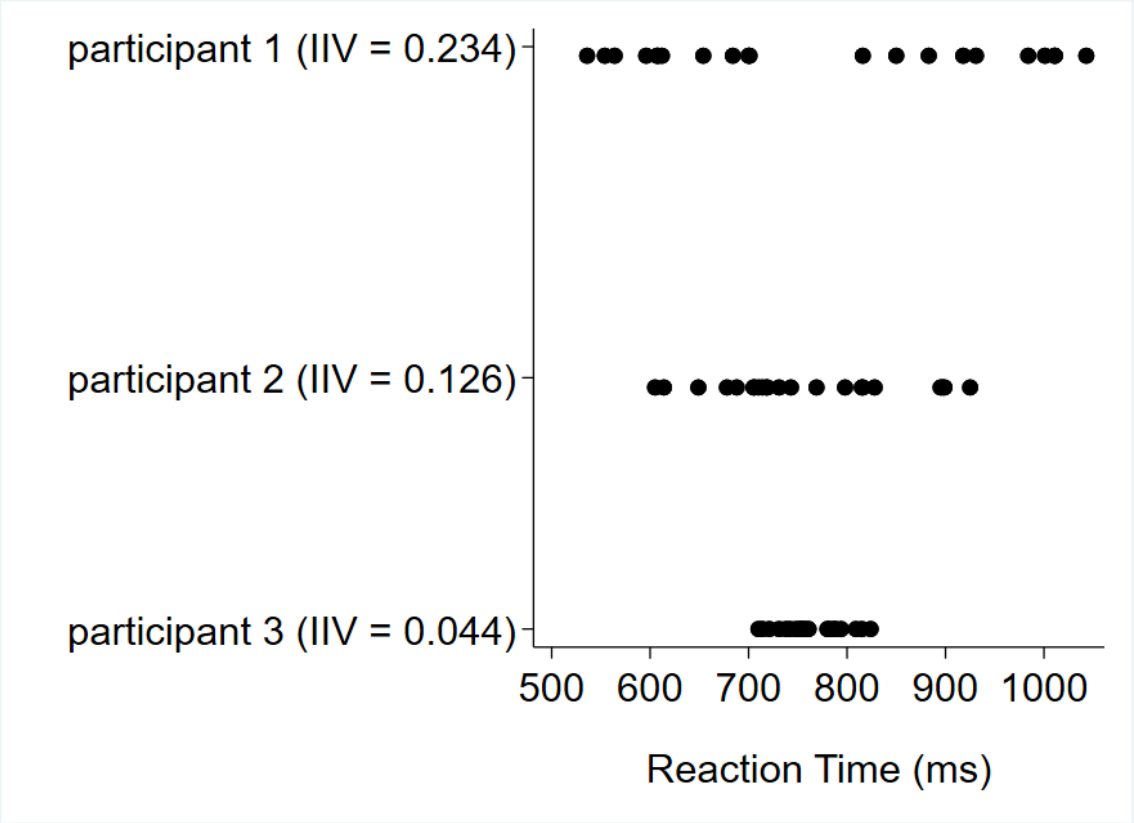


Figure 6-6. Distribution of reaction times for three participants with the maximum, median and minimum intra-individual variability scores

Each participant's reaction times to the 22 trials are plotted (correct responses only). Intra-individual variability (IIV) in RT (for correct responses only) is defined as the coefficient of variation (SD/mean). These three participants have been selected to illustrate the meaning of the IIV score and to put its magnitude in context: participant 1 has the maximum IIV, participant 2 has the median IIV and participant 3 has the minimum IIV. It can be seen that these three participants have similar mean RTs (774ms, 755ms and 763ms respectively) but the variability of their responses is very different.

Table 6-3. Associations between demographic and life-course predictors and intra-individual variability in RT on the Choice RT task (n = 501)

Predictor	Intra-individual variability in RT [a]: coefficient and 95% CIs
Sex (female as reference)	0.0046 (-0.0106, 0.0013)
Age at assessment (per year)	**0.0085*** (0.0043, 0.0128)
Childhood cognitive ability (per z-score)	-0.0038 (0.0084, 0.0007)
Education (per category) [b]	**-0.0030** (-0.0058, -0.0002)
Adult socioeconomic position (per category) [b]	0.0024 (-0.0008, 0.0057)
Neurological or psychiatric condition [c] (cognitively-normal as reference)	**0.0149*** (0.0050, 0.0257)

Coefficients in bold are significant at p < 0.05 and asterisks indicate significance at p < 0.01. [a] Intra-individual variability in RT (for correct responses only) is defined as the coefficient of variation (SD/mean). Multivariable regression models were used so each association is independent of all others. [b] See section 3.2.4 for definition of categories. [c] See section 3.2.3 for definitions. RT = reaction time.

On average, participants with neurological and psychiatric conditions had higher IIV than cognitively-normal participants (Table 6-3).

There was an association between older age at assessment and greater IIV in RT (Table 6-3, Figure 6-7).

Higher educational attainment was associated with reduced IIV (Table 6-3). The effect size was relatively small: the predicted IIVs of the lowest and highest educational categories (with all other variables held at average values) were 0.133 and 0.121 respectively, equivalent to a difference of 0.35 *SD*.

There was no evidence of an association with childhood cognitive ability, sex or adult socioeconomic position (Table 6-3).

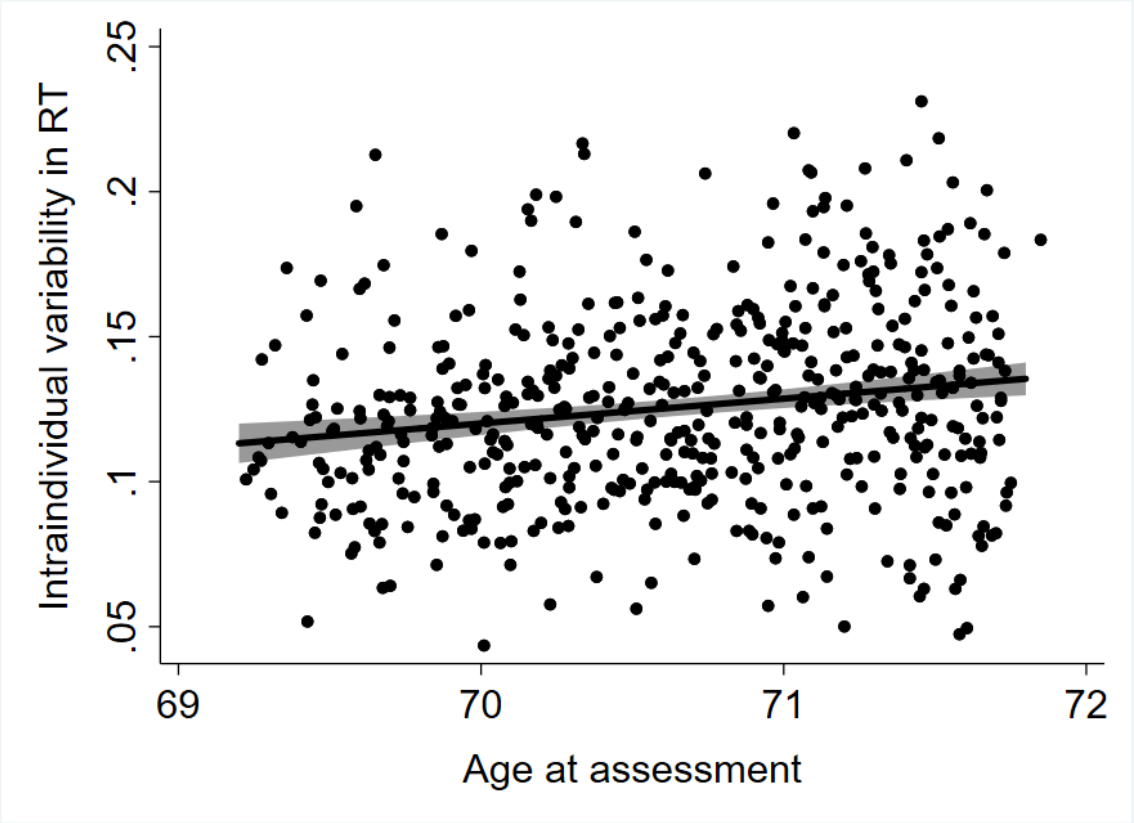


Figure 6-7. Intra-individual variability in reaction time against age at assessment

Scatter plot shows the raw data. The solid line is the predicted line of best fit from the multivariate regression model (adjusted for sex, childhood cognitive ability, education, adult socioeconomic position and presence of a neurological or psychiatric condition), and the shaded grey area represents its 95% CIs. Intra-individual variability in RT (for correct responses only) is defined as the coefficient of variation (SD/mean).

Adding error rate into the model confirmed that higher error rate was associated with greater intra-individual variability in RT for correct responses (regression coefficient = 0.0021 per percentage point increase in error rate, *95% CIs* 0.0014 to 0.0029, $p < 0.0001$). This could be due to post-error slowing, where a more cautious response speed is likely to be adopted for the next response after an error, which could make RTs more variable overall.

6.4. Associations with biomarkers and *APOE*-ε4

Following the format laid out in section 3.5, the second part of this chapter investigates associations between Choice RT performance and biomarkers of AD in cognitively-normal participants for whom complete biomarker data are available. The number of

participants meeting these criteria who also had usable data from the Choice RT task was 406 (see section 3.3).

As explained in 3.5.2.1, I wanted to derive some summary scores that capture the key aspects of performance on each task, to use for comparing results across the different cognitive tests in the Insight 46 battery (see Chapter 9). For the Choice RT task, I calculated the following three summary outcomes for each participant, combined across word and arrow stimuli:

i) **Mean RT** for correct responses
ii) **Error rate** (% incorrect responses)
iii) **Intra-individual variability in RT** for correct responses (*SD* / mean, as defined in section 6.3.1.4)

Combining outcomes across the word and arrow stimuli is justified because the magnitude of differences in RT and error rate between the two stimulus types was small, and my hypotheses about associations between AD pathology and RT (see section 6.2.3) were general rather than specific to certain types of stimuli.

For each of these three outcomes, I tested for associations with the same biomarkers as in the previous chapter (see 3.5.2). I did also run analyses using the trial-by-trial data to check whether there were any interactions with stimulus type (arrow vs. word), but there were none, so I have only reported the analyses using the summary outcomes.

6.4.1. Statistical Analyses

Mean RTs were analysed using a linear regression model. The distribution of mean RTs had a positive skew. As above (see 6.3.1.3), two transformations were considered – log-transforming the mean RTs or calculating the mean of the log-transforming raw RTs – but neither of these removed the skew so the untransformed data were used for ease of interpretation, and bootstrapping was used to produce bias-corrected and accelerated 95% CIs from 2000 replications.

Error rate was analysed using a GEE logistic regression model with an independent correlation structure and robust standard errors. The outcome was number of errors, which was treated as a proportion of the number of trials (22 for most people, but 21 for the 40 participants who had a trial excluded during the data cleaning process – see 6.2.5). Results are expressed as odds ratios for ease of interpretation.

Intra-individual variability in RT was analysed using a linear regression model as before (see section 6.3.1.4).

All models included predictors of amyloid status (positive *vs.* negative), whole brain volume, WMHV and *APOE* genotype (ε4-carrier *vs.* non-carrier). To adjust for the correlation between whole brain volume and head size, total intracranial volume (TIV) was included in all models, as were the demographic factors investigated in section 6.3 (sex, age at assessment, childhood cognitive ability, education and adult socioeconomic position).

The models were additionally rerun replacing dichotomised amyloid status with SUVR to test whether increasing Aβ deposition was associated with differences in performance. To check whether associations between SUVR and cognition were sensitive to the inclusion of the imputed SUVR values (see section 3.2.2), the analyses were rerun excluding the 26 participants with imputed SUVR data.

6.4.2. Results

Results of the regression models for the three outcomes are reported in Table 6-4. Results of interactions are reported in the text. Results for the demographic and life-course factors (sex, age at assessment, childhood cognitive ability, education and adult socioeconomic position) are not reported as they are essentially unchanged from the first analysis section (6.3.2).

Table 6-4. Associations between biomarkers and Choice RT outcomes in cognitively-normal participants (n = 406): mean RT, error rate and intra-individual variability in RT

Predictor	Mean RT (ms): coefficient (95% CIs)	Odds ratio for making an error (95% CIs)	Intra-individual variability in RT: coefficient (95% CIs)
β-amyloid status (negative as reference)	11.71 (-7.75, 30.77)	1.37 (0.93, 2.01)	**0.013*** (0.004, 0.021)
WMHV (per 10 ml)	-3.87 (-15.11, 7.59)	0.90 (0.70, 1.18)	-0.003 (-0.008, 0.003)
Whole brain volume (per 10 ml)	-1.70 (-3.59, 0.09)	1.01 (0.97, 1.04)	-0.000 (-0.001, 0.000)
***APOE*-ε4** (non-carriers as reference)	-2.25 (-17.96, 14.76)	0.94 (0.68, 1.30)	-0.004 (-0.011, 0.003)

Multivariable regression models were used so each association is independent of all others. In addition to the predictors listed, models also included sex, age at assessment, childhood cognitive ability, adult socioeconomic position and total intracranial volume.

Coefficients in bold are significant at p < 0.05 and asterisks indicate significance at p < 0.01.

CI = confidence interval; RT = reaction time; WMHV = white matter hyperintensity volume.

On average, Aβ+ participants had greater intra-individual variability in RT (adjusted means from the regression model: 0.135 vs. 0.122, a difference of 0.37 *SD*) (Table 6-4, Figure 6-8). Aβ+ participants also had slightly slower mean RTs and higher error rates than Aβ- participants, but these differences were not statistically significant (RT: Aβ+ = 789 ms; Aβ- = 777 ms; error rate: Aβ+ = 3.1%; Aβ- = 2.3% ms) (Table 6-4). Given the evidence that performance was influenced by speed-accuracy trade-offs (see section 6.3.2.3), I conducted a post-hoc analysis to investigate whether Aβ+ and Aβ- participants differed in their error rates after accounting for their mean RT, by rerunning the regression model for error rate with mean RT included as an additional co-variate. The results indicated that Aβ+ participants had 48% higher odds of making an error (95% CIs 1.03 to 2.13, $p = 0.035$, adjusted means: Aβ+ = 3.3%; Aβ- = 2.3% ms) after adjustment for their mean RTs.

When rerunning the models replacing dichotomous amyloid status with continuous SUVR, the same pattern emerged, as SUVR was associated with intra-individual variability in RT (regression coefficient = 0.062, *95% CIs* 0.014 to 0.110, $p = 0.012$) but

not with mean RT (regression coefficient = 59.0 ms, *95% CIs* -34.1 to 158.7, $p > 0.10$) or error rate (Odds ratio = 2.34, *95% CIs* 0.25 to 21.45, $p = 0.45$). These results were unchanged in a sensitivity analysis excluding the individuals with imputed SUVR values (see section 3.2.2).

Whole brain volume, WMHV and *APOE* genotype showed no evidence of associations with any of the outcomes, although there was a trend towards an association between larger whole brain volume and faster RT ($p < 0.07$) (Table 6-4).

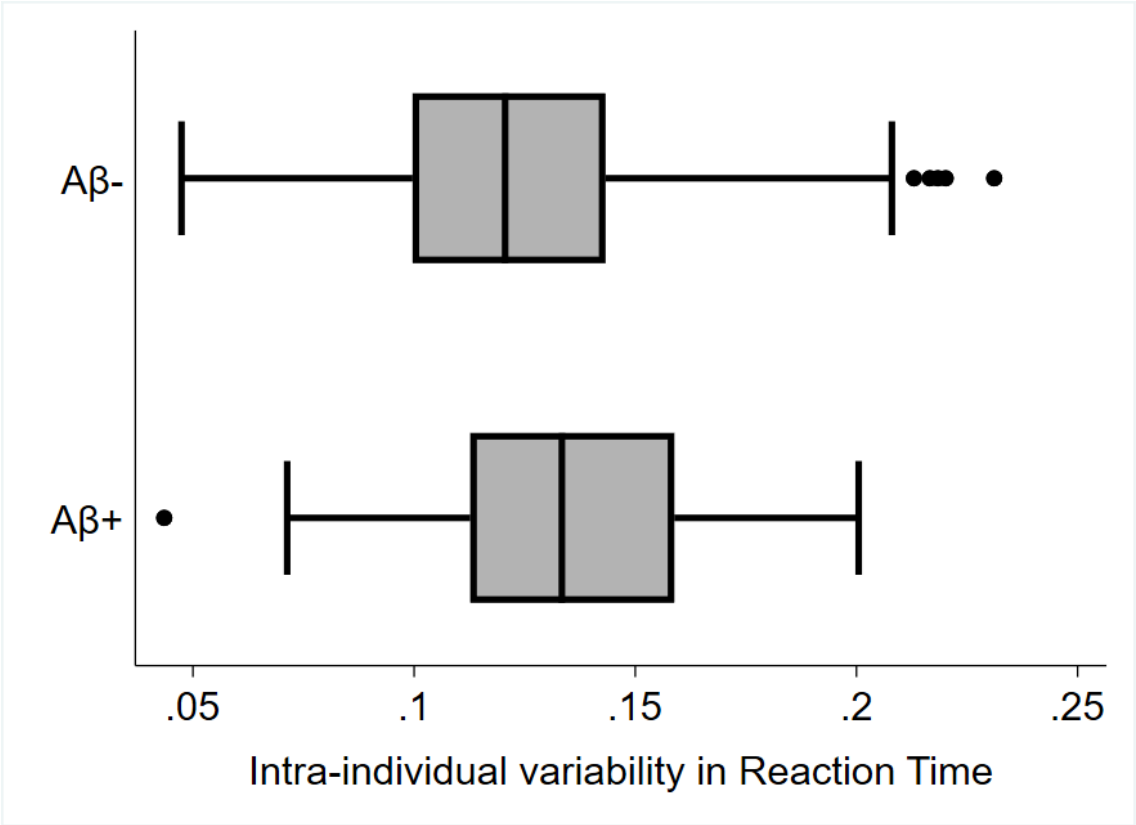


Figure 6-8. Box plots of intra-individual variability in reaction time for β-amyloid negative (n=332) and β-amyloid positive (n=74) participants

Intra-individual variability in RT (for correct responses only) is defined as the coefficient of variation (SD/mean).

6.5. Discussion

6.5.1. Summary

As hypothesised, cognitively-normal Aβ+ participants had greater intra-individual variability in Choice RT. To my knowledge, this is the first time this finding has been reported. However, there was no evidence in support of the hypothesis that Aβ positivity would be associated with slower RT. As expected, there were some effects of childhood cognitive ability and education on performance, but there was no evidence of the sex differences reported by others. Speed-accuracy trade-offs were identified as an important factor in interpreting response patterns. These results are discussed in greater detail in the following sub-sections.

6.5.2. Patterns of performance

Responses were slower to words than arrows. A possible explanation is that word stimuli place greater cognitive demands on the participant because they require higher-order processing of the concepts of left and right, whereas arrow stimuli simply require the participant to press the button on the same side as the arrow. A similar effect was found in a previous study which compared Choice RT to numbers and lights, and concluded that the slower responses to numbers could be attributed to the higher-order cognitive processing required (Nissan, Liewald and Deary, 2013). One might expect that stimuli requiring greater cognitive processing (and consequently slower RT) would also elicit a higher error rate. This was the case in the results of Nissan *et al.*, but the opposite pattern was observed in this study – a higher error rate for arrows than words. My analyses suggest that this was due to a speed-accuracy trade-off (i.e. errors occurred with hasty responses – the classic pattern in decision-making) because after adjusting for RT, the relationship was reversed such that words were associated with a higher error rate than arrows. This illustrates the importance of considering speed-accuracy trade-offs when interpreting error rates in Choice RT tasks.

The finding that incorrect responses were typically faster than correct responses suggests that some errors arose from anticipatory responses. This is consistent with the observation of a between-subjects speed-accuracy trade-off: the most error-prone participants tended to have fast mean RTs, and the slowest-responding participants tended to be highly accurate. Thus, the tendency to prioritise speed or accuracy was a relevant factor in explaining individual differences in performance, although its impact was limited by the ceiling effect on accuracy. Similar speed-accuracy trade-offs in Choice

RT tasks have been observed before (Nissan, Liewald and Deary, 2013) but other studies have reported the reverse (a correlation between speed and accuracy) (Vaportzis, Georgiou-Karistianis and Stout, 2013). I would hypothesise that the difference is most likely to be explained by the design of the experiment and the precise stimuli used, with easier tasks being more susceptible to speed-accuracy trade-offs, and more complex tasks tending to show speed-accuracy correlations (as in the following experiment, see Chapter 7). Regarding the factors which predict whether an individual tends to prioritise speed or accuracy, strategies have been reported to differ on a range of cognitive tasks by age (Brébion, 2001), sex (Reimers and Maylor, 2006; Thakkar *et al.*, 2014), risk-aversion (Nagengast, Braun and Wolpert, 2011) and mental fatigue (Rozand *et al.*, 2015). However, it should be noted that an individual may alter their speed-accuracy strategy during the task, which was not captured by the between-subjects analysis. For example, after making an error participants may shift their strategy to place an increased priority on accuracy, or after a run of correct responses they may shift their strategy to place an increased priority on speed; toggling between the two priorities is a legitimate strategy for maximising both over the course of the experiment (Dang, Figueroa and Helton, 2018).

6.5.3. Demographic and life-course predictors

6.5.3.1. Associations with age at assessment

The findings that older age was associated with slower RT, greater intra-individual variability in RT and lower error rate are consistent with the literature, where response times are widely-reported to become slower and more variable with age (Anstey *et al.*, 2005; Deary and Der, 2005; Williams *et al.*, 2005; Bugg *et al.*, 2006; Der and Deary, 2006; Dykiert *et al.*, 2012a; Garrett, Macdonald and Craik, 2012; Vincent *et al.*, 2012; Nissan, Liewald and Deary, 2013; Woods *et al.*, 2015). However, the fact that they were detected in the Insight 46 cohort was unexpected, since the age range is so narrow (69.2 to 71.8 years – reflecting the time it took to collect the data, since participants were all born in the same week). To my knowledge no studies have investigated age effects on Choice RT over such a small interval, but estimates of the rate of decline per year have been reported in cross-sectional studies: a recent comparison of large-scale Choice RT studies concluded that RT increases by 2.0-3.4 ms per year after young adulthood (Woods *et al.*, 2015), and another study of adults aged 60-85 estimated that processing speed declines by about 0.06 *SD* per year (Harrington *et al.*, 2018). A longitudinal study which tested adults repeatedly at ages 70, 73 and 76 reported a decline of -0.096 *SD*

per year (Ritchie *et al.*, 2016). The effect size reported here, which equates to an increase in RT of about 24ms per year, or 0.30 *SD*, is implausibly large in comparison.

Regarding the association between older age and lower error rate, this is consistent with some previous reports that older people tend to reduce errors at the expense of speed (i.e. they adopt a more cautious balance on the speed-accuracy trade-off) (Brébion, 2001)) but other studies contradict this (Vaportzis, Georgiou-Karistianis and Stout, 2014; Woods *et al.*, 2015), and it may depend greatly on the precise characteristics of the task.

As mentioned in Chapter 4 in the context of the apparent age effect on the Matrix Reasoning task, while it is conceivable that a genuine subtle effect of age may be detectable over a 2.6-year range, especially when carefully controlling for other factors that affect performance such as childhood cognitive ability and education, I considered the possibility of a recruitment bias whereby participants seen towards the beginning of the data collection period may have differed in some ways to those seen towards to the end. This is discussed in greater detail in Chapter 10.

6.5.3.2. Associations with childhood cognitive ability, education and adult socioeconomic position

Consistent with the literature on the relationship between RT and intelligence, higher childhood cognitive ability was associated with faster RT, but the effect size was small. Higher childhood cognitive ability was also associated with lower error rate, even though error rate did not vary much due to the ceiling effect on accuracy. Following on from my earlier conclusion that errors were likely to be anticipatory or hasty responses (section 6.5.2), this may suggest a link between higher childhood cognitive ability and inhibition, which was assessed more directly in a related experiment presented in Chapter 7. Contrary to hypotheses, childhood cognitive ability was not associated with intra-individual variability in RT. Conversely, educational attainment had no independent effect on RT and error rate, but was associated with reduced intra-individual variability in RT – an effect which is consistent with a previous report (Christensen *et al.*, 2005) but was modest in magnitude and unlikely to be clinically meaningful. These findings are broadly consistent with the conclusion of Der and Deary (2017) that intra-individual variability in RT shows little association with mental ability when its dependence on mean RT is accounted for, as was the case in my analyses.

Overall, the effects of childhood cognitive ability and education on this task were relatively small when compared to their effects on the standard cognitive tests discussed in the previous chapter. The fact that adult socioeconomic position was not a significant

predictor of performance on any outcome is not unexpected in light of previous analyses of predictors of cognition across adulthood in the NSHD, which have found adult socioeconomic position to be less important than childhood cognitive ability and education (Richards and Sacker, 2003; Richards *et al.*, 2019).

6.5.3.3. Sex differences

Contrary to some previous studies (Deary and Der, 2005; Der and Deary, 2006; Dykiert *et al.*, 2012b; Vincent *et al.*, 2012; Phillips *et al.*, 2013), in this study, which benefits from a large sample size, very narrow age range, and adjustment for confounding variables of prior cognitive ability, there was no evidence of sex differences on any outcome.

6.5.4. Associations with biomarkers and APOE-ε4

The second part of the analyses investigated associations between performance on the key summary outcomes of the Choice RT task and various neuroimaging and genetic biomarkers of AD pathology. The hypothesis that Aβ+ participants would have greater intra-individual variability in RT than Aβ- participants was supported, but there was no evidence for the hypothesis that their RT would be slower overall.

To my knowledge, the literature on changes in response time and variability preceding the onset of AD dementia comes from investigation of individuals with MCI (Christensen *et al.*, 2005; Tales *et al.*, 2012; Phillips *et al.*, 2013). As the Aβ+ participants in this analysis were all cognitively-normal and on average perhaps a decade or more away from the time when a significant proportion of them may be expected to develop dementia, this suggests the possibility that intra-individual variability in RT may be a particularly sensitive measure to accumulating AD pathology, whereas the slowing of mean RT may perhaps occur at a later stage.

It is important to remember that changes in response speed and variability are by no means specific to AD. These measures appear to be general markers of the integrity of the central nervous system, vulnerable to disruption across a range of neurodegenerative diseases and brain injuries, and correlating with diverse physical health measures such as grip strength, forced expiratory volume and waist circumference (Anstey *et al.*, 2005; MacDonald, Nyberg and Bäckman, 2006; Masi *et al.*, 2018). Further work is needed to investigate how changes in RT associated with AD pathology may relate to such physical health measures and to other brain pathologies that accumulate with age. One study comparing intra-individual variability of RT in patients with AD and Parkinson's disease reported that the AD patients were more

variable, controlling for overall severity of cognitive impairment (Burton *et al.*, 2006), but more research is needed to determine whether there is a pattern to the timing and nature of changes in RT (both response speed and variability) as AD pathology accumulates.

There was a suggestion that Aβ pathology was associated with higher error rate on this task, as Aβ+ participants made more errors than Aβ- participants and this difference was statistically significant after adjustment for the trade-off between RT and error rate. As Aβ+ participants had slightly slower RTs on average (although this difference was not statistically significant), I would speculate that they might have been less likely to make errors due to responding in haste and more likely to make errors for other reasons, such as failures in the executive functions of monitoring and inhibition. However, this experiment was limited in its ability to measure subtle differences in accuracy due to the small number of trials, as discussed in the following section.

No evidence was found of statistically significant associations between outcomes on this task and whole brain volume, white matter hyperintensity volume (WMHV) or *APOE*-ε4 (accounting for amyloid status). In the previous chapter I reported the finding that Insight 46 participants with smaller whole brain volumes and larger WMHV showed evidence of slightly slower perceptual processing speeds on the Digit-Symbol task (see Chapter 4), and discussed why the Digit-Symbol test may be particularly sensitive to overall brain health. As the detrimental effects of white matter disease on processing speed are well established (Prins and Scheltens, 2015) and compromised white matter integrity has been reported to predict slower and more variable RT (Fjell *et al.*, 2011; Jackson *et al.*, 2012), the lack of an association between WMHV and Choice RT is probably due to the fact that the Insight 46 participants on the whole had very low levels of white matter disease.

In summary, this study builds on the results of the previous chapter by providing further evidence that subtle differences in cognition can be detected cross-sectionally between cognitively-normal Aβ+ and Aβ- individuals in a large well-characterised sample, and that such differences are present in multiple cognitive domains.

6.5.5. Strengths and limitations

A strength of the experimental design is that RTs were collected for every response, in contrast to some RT experiments which only collect the mean RT, *SD* and error rate (for example the numbers-based device that has been widely used (e.g. (Der and Deary, 2006)) including in the NSHD cohort when they were aged 60-64 (Silverwood *et al.*, 2014). By collecting data on individual responses, we were able to exclude extreme

responses thus reducing noise in the data, as well as being able to investigate within-subject speed-accuracy trade-offs by comparing the RTs of correct and incorrect responses. Without this information it would have been difficult to interpret the fact that arrow stimuli had a higher error rate than word stimuli, which turned out to be explained by a speed-accuracy trade-off (see section 6.5.2).

There were a number of limitations to the experimental design, the main one being the small number of trials (22). This had a particular impact on the error rate measure, as a single incorrect response corresponded to a large proportionate increase in error rate. The fact that odds of making an error were associated with childhood cognitive ability, age at assessment and presence of a neurological or psychiatric condition – despite the ceiling effect on accuracy – suggest that this measure may capture something meaningful about performance on this task, rather than just signifying random mistakes. The fact that errors tended to be faster than correct responses suggests that impulsivity or a lack of inhibition may be playing a role. The numbers of participants with each neurological or psychiatric diagnosis are too small to compare between them (see section 3.6), but some of the conditions have motor symptoms (e.g. Parkinson's disease and multiple sclerosis) so motor errors could have been a contributing factor. The apparent difference in accuracy between cognitively-normal Aβ+ and Aβ- participants would be interesting to investigate further. The best strategy for replicating these results would be to repeat the experiment with a much greater number of trials. Any attempt to reduce the ceiling effect on accuracy by making the task more difficult (e.g. by changing the number of response options from two to three) could disrupt the phenomena being studied, since the ceiling effect is integral to the type of errors that are being captured (theoretically the only requirement for accuracy is to know one's left from one's right so errors are likely to be due to anticipations or occasional lapses of attention).

It is worth noting that many Choice RT studies place little focus on error rates – and sometimes they are not even reported – since the interest is naturally in investigating RT, which is typically only analysed for correct responses. To study RT in as pure a form as possible, some studies exclude participants with high error rates, for example Der and Deary (2006) excluded participants with error rates of 20% or higher, noting that their performance *"suggests problems in correctly carrying out the task"*. However, for studying brain health in older age it might be the case that errors on Choice RT tasks could be a relevant indicator.

There were a number of other limitations to the experimental design. Below is a list of recommendations (in no particular order) for how I think the task should ideally be altered to address these limitations.

i) The first trial of each block began automatically after a pause of about 5 seconds, following the practice trials. Although participants were warned about this, they were not always ready. This resulted in me having to exclude the first trial of each block. A better experimental design would be for the tester or participant to initiate the main block by pressing a button when ready, or to designate the first few trials of the block as warm-up trials, as in Aron *et al.* (2004).

ii) Button presses of longer than a certain duration were registered as multiple presses. Although participants were instructed to press the button quickly, some had a tendency to hold the button down (e.g. for 500ms). If this happened for a correct response, it did not matter because the next trial would be immediately initiated after the first button press had registered, and any "extra" presses would fall during the cue of the next trial, when they would not be registered. However, if a participant held the button down for too long on an incorrect response, this meant that they would not have the chance to correct their error since their incorrect response registered as both the first and second attempt. Ideally a single button press should be registered until the button is released, however long that takes.

iii) When the instructions were on the screen at the start of each block, it was possible for the participant to accidentally initiate the block by pressing one of the response buttons before the tester had finished explaining the task. Occasionally this happened when the participant held down the button too long when responding to the final stimulus of the previous block and it registered as a double-press (see point ii). Ideally an extra blank screen should be added before the instructions screen, and it should be impossible to initiate the block using the response buttons (e.g. by assigning a key on the keyboard instead).

Strengths and limitations that apply to all the analyses presented in this thesis, such as considerations relating to the generalisability of the sample, are discussed in Chapter 10.

6.5.6. Conclusions

In summary, this short Choice Reaction Time experiment has provided novel evidence that cognitively-normal older adults with Aβ pathology are, on average, less consistent in their response times. This adds to previous literature that has shown similar deficits in

individuals with MCI and AD, and suggests that this may be a sensitive measure of early cognitive changes that could be included in future studies and clinical trials targeting individuals with preclinical AD.

7. RESPONSE INHIBITION

7.1. Introduction

Executive function is a term that describes the higher-order cognitive processes – controlled by the prefrontal cortex – which are necessary for problem-solving, decision-making, planning and focusing attention in an effortful manner to accomplish tasks (Otero and Barker, 2014). A key aspect of executive function is inhibitory control, which has both behavioural and cognitive dimensions. Behavioural inhibition describes the ability to resist impulsive actions (self-control), whereas cognitive inhibition describes the ability to selectively focus attention on stimuli or thought processes that are relevant for the current goal and to resist interference from irrelevant stimuli or thought processes that would interfere with achieving the goal (Diamond, 2013). Often this involves inhibiting automatic responses, which are habitual or "overlearned". Inhibitory control is a vital skill for all aspects of life including health, social interactions and employment.

Inhibitory control has been widely researched using **response inhibition** tasks, which create a conflict between an automatic (but incorrect) response and the correct response. An effective way of creating this conflict is to use incongruent stimuli, such as in the Stroop task (see MacLeod (1991) for a review) or the Simon task (Simon, 1969). In the Stroop task participants are presented with incongruent stimuli (e.g. the word "blue" printed in red-coloured ink) and congruent stimuli (e.g. the word "blue" printed in blue-coloured ink) and asked to name the colour of the ink. Responses to incongruent stimuli are typically slower and less accurate than responses to congruent stimuli. In a computerised version of the Simon task (e.g. Stoet, 2017), arrows are presented pointing left or right, on either the left-hand or right-hand side of the screen. If the direction of the arrow is incongruent with its position on the screen (e.g. arrow pointing left on the right-hand side of the screen), participants are generally slower and less accurate at identifying the direction of the arrow. There are other methods of creating conflict between an automatic response and the required response; for example, the Stop-signal or Go/No-Go tasks (see Rey-Mermet and Gade, 2018) require participants to repeatedly press a button in response to a stimulus, except on rare occasions when a "stop signal" is presented. The response is difficult to inhibit because it has become automatic through repetition. The Insight 46 cognitive battery contained a response inhibition task with congruent and incongruent stimuli, somewhat akin to the Simon task.

Impaired inhibitory control can be a feature of many neurodevelopmental and psychiatric conditions affecting the frontal lobes, including attention deficit hyperactivity disorder,

autism, schizophrenia, obsessive compulsive disorder and addictions, and is also seen subsequent to frontal lobe damage due to traumatic brain injury or neurodegenerative diseases including the dementias (Goldstein and Naglieri, 2014). It is a prominent and early feature of frontotemporal dementia – a spectrum of neurodegenerative diseases where the frontal and temporal lobes are subject to focal atrophy (O'Callaghan, Hodges and Hornberger, 2013). Although it is not generally considered to be a defining symptom of typical AD, a recent meta-analysis estimated that 17% of patients with AD show disinhibition, as rated by a care-giver on the Neuropsychiatric Inventory (Zhao *et al.*, 2016). There is some evidence that patients with AD show impairments on the response inhibition tasks described above: patients with mild AD have been reported to respond disproportionately slowly and inaccurately to incongruent stimuli on the Stroop task (Bélanger, Belleville and Gauthier, 2010) and the Simon task (Castel *et al.*, 2007), as well as a similar interference task (Aschenbrenner *et al.*, 2015). There is also some evidence to suggest that such impairments in inhibitory control may emerge prior to the onset of AD dementia, as individuals with MCI have been reported to show increased interference from incongruent stimuli compared to healthy older adults (Bélanger, Belleville and Gauthier, 2010).

As to whether impairments in inhibitory control may emerge even earlier – in the preclinical stage of AD when individuals are cognitively-normal but have accumulating AD pathology – there has been little research to date. While there is growing evidence that subtle changes in executive function can be detected in the preclinical stage of AD and can predict progression to dementia diagnosis (i.e. in cognitively-normal individuals who show biomarker evidence of AD pathology) (Grober *et al.*, 2008; Clark *et al.*, 2012; Hassenstab *et al.*, 2015; Baker *et al.*, 2017; Duke Han *et al.*, 2017; Mortamais *et al.*, 2017), executive function is a broad term with no consistent definition. The conclusions of these studies are based on diverse cognitive tests including processing speed (e.g. Digit-Symbol test, see Chapter 4), verbal fluency (e.g. generating items within a category such as animals or vegetables), cognitive flexibility (e.g. the Trail Making Test) and working memory (e.g. Digit Span). A small number of studies have specifically examined associations between inhibitory control and biomarkers of AD pathology in cognitively-normal individuals. One such study reported that individuals with pathological β-amyloid$_{42}$/tau ratios (based on CSF sampling) tended to show greater interference on the Stroop task than controls (Harrington *et al.*, 2013) and another study observed that interference on a different response inhibition task was correlated with quantity of β-amyloid plaques (as measured by PET) and CSF levels of β-amyloid$_{42}$ and tau (Aschenbrenner *et al.*, 2015).

As I have argued in previous chapters, when investigating potential differences in inhibitory control that may be associated with AD pathology, it is important to account for other factors which may contribute to individual differences such as age effects, sex and prior cognitive ability. Factors associated with individual differences in inhibitory control are briefly commented on below.

A recent meta-analysis of 176 studies comparing older and younger adults on response inhibition tasks (including the Stroop, Simon and stop-signal tasks) concluded that the prevailing hypothesis of a deficit in inhibitory control in older age was generally not supported, although there was some variation between tasks (Rey-Mermet and Gade, 2018).

In terms of sex differences on response inhibition tasks, the evidence is limited as most studies adjust for sex in their analyses without reporting whether it was associated with the outcomes. Sex differences are generally not observed on the Stroop task (MacLeod, 1991). A recent review concluded that women may show greater interference on the Simon task (i.e. greater slowing of responses to incongruent stimuli), but the evidence is mixed (Stoet, 2017).

As with many cognitive tests, there is some evidence of a relationship between education and performance on response inhibition tests, although it does not seem to have been widely investigated: one study reported an association between greater education and reduced interference (Aschenbrenner *et al.*, 2015), and another study found that increased interference on the Stroop task was correlated with lower verbal IQ in older adults (Puccioni and Vallesi, 2012).

As for the Choice Reaction Time task reported in the previous chapter, it is important to consider the relationship between speed and accuracy on response inhibition tasks, rather than only considering them as two separate outcomes. Speed-accuracy trade-offs have been reported on “stop-signal” tasks (Dang, Figueroa and Helton, 2018), so the speed of an individual’s response may dictate their ability to suppress interference from irrelevant properties of the stimulus.

Following the structure of the previous chapter, the aims of this study were firstly to understand patterns of performance on the Response Inhibition task and characterise associations between task performance and demographic and life-course predictors, and secondly to investigate associations between performance and biomarkers of brain pathologies among cognitively-normal participants.

7.2. Methods

7.2.1. Stimuli and Procedure

Two related experiments were combined into one task. The first experiment – a Choice Reaction Time task – was described in the previous chapter. This chapter concerns the second experiment which was a response inhibition task.

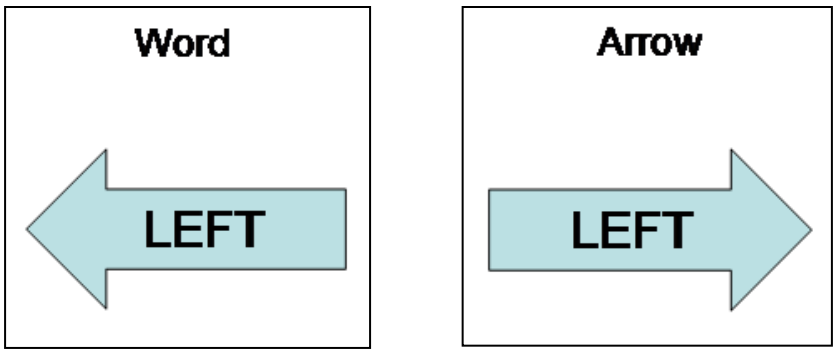
The response inhibition task was presented on the same DELL computer with the same 2-button response box (see section 6.2.1). All stimuli were displayed in the centre of the screen and consisted of an arrow pointing left or right with the word 'LEFT' or 'RIGHT' inside it. In 50% of trials, the arrow and the word were congruent (e.g. an arrow pointing left with the word 'LEFT' inside it) and in the other 50% of trials the arrow and the word were incongruent (e.g. an arrow pointing right with the word 'LEFT' inside it) (Figure 7-1). Each stimulus was preceded by a cue indicating whether the participant should respond to the arrow or to the word.

There were 2 blocks of 48 trials each. In Block 1, the cue duration was 200ms (i.e. the cue was displayed on its own for 200ms before the stimulus appeared below it). In Block 2 the cue duration was 1500ms. Throughout both blocks, the order of the trials was A-A-A-W-W-W-A-A-A-W-W-W etc., where A indicates an arrow cue and W indicates a word cue. Thus, there were equal numbers of arrow cue and word cue trials. Within each A-A-A and W-W-W sequence the trials are referred to as being in run-positions 1, 2 and 3. Within each block there were equal numbers of trials in each of the four possible combinations of congruency and cue type (congruent arrow, incongruent arrow, congruent word, incongruent word). However, these factors were not perfectly counterbalanced across each run-position; the number of trials in each combination of *run-position x cue type x congruency* varied from 2 to 6 within each block, because the order was random with respect to congruency.

The order of trials was the same for all participants. Six practice trials preceded each block. The practice trials were followed by a gap of about 5 seconds before the main block started automatically.

As in the Choice RT task (see Chapter 7), participants were instructed to press the correct button on the response box (left or right) as quickly as possible, using the index and middle fingers of their dominant hand. If their response was correct, the next trial was initiated; if their response was incorrect, the stimulus remained on the screen and an error tone sounded to signal that they should respond again as quickly as possible.

Regardless of whether their second response was correct or incorrect, no feedback was given and the next trial was initiated.



Congruent trial Incongruent trial

Figure 7-1. The Response Inhibition task

The cue ("Arrow" or "Word") appeared on its own for 200ms (Block 1) or 1500ms (Block 2) before the stimulus appeared underneath.

7.2.2. Outcome variables

Two variables were recorded for each response: **RT (ms)** and **accuracy (correct or incorrect)**. All analyses were based on the initial response to each stimulus; second attempts were not analysed as their purpose was to reorient participants to the task.

As explained in section 6.2.2, standard practice on tasks measuring RT is to analyse RT for correct responses only. This approach was adopted here.

In the Choice RT task, an outcome of interest was intra-individual variability in RT. This would not be a very meaningful measure in the Response Inhibition task because there are large differences in RT depending on congruency, cue duration and cue type. A variability measure would be heavily influenced by the extent to which participants were affected by these factors, rather than capturing their variability in executing a repeated response. I considered calculating intra-individual variability for each combination of these factors separately, but there are not many trials in each and it is not the primary interest of this experiment. The primary interest is the ability to inhibit automatic responses to incongruent stimuli.

7.2.3. Exploration of task-set switching

This experiment is based on the paradigm used by Aron *et al.* (2004) which was designed to measure task-set switching – the ability to switch backwards and forwards between two different sets of instructions, or "task-sets", within the same experiment. As in our experiment, they ordered the trials in groups of three (A-A-A-W-W-W-A-A-A-W-W-W etc. where A indicates an arrow cue and W indicates a word cue) and their aim was to examine task-set switching by comparing responses in run-positions 1, 2 and 3. They assessed whether responses were slower and less accurate to trials in run-position 1 (i.e. trials where the cue was different to the previous trial, requiring a switch of task-set) compared to run-positions 2 and 3 (where the cue was the same as the previous trial). I was not involved in the design of the experiment for Insight 46, and when I came to analyse it, I came to the conclusion that it is not suitable for addressing the question of task-set switching because the experimental design differs from that of Aron *et al.* in several important ways.

Firstly, in order to accurately assess task-set switching, trials should be excluded from analysis if the participant's response to the preceding trial was incorrect, because their incorrect response may be due to adopting the wrong task-set. For example, if a trial in run-position 3 is an incongruent stimulus with the cue 'arrow', but the participant responds to the word (thus making an error), then the following trial (a trial in run-position 1 with the cue 'word') would not actually require the participant to switch their task-set. Aron et al. excluded trials following an error on either of the preceding two trials for this reason. This was feasible because their experiment contained a large number of trials (288). This experiment only contained 96 trials due to time constraints, but contained a requirement for participants to make a second response if they received feedback that their first attempt was incorrect – a feature not present in the experiment of Aron *et al.* – which arguably functions as an alternative mechanism to ensure that participants were reoriented to the correct task-set before the next trial appeared. However, on reflection I do not think this modification is sufficient for the following reasons:

i) To correct an incorrect response, all that is required is to press whichever button was not pressed on the first attempt. This may be cognitively different to having to process the cue and stimulus, so it cannot be assumed that correcting an error equates to being reoriented to the task.
ii) Some participants may have got the second attempt wrong due to perseverating with the incorrect task-set despite the feedback that their first attempt was incorrect – anecdotally I believe I sometimes observed this happening when

administering the test. As no feedback was given for the second attempt, this would not be drawn to their attention.

iii) As discussed in section 6.5.5, one limitation of the response box used for the Choice RT and Response Inhibition experiments was that button presses of longer than a certain duration registered as multiple presses. Although this issue only affected a minority of participants, it reduces confidence in the effectiveness of the second attempt for reorienting participants to the correct task-set.

The second main reason that this experiment does not address task-switching in a comparable manor to Aron et al. is that their format made the A-A-A-W-W-W pattern obvious to participants by the position of the stimuli on the screen (see Figure 7-2) as run-positions 1, 2 and 3 appeared in different locations on the screen, moving round in a circle. This allowed the authors to conclude that task-switching was associated with a cost (slower and less accurate responses) even when each upcoming switch could be anticipated. In contrast, the design of the experiment used in Insight 46 meant that the A-A-A-W-W-W pattern was not made obvious to participants and, based on anecdotal observations, I suspect that few of them spotted it, although there may have been some implicit learning. Therefore, participants may have been more likely to respond to each trial with the same degree of surprise rather than allowing the task-set to become established by knowing when the switches were due.

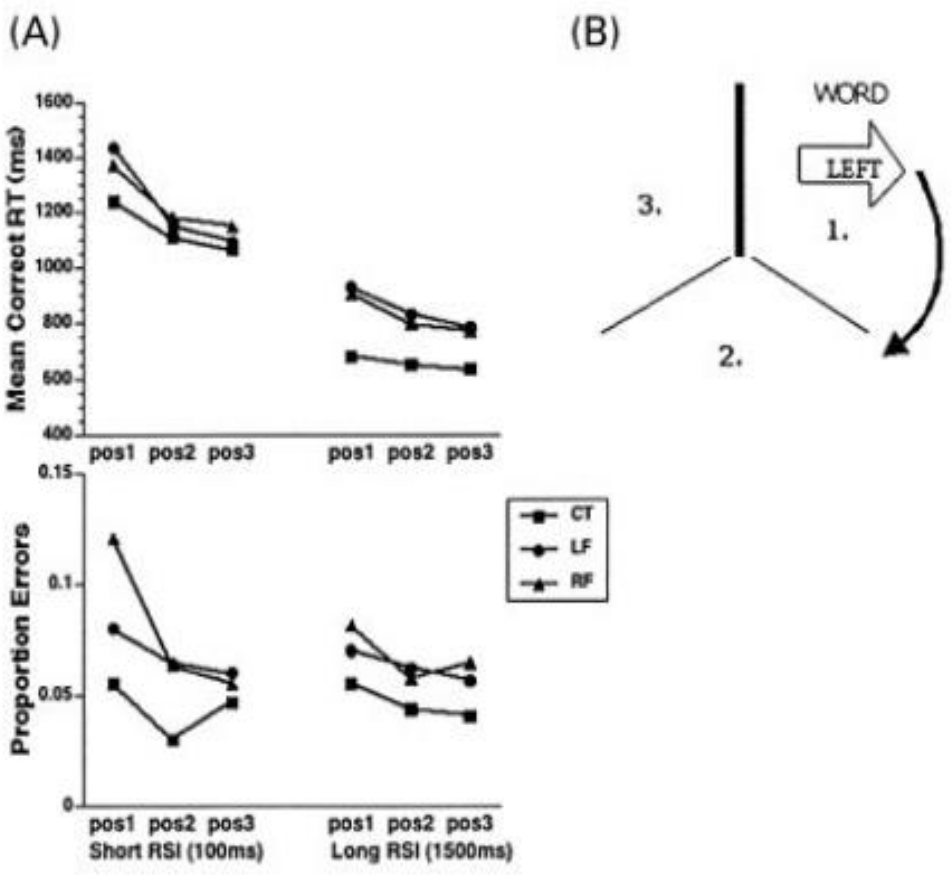


Figure 7-2. Run-position effect and task design in the experiment used by Aron *et al.*

This figure is reprinted from Aron et al. (2004) with permission from Oxford University Press. (A) shows that run-position 1 elicited significantly slower and less accurate responses, indicating the presence of a task-set shifting effect. (B) shows that stimuli in the three run-positions were presented in different places on the screen, making it obvious when the shift of task-set would occur.
CT = controls; LF = left frontal lesion patients; RF = right frontal lesion patients; RSI = duration of cue before stimulus appeared; pos = run-position.

Thirdly, in the original study, the simple arrow and word trials (which the Insight 46 Choice RT experiment was based on – see Chapter 6) were not administered in a separate experiment but were mixed in with the congruent and incongruent trials, and were referred to as 'neutral' trials. By comparing neutral and congruent trials the authors were able to address a specific question about the processes involved in task-set switching because neither neutral nor congruent trials trigger any competing response which needs to be inhibited, yet there is typically a difference in RT between the two, with responses to congruent trials being slower than responses to neutral trials. As congruent stimuli contain elements that are associated with the irrelevant task-set (e.g. the presence of a word when the relevant task-set is 'arrow'), this slowing of responses compared to

neutral trials can only be attributed to competition arising from endogenous activation of the irrelevant task-set. This is different to the activation of the irrelevant task-set that can be triggered exogenously by incongruent stimuli. Therefore, the original study was able to analyse both the endogenous and exogenous aspects of task-set shifting, whereas our experiment does not permit investigation of the endogenous aspect because of the lack of 'neutral' trials.

In summary, I concluded that the Insight 46 Response Inhibition task could not provide a valid measure of task-set shifting. However, the fact that the trials were ordered in a pattern of groups of three (A-A-A-W-W-W, where A indicates an arrow cue and W indicates a word cue) meant that the influence of run-position on the results still needed to be considered. My methods for considering this are outlined in 7.3.1.1.

7.2.4. Hypotheses

For the purpose of assessing inhibitory control, the outcomes of interest were RT and error rate on incongruent trials, compared to congruent trials.

Based on the literature, I hypothesised that higher childhood cognitive ability and educational attainment would be associated with greater inhibitory control i.e. the tendency to respond slower and less accurately on incongruent trials would be reduced.

I aimed to test the hypothesis that cognitively-normal Aβ+ participants would show reduced inhibitory control compared to Aβ- participants.

7.2.5. Participants

All 502 participants completed the Response Inhibition task, but one participant's data-file was not successfully saved (see section 3.3). Participant characteristics are reported in section 3.6.

7.2.6. Data processing

As for the Choice RT task, the first trial in each block was excluded from analyses because participants did not always realise that the block had started and this affected RT (see sections 6.2.5 and 6.5.5).

In line with the policy of keeping the sample as representative as possible, the only reason to exclude participants with outlying performance was deemed to be a clear indication that they deviated from the protocol e.g. a fundamental misunderstanding of the instructions. As mentioned in the previous chapter (see section 6.2.5), one participant had a lifelong condition affecting cognition and was noted to have difficulty in understanding the task, so this participant was excluded from the analyses, reducing the sample size to 500.

As explained in the previous chapter (see section 6.2.5), an error rate of greater than 50% in a 2-choice task would normally be suggestive of guessing at chance level, and this would be grounds to consider excluding participants from the analyses. However, this task contains incongruent trials where it is necessary for participants to inhibit incorrect responses which are prompted by the non-cued property of the stimulus. Therefore, an error rate of greater than 50% on incongruent trials would not necessarily be suggestive of guessing at chance level – it could indicate a failure to inhibit incorrect responses despite understanding the instructions. Anecdotally, from observing participants completing this task, I concluded that this was the case, so it would not be appropriate to exclude participants with error rates of greater than 50%. In fact, although no participants had an overall error rate of greater than 50%, 12 participants had an error rate of greater than 50% on incongruent trials alone (max 67%). Of these, three met criteria for a neurological or psychiatric condition and nine were classified as cognitively-normal (see definitions in section 3.2.3)

As in the previous chapter, I have considered both correct and incorrect responses for the purpose of checking for outlying RTs. The distribution of each participant's mean RT shows that there a few participants with outlying slow RTs (Figure 7-3). Of those with a mean RT greater than 4000 ms, one met criteria for MCI (see section 3.2.3) and the other three were cognitively-normal. There was no evidence that any participants did not complete the task according to the instructions, so none were excluded based on their mean RT.

Reaction times to individual trials were checked for outliers. As in the previous chapter (see section 6.2.5), I adopted a threshold of 300 ms as the fastest possible time for a valid response. Four participants had one response each excluded for being faster than

300ms. There were a few trials with extremely slow outlying RTs, up to 40 seconds (Figure 7-4). It is clear that some of these should be excluded as they are outside a reasonable time frame and correspond to notes made by the testers about deviations from the protocol; for example, I observed instances of participants stopping to cough or drink water. As in the previous chapter (see section 6.2.5) I excluded trials where the RT was more than 3 *SD* above each participant's own mean. This applied to 558 trials (1.2%) from 379 participants, with between 1-4 trials each. After these exclusions, the slowest response was 13755 ms. Even though slower responses are more likely to occur in certain conditions (e.g. incongruent stimuli, short cue duration – see results section 7.3.2.2.1) I concluded that it was fair to exclude this small minority of trials as a way of reducing noise in the data, especially as there are not enough trials in the different conditions to define outliers within each one.

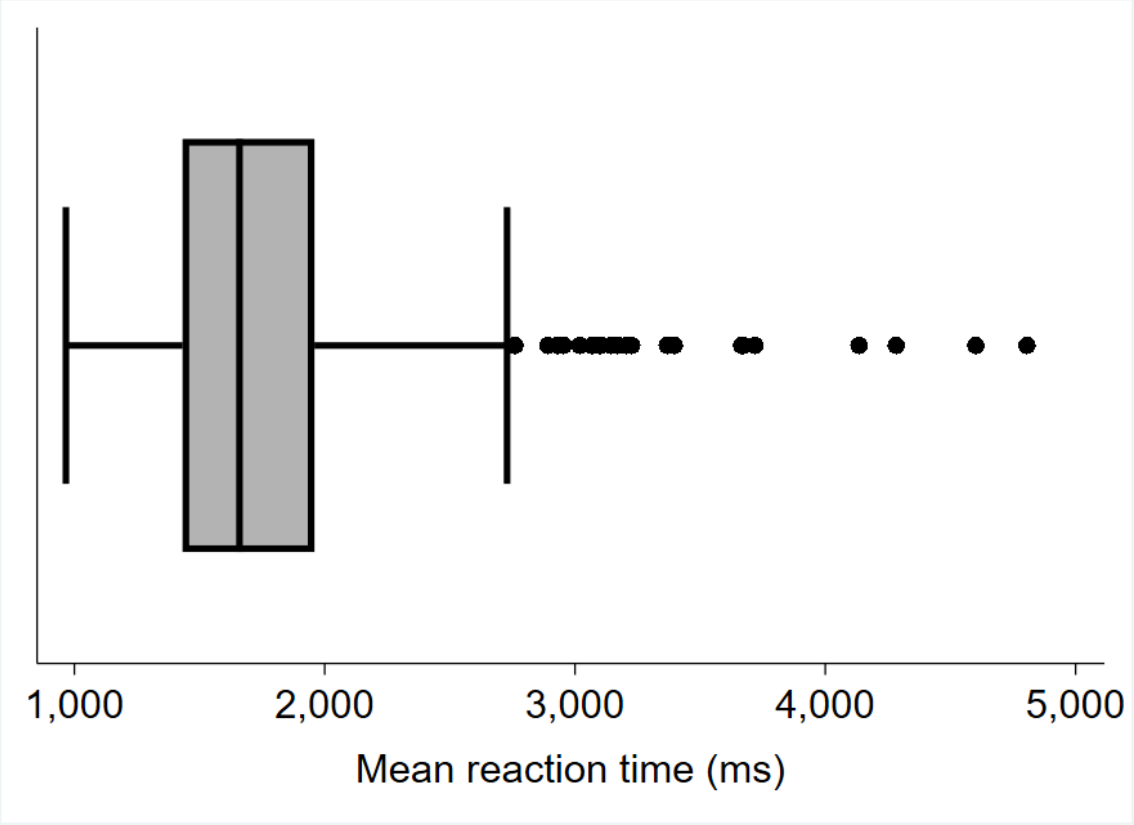


Figure 7-3. Box and whisker plot of mean reaction times (correct and incorrect responses) on the Response Inhibition task, before data cleaning

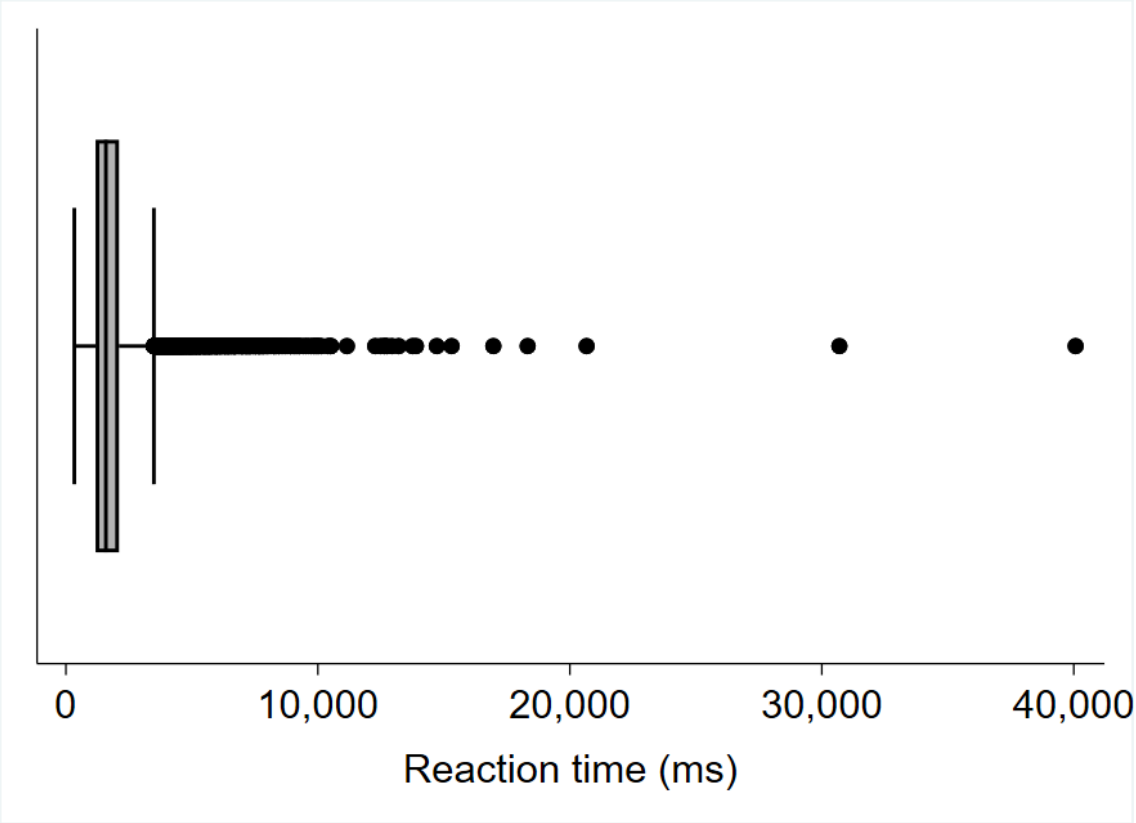


Figure 7-4. Box and whisker plot of individual reaction times (correct and incorrect responses) on the Response Inhibition task, before data cleaning

7.3. Patterns and predictors of performance

Following the format laid out in section 3.5, in the first part of this chapter I aimed to describe patterns of performance across the various outcomes and conditions of the task, and to investigate the effects of demographic and life-course predictors on performance in the full Insight 46 sample. Specifically I aimed to investigate whether RT and error rate differ in the different conditions (word *vs.* arrow stimuli, congruency, cue duration) and explore the relationship between RT and error rate. The demographic and life-course predictors (sex, age at assessment, childhood cognitive ability, educational attainment, adult socioeconomic position and presence of a neurological or psychiatric condition) are defined in sections 3.2.4 and 3.2.3 respectively.

7.3.1. Statistical Analyses

7.3.1.1. Run-position and practice effects

As described in section 7.2.3, the fact that trials were presented in an A-A-A-W-W-W sequence (where A indicates an arrow cue and W indicates a word cue) means that the speed and accuracy of a given response may potentially be affected by its position within this sequence. As the three experimental factors (congruency, cue type and cue duration) were not perfectly counterbalanced across the run-positions (see 7.2.1), the interpretation of the effects of these factors could be confounded by the influence of run-position.

Practice effects could also confound interpretation of the effects of these factors, particularly the effect of cue duration since the short cue block always preceded the long cue block, so the long cue block would theoretically benefit more from practice effects. As well as general improvement in speed and accuracy, practice effects could have led to implicit or explicit learning of the A-A-A-W-W-W sequence, which may have helped participants to improve their performance as the task went on.

Therefore, to account for these issues, run-position (1, 2 or 3) and trial number (1 to 47 within each block) were included in the analysis models as described below, and their effects are reported first (see section 7.3.2.1) to provide context for the other results which follow.

7.3.1.2. RT and error rate

Rather than using summary scores for each participant (e.g. mean RT for each condition), trial-by-trial responses were analysed to avoid losing information.

As in the previous chapter, reaction times were analysed using a GEE model assuming a normal distribution for the dependent variable and an identity link (as with standard linear regression), with an exchangeable correlation structure and robust standard errors to allow for the correlation between repeated measures of the same participant. RTs were first log-transformed so that the distribution more closely approximated the normal distribution.

As in the previous chapter, response accuracy (correct *vs.* incorrect) was analysed using a GEE logistic regression model with an independent correlation structure and robust standard errors. Results are expressed as odds ratios for ease of interpretation.

Predictors in the models were cue type (arrow *vs.* word), congruency (congruent *vs.* incongruent), cue duration (short *vs.* long), run-position (1, 2 or 3), trial number (1 to 47), sex, age at assessment, childhood cognitive ability, education, adult socioeconomic position and presence of a neurological or psychiatric condition (yes *vs.* no).

To investigate how cue type, congruency and cue duration combine to affect performance, I tested for interactions between these variables. I also tested for interactions between these variables and each between-subject predictor (sex, education etc.). In the case of congruency, these interaction tests allowed me to assess differences in inhibitory control, in accordance with my hypotheses. In the case of cue type and cue duration, these interaction tests were exploratory as I did not have prior hypotheses about how the between-subject predictors may be associated with differences in the degree to which participants would be disadvantaged by the short cue duration (compared to the long duration) or differences in performance on arrow and word trials.

7.3.1.3. Relationship between speed and accuracy

As in the previous chapter, I investigated the relationship between speed and accuracy both between-subjects and within-subject, to determine whether there was a trade-off between speed and accuracy as there was on the Choice Reaction Time task (see sections 6.3.1.3 and 6.3.2.3).

To address the between-subject questions (e.g. are the fastest-responding participants likely to make the most errors?), each participant's mean RT for correct responses was compared to their error rate (percentage of incorrect responses).

As in the Choice RT task, the distributions of mean RTs and error rates both had a positive skew. Two transformations were considered for mean RT: log-transforming the mean RTs or calculating the mean of the log-transformed raw RTs. As neither of these removed the skew, the untransformed data were used for ease of interpretation and for consistency with the analysis of the Choice RT data (see section 6.3.1.3). Spearman's rank correlation was used to examine the relationship between mean RT and error rate.

As in the Choice RT task, the within-subjects questions were addressed by analysing the trial-by-trial data to see whether the speed of a response predicted whether that response would be correct or incorrect. This was done by rerunning the GEE model for the odds of making errors (see 7.3.1.2) with RT included as an additional covariate. RT was not log-transformed this time, to aid interpretability of its effects as a predictor (it was

previously transformed to comply with the assumption of a normal distribution for linear regression when it was being modelled as an outcome).

7.3.2. Results

Descriptive statistics are presented in Table 7-1.

Table 7-1. Descriptive statistics for the Response Inhibition task

		Congruent	Incongruent	Congruent and Incongruent combined
RT for correct responses (ms)	Median	1491	1645	1560
	IQR	1111 - 1936	1225 - 2185	1161 - 2052
	Range	513 - 12273	353 - 13755	353 – 13755
Error rate (%)		0.8	10.7	5.6

Statistics are based on 500 participants who each completed 94 trials. Results for congruent and incongruent trials are presented separately as this was the main factor of interest. IQR = interquartile range

7.3.2.1. Run-position and practice effects

Run-position effects on RT are illustrated in Figure 7-5. Correct responses got faster with increasing run-position (regression coefficient = 0.982 (i.e. 1.8% decrease in RT per run-position), 95% CIs 0.978 to 0.986, $p < 0.0001$). As this effect was independent of the effect of trial number, it cannot be explained by the fact that run-positions 2 and 3 would benefit more on average from the general practice effect. Therefore, it is likely that some implicit and/or explicit learning of the A-A-A-W-W-W pattern facilitated faster responses across each group of three trials. It is worth noting that this does not constitute a "task-set switching" effect as described by Aron *et al.* (see 7.2.3), as pairwise comparisons of consecutive run-positions revealed that RT improved steadily from position 1 to 2 (regression coefficient = 0.982 (i.e. 1.8% decrease in RT), 95% CIs 0.974 to 0.989, $p < 0.0001$) and position 2 to 3 (regression coefficient = 0.985 (i.e. 1.5% decrease in RT), 95% CIs 0.980 to 0.991, $p < 0.0001$), whereas a task-set switching effect would manifest as a large decrease in RT from run-positions 1 to 2 only (see Figure 7-2).

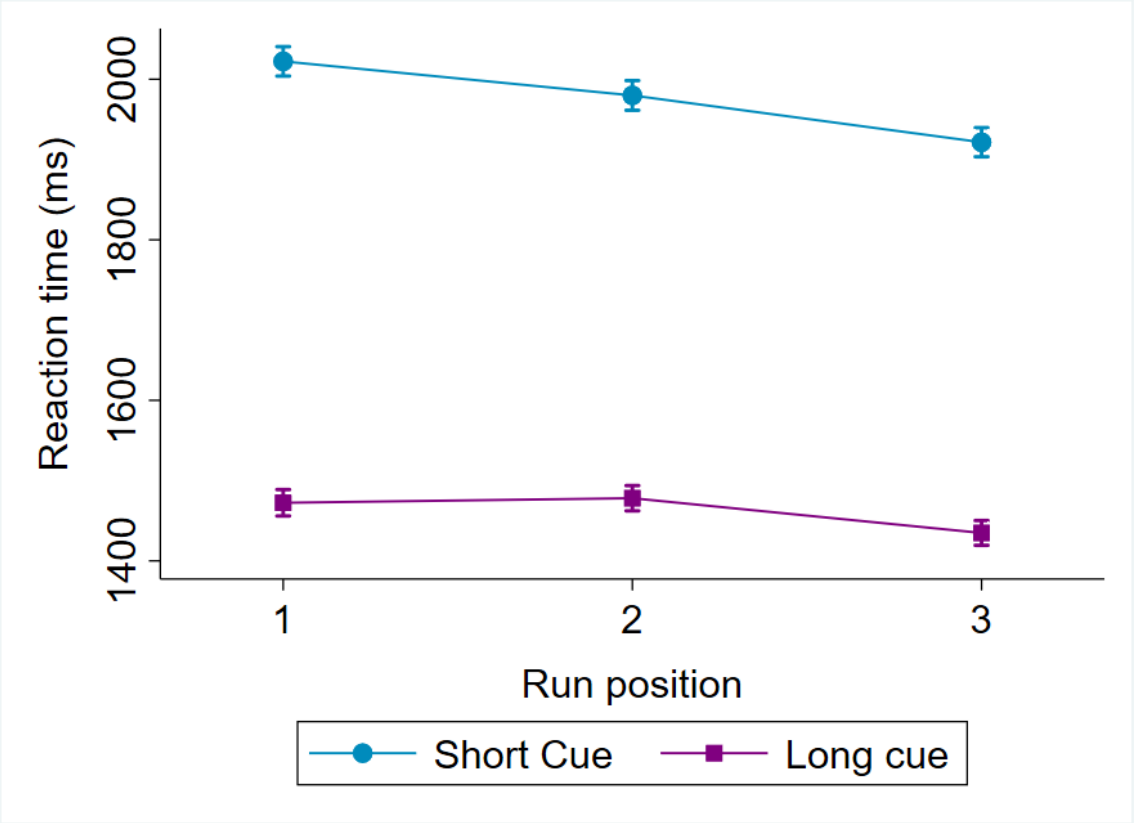


Figure 7-5. Mean reaction time at each run-position in the short cue and long cue blocks of the Response Inhibition task

Markers show unadjusted means and error bars show 95% confidence intervals.

Run-position effects on error rate are illustrated in (Figure 7-6). The odds of making an error decreased with increasing run-position (*OR* = 0.935, 95% CIs 0.890 to 0.982, *p* = 0.007). However, there were contrasting profiles in the short cue duration and long cue duration blocks: in the short cue duration block there was no main effect of run-position (*OR* = 0.954, 95% CIs 0.893 to 1.019, *p* = 0.16), whereas in the long cue duration block error rate decreased with increasing run-position (*OR* = 0.862, 95% CIs 0.801 to 0.928, *p* < 0.0001). The lack of a consistent relationship between run-position and error rate is probably due to the fact that error rate was heavily influenced by congruency and stimulus cue (arrow *vs.* word) (see 7.3.2.2.1) and these factors were not counterbalanced across the run-positions (see 7.2.1). Again, there is no evidence of the task-set switching effect described by Aron *et al.* (which would be indicated by substantially higher error rates at run-position 1 relative to the positions 2 and 3).

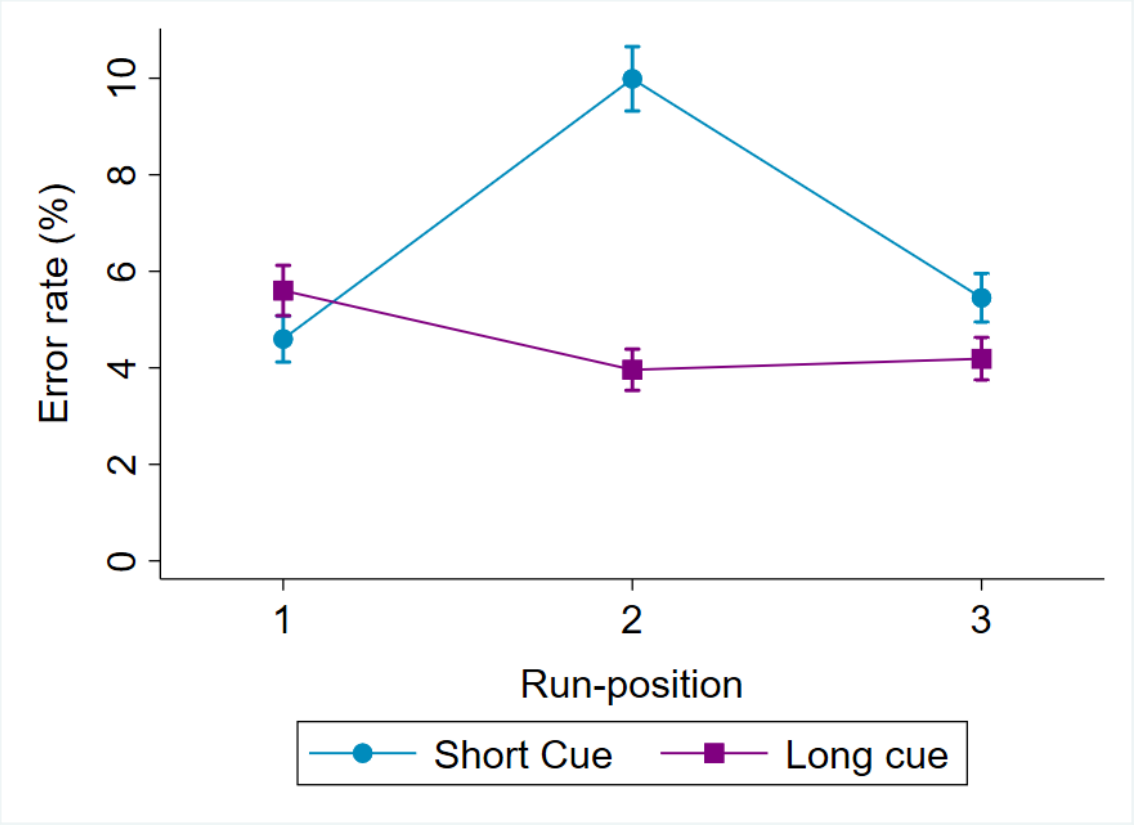


Figure 7-6. Error rate at each run-position in the short cue and long cue blocks of the Response Inhibition task

Markers show unadjusted means and error bars show 95% confidence intervals.

Practice effects on RT are illustrated in Figure 7-7. There was a general practice effect whereby RT got slightly faster with increasing trial number (regression coefficient = 0.999 (i.e. 0.1% decrease in RT per trial), 95% CIs 0.998 to 0.999, $p < 0.0001$). However, when analysing the 'short cue' and 'long cue' blocks separately, it is clear that this effect derives from the short cue block (regression coefficient = 0.997 (i.e. 0.3% decrease in RT per trial), 95% CIs 0.997 to 0.997, $p < 0.0001$), and there is no statistically significant effect in the long cue block (regression coefficient = 1.000, 95% CIs 0.999 to 1.000, $p = 0.42$). This could be partly because the long cue duration block was administered after the short cue duration block, so participants may have already had sufficient practice to reach their optimum level of performance, but also because the long cue duration block is easier (see next section) so participants may have reached a plateau very quickly.

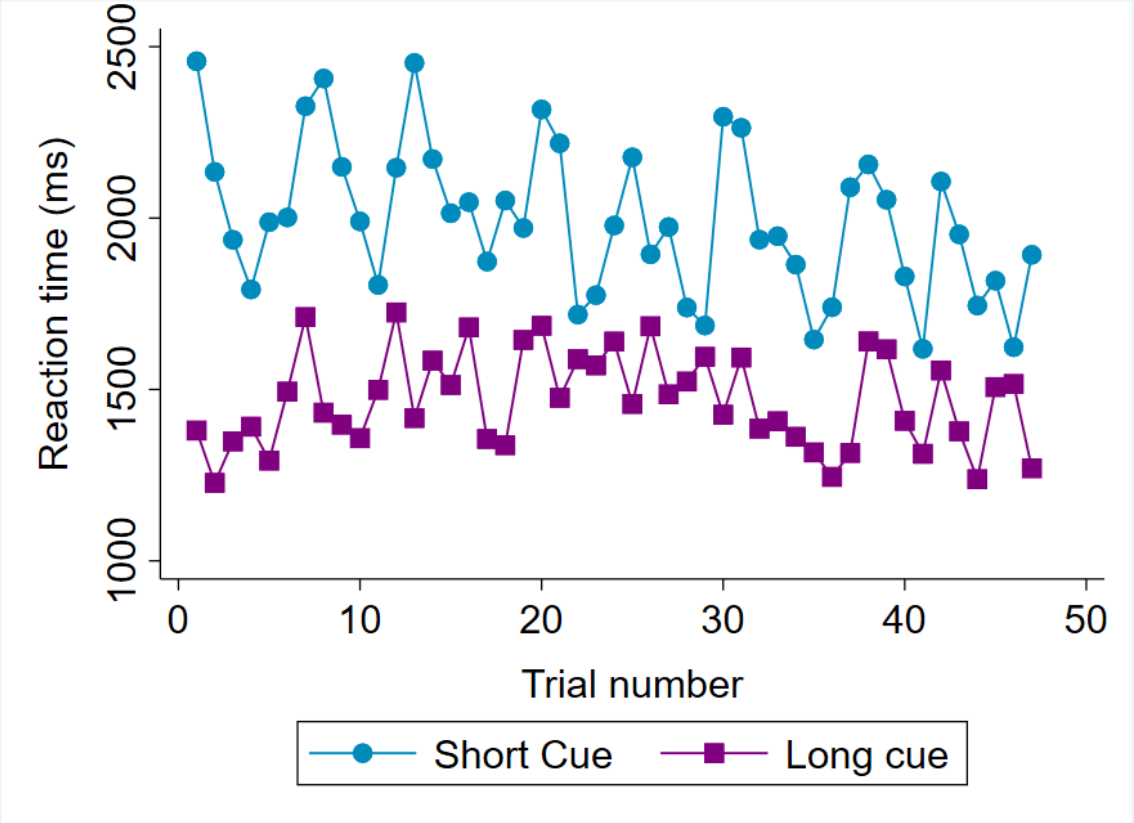


Figure 7-7. Mean Reaction Time for each trial in the short cue and long cue blocks of the Response Inhibition task

There were 47 trials within each condition. The long cue block was always administered first.

Practice effects on error rate are illustrated in Figure 7-8. A slight practice effect (i.e. a reduction in error rate with increasing trial number) was observed overall (*OR* = 0.993, *z* = -4.54, *p* < 0.0001). As with RT, this was observed in the short cue block *(OR* = 0.982, *z* = -7.62, *p* < 0.0001) but not the long cue duration block, which had evidence of a slight worsening of performance over time *(OR* = 1.008, *z* = 3.09, *p* = 0.002).

Overall, run-position and trial number had small independent effects on RT and error rate so it is justified to account for them in the following analyses, but there is no cause to be concerned that the comparison between the short and long cue durations will be significantly compromised by practice effects.

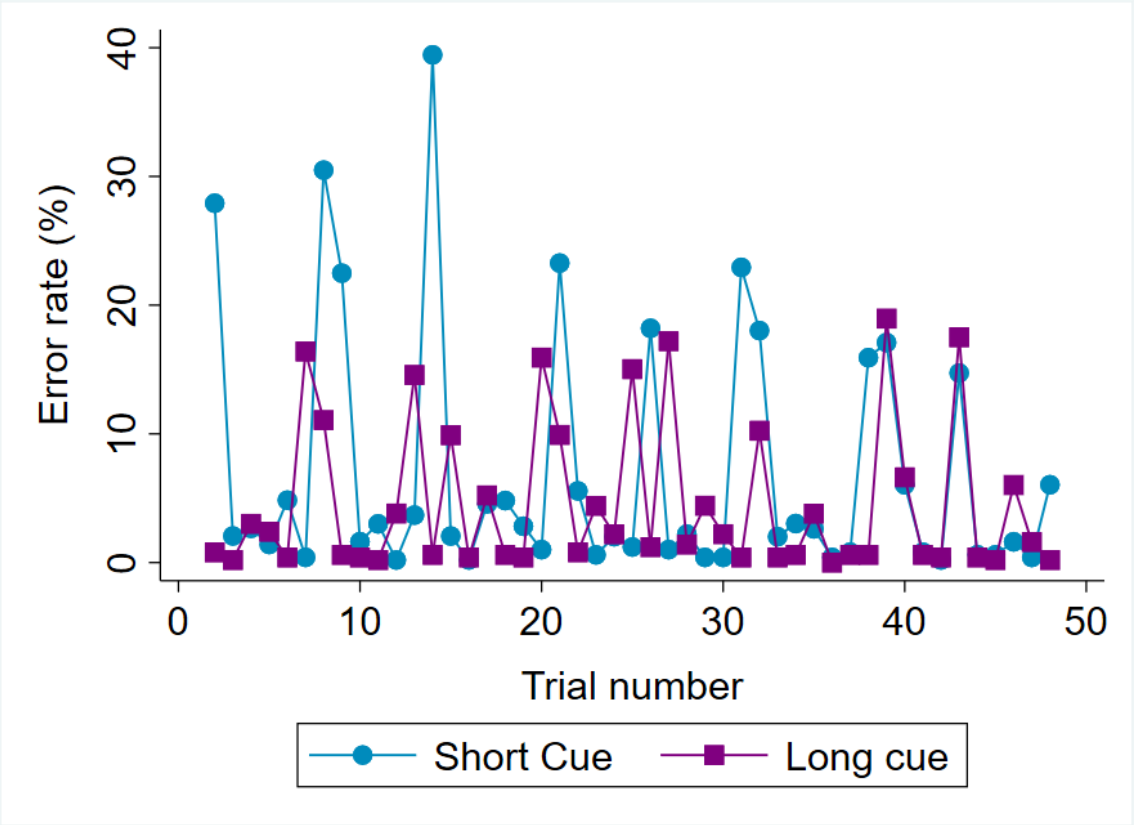


Figure 7-8. Error rate for each trial in the short cue and long cue blocks of the Response Inhibition task

There were 47 trials within each condition. The long cue block was always administered first.

7.3.2.2. RT and error rate

7.3.2.2.1. Cue duration, congruency and cue type

Results of the multivariable regression models for RT and error rate are given in Table 7-2 and Table 7-3 respectively, along with results of interaction tests between predictors.

Responses were 39% slower on average when the cue was of short duration compared to long duration (Table 7-2), and also 1.65 times more likely to be incorrect (Table 7-3). This is as expected, because the long cue duration gives people more time to prepare the appropriate task set for their response. As mentioned above, a contribution of practice effects to this result cannot be ruled out, as the long cue duration block came after the short cue duration block.

Incongruent trials were associated with 13% slower responses than congruent trials (Table 7-2) and the odds of making an error were 17.17 times greater (Table 7-3).

Responses were slightly slower (6%) to arrow cues than word cues (Table 7-2) and also 5.32 times more likely to be incorrect (Table 7-3).

Several interaction effects were observed between these factors; as an aid to the interpretation of these interaction effects, described below, mean RT and error rate for each combination of cue duration, congruency and cue type are shown in Figure 7-9 and Figure 7-10.

The slowing of responses and the increased error rate associated with incongruent trials were both disproportionately greater when the cue type was arrow rather than word (Table 7-2, Table 7-3, Figure 7-9, Figure 7-10). The slowing of responses associated with incongruent trials was slightly reduced when the cue was of short duration compared to long duration (Table 7-2, Figure 7-9), but there was no evidence that the increase in error rate on incongruent trials differed between the short and long cue duration conditions (Table 7-3, Figure 7-10). Finally, the slowing of responses and the increased error rate associated with arrow cues were both disproportionately greater when the cue was of short duration compared to long duration (Table 7-2, Table 7-3, Figure 7-9, Figure 7-10).

Overall these results show that both RT and error rate were influenced by all three experimental factors, with the most difficult combination being an arrow cue of short duration for an incongruent stimulus. The biggest influence on RT was cue duration whereas the biggest influence on error rate was congruency. It is clear from Figure 7-10 that the error rate for congruent trials was extremely low, regardless of cue duration and cue type, whereas incongruent trials were much more error-prone, especially when participants were cued to respond to the arrow. This suggests that the word was the most salient part of the stimulus and participants found it difficult to inhibit a response to it.

Table 7-2. Associations between demographic and life-course predictors and RT in the Response Inhibition task (n = 500)

Predictor	Coefficient [a] (95% confidence intervals)			
	Main effect of predictor	**Interaction between predictor and cue duration**	**Interaction between predictor and congruency**	**Interaction between predictor and cue type**
Short cue duration (long cue as reference)	**1.39*** (1.37, 1.41)	N/A	**0.97*** (0.96, 0.98)	**1.13*** (1.11, 1.14)
Incongruent stimulus (congruent as reference)	**1.13*** (1.12, 1.14)	**0.97*** (0.96, 0.98)	N/A	**1.08*** (1.07, 1.10)
Arrow cue type (word as reference)	**1.06*** (1.05, 1.07)	**1.13*** (1.11, 1.14)	**1.08*** (1.07, 1.10)	N/A
Sex (female as reference)	**0.92*** (0.88, 0.96)	0.98 (0.95, 1.01)	0.99 (0.98, 1.01)	**0.98** (0.96, 1.00)
Age at assessment (per year)	**1.05*** (1.02, 1.08)	0.98 (0.96, 1.00)	1.00 (0.99, 1.01)	1.00 (0.99, 1.01)
Childhood cognitive ability (per z-score)	**0.95*** (0.92, 0.98)	1.00 (0.98, 1.01)	**0.98*** (0.97, 0.99)	1.00 (0.99, 1.01)
Education (per category) [b]	0.98 (0.96, 1.01)	0.99 (0.98, 1.00)	**0.99*** (0.98, 1.00)	1.00 (0.99, 1.01)
Adult SEP (per category) [b]	0.99 (0.97, 1.02)	1.00 (0.98, 1.01)	**0.99*** (0.98, 1.00)	1.00 (0.99, 1.01)
Neurological or psychiatric condition [c] (cognitively-normal as reference)	1.05 (0.97, 1.14)	1.01 (0.96, 1.96)	1.03 (1.00, 1.06)	0.99 (0.96, 1.03)

Coefficients in bold are significant at p < 0.05 and asterisks indicate significance at p < 0.01. Multivariable regression models were used so each association is independent of all others. [a] As the data were log-transformed, the coefficients are quoted in exponentiated form for ease of interpretation; for example, a coefficient of 1.05 would mean that the factor was associated with 5% longer response time. [b] See section 3.2.4 for definition of categories. [c] See section 3.2.3 for definitions.

SEP = socioeconomic position.

Table 7-3. Associations between demographic and life-course predictors and error rate in the Response Inhibition task (n = 500)

Predictor	Odds Ratio for making an error (95% CIs)			
	Main effect of predictor	**Interaction between predictor and cue duration**	**Interaction between predictor and congruency**	**Interaction between predictor and cue type**
Short cue duration (long cue as reference)	**1.65*** (1.48, 1.83)	N/A	0.75 (0.50, 1.14)	**1.79*** (1.36, 2.34)
Incongruent stimulus (congruent as reference)	**17.17*** (12.96, 22.75)	0.75 (0.50, 1.14)	N/A	**3.51*** (2.45, 5.04)
Arrow cue type (word as reference)	**5.32*** (4.34, 6.52)	**1.79*** (1.36, 2.34)	**3.51*** (2.45, 5.04)	N/A
Sex (female as reference)	1.25 (0.98, 1.59)	0.87 (0.71, 1.08)	0.71 (0.42, 1.20)	0.83 (0.55, 1.23)
Age at assessment (per year)	0.97 (0.82, 1.14)	1.00 (0.86, 1.16)	1.33 (0.94, 1.90)	**1.37** (1.02, 1.85)
Childhood cognitive ability (per z-score)	**0.70*** (0.59, 0.84)	1.06 (0.93, 1.22)	0.83 (0.63, 1.11)	**1.36** (1.06, 1.76)
Education (per category) [a]	**0.82*** (0.73, 0.91)	1.06 (0.98, 1.14)	1.05 (0.86, 1.29)	**1.15** (1.00, 1.32)
Adult SEP (per category) [a]	**0.87** (0.78, 0.98)	0.99 (0.91, 1.07)	1.06 (0.86, 1.30)	1.07 (0.91, 1.26)
Neurological or psychiatric condition [b] (cognitively-normal as reference)	1.33 (0.88, 2.03)	0.91 (0.67, 1.24)	0.84 (0.47, 1.49)	0.80 (0.42, 1.56)

Coefficients in bold are significant at p < 0.05 and asterisks indicate significance at p < 0.01. Multivariable regression models were used so each association is independent of all others.
[a] See section 3.2.4 for definition of categories. [b] See section 3.2.3 for definitions.

CI = confidence interval; SEP = socioeconomic position.

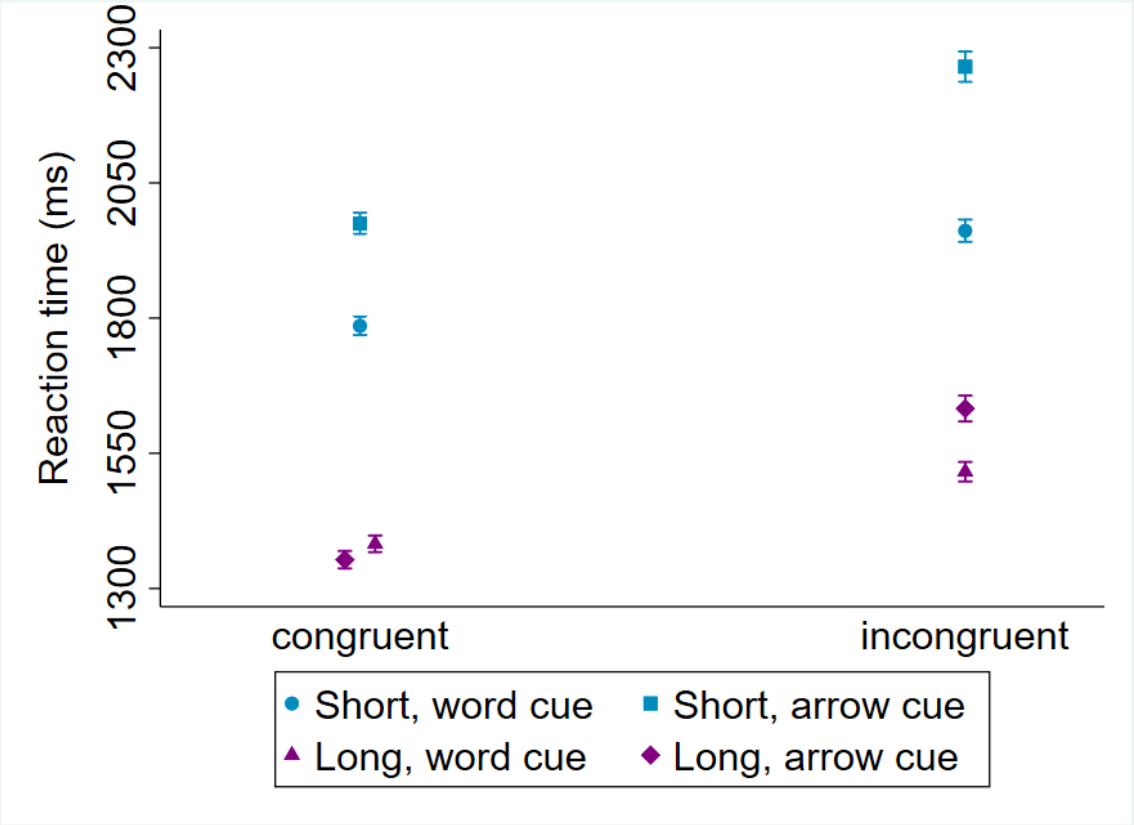


Figure 7-9. Mean reaction time by congruency, cue type (arrow vs. word) and cue duration (short vs. long) on the Response Inhibition task

Markers show unadjusted means and error bars show 95% confidence intervals.

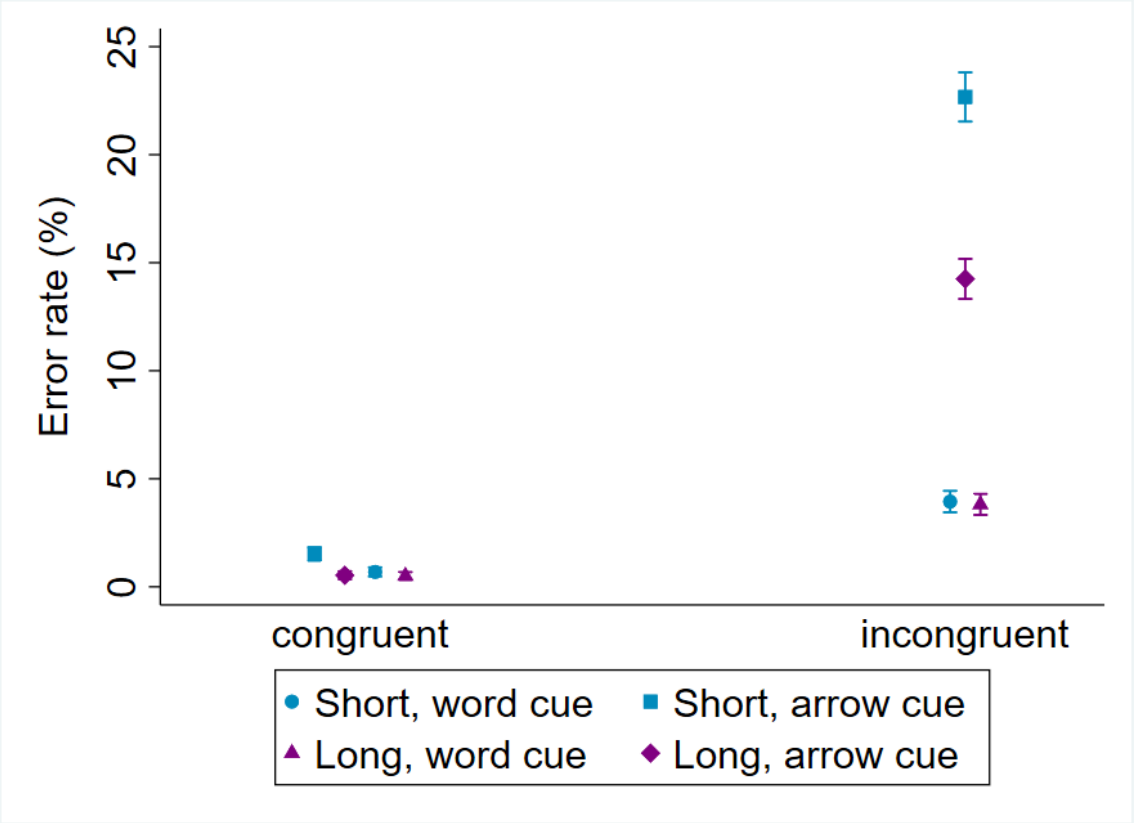


Figure 7-10. Error rate by congruency, cue type (arrow vs. word) and cue duration (short vs. long) on the Response Inhibition task

Markers show unadjusted means and error bars show 95% confidence intervals.

7.3.2.2.2.Demographic and life-course predictors

Results of the multivariable regression models for RT (correct responses only) and error rate are given in Table 7-2 and Table 7-3 respectively, along with results of interaction tests between predictors.

On average, males responded 8% faster than females (1640 *vs.* 1788 ms, Table 7-2). There was evidence of an interaction between sex and cue type, such that the tendency to respond more slowly to arrow cues than words cues was reduced in males, i.e. their RTs to arrows and words were more similar. There was no evidence of statistically significant sex differences in overall error rates (males = 6.0%; females = 5.3%) nor in the extent to which accuracy was influenced by cue duration, congruency or cue type (Table 7-3).

Older age at assessment was associated with slightly slower RT (5% slower per year), an effect which appeared to generalise across all conditions of the task as there was no evidence of interactions between age and cue duration, congruency or cue type (Table

7-2). There was no evidence of a general ageing effect on error rate, but the tendency to make more errors in response to arrow cues (as opposed to word cues) was exaggerated in older participants (Table 7-3).

Participants with higher childhood cognitive ability tended to respond faster overall (5% faster per z-score unit of childhood cognitive ability) and had a reduced tendency to respond more slowly to incongruent stimuli, i.e. their RTs to congruent and incongruent stimuli were more similar, indicating better inhibitory control (Table 7-2). While higher educational attainment and higher adult socioeconomic position did not have a statistically significant effect on RT overall, these factors were also associated with a reduction in the tendency to respond more slowly to incongruent stimuli, indicating better inhibitory control (Table 7-2).

Childhood cognitive ability, education and adult socioeconomic position all had notable independent effects on error rate (odds of making an error reduced by 30% per unit z-score of childhood cognitive ability, reduced by 18% per category of educational attainment, and reduced by 13% per category of adult socioeconomic position, Table 7-3, Figure 7-11). There was no evidence of interactions between these factors and the tendency to make more errors on incongruent trials, although as error rates in the congruent condition were so low it should be noted that the overall error rates were driven almost entirely by errors on incongruent trials. There was also no evidence of interactions between these factors and the tendency to make more errors when the cue was of short duration, but there were interaction effects with cue type such that those with higher childhood cognitive ability and higher education attainment made relatively more errors in response to arrow cues (as opposed to word cues). In other words, their accuracy advantage was greater when they were required to respond to the word.

There was no evidence of any statistically significant differences between participants with neurological or psychiatric conditions and cognitively-normal participants in terms of RT or error rate (Table 7-2, Table 7-3).

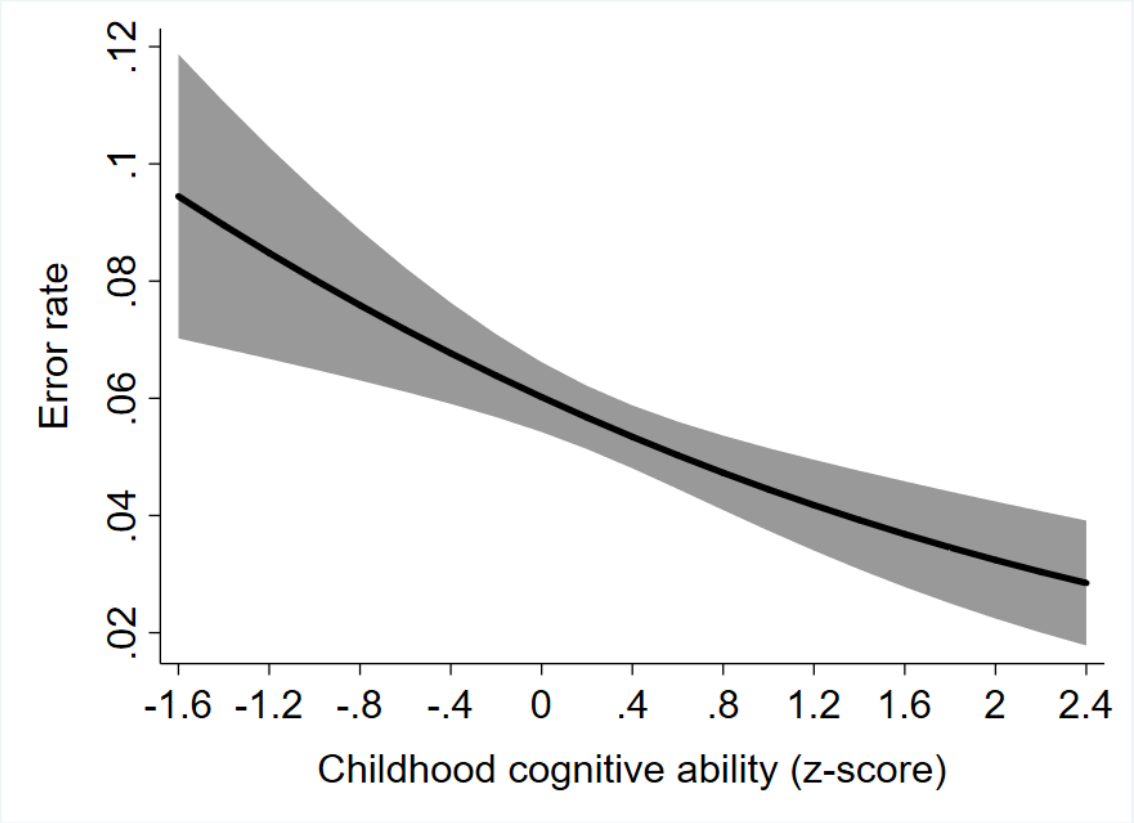


Figure 7-11. Association between childhood cognitive ability and error rate on the Response Inhibition task

Error rate is represented as a proportion between 0 and 1. Solid line indicates predictions from the multivariate regression model, adjusted for sex, age at assessment, education, adult socioeconomic position and presence of a neurological or psychiatric condition. Shaded area represents 95% confidence intervals. For an explanation of the childhood cognitive ability variable, see section 3.2.4.

7.3.2.3. Relationship between speed and accuracy

Figure 7-12 illustrates the relationship between mean RT and accuracy. There was a ceiling effect on accuracy with many participants having an error rate close to zero regardless of their speed, but there was a positive association (rather than a trade-off) between speed and accuracy, such that slower mean RT for correct responses was predictive of a higher error rate (Spearman's $\rho = 0.33$, $p < 0.0001$).

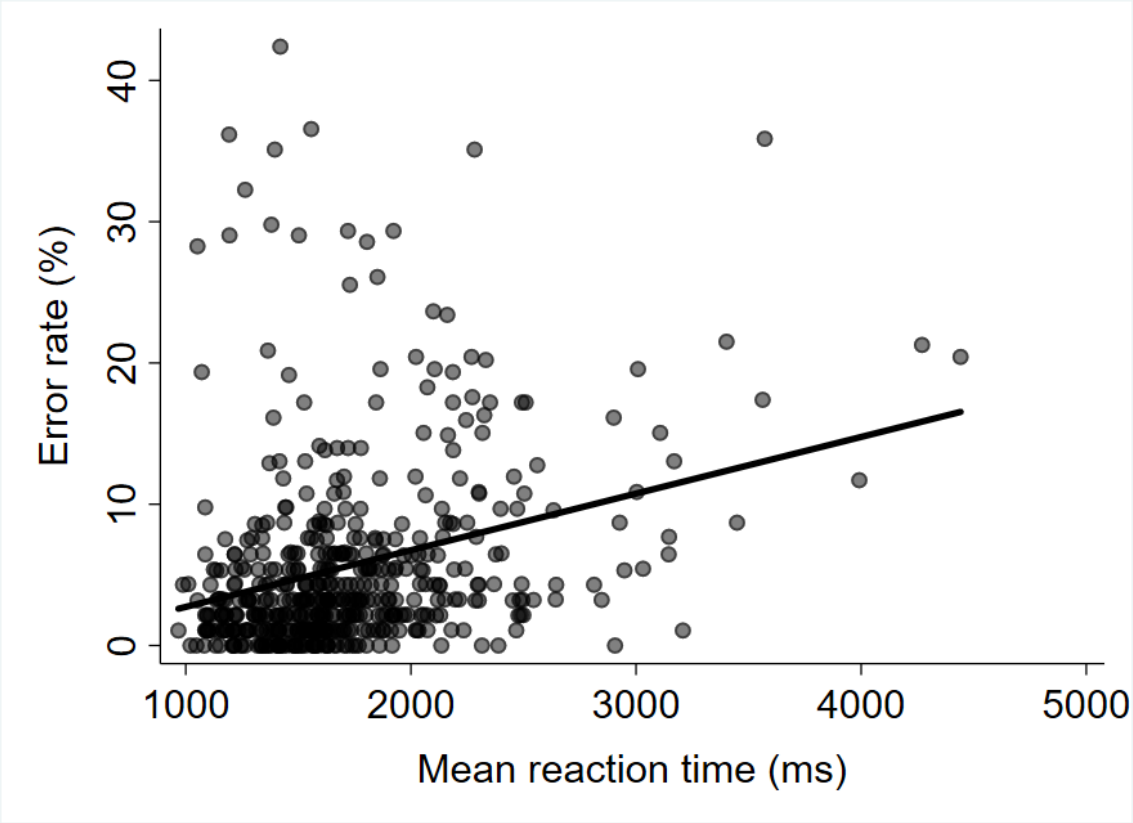


Figure 7-12. Scatter plot of each participant's mean reaction time against their error rate on the Response Inhibition task

Transparent markers have been used so that the density is visible where there are markers on top of each other. Solid line indicates line of best fit.

The within-subjects analysis of trial-by-trial responses found that errors were more likely to occur with increasing RT, with a 2% increase in the odds of making an error per additional 100 milliseconds (*OR* = 1.02, 95% CIs 1.01 to 1.03, $p < 0.0001$). To put this in context, the mean RTs for correct and incorrect responses were 1712 ms and 2175 ms respectively. The predictors of higher error rate remained unchanged when adjusting for RT: incongruent stimuli (*OR* = 16.21, 95% CIs 12.19 to 21.54, $p < 0.0001$), short cue duration (*OR* = 1.48, 95% CIs 1.31 to 1.66, $p < 0.0001$), and arrow cues (*OR* = 5.11, 95% CIs 4.16 to 6.26, $p < 0.0001$). This provides evidence that RT and error rate were mainly dependent on the properties of the stimuli (i.e. certain conditions were more difficult which had a detrimental effect on both speed and accuracy), rather than errors being due to responding in haste.

7.4. Associations with biomarkers and *APOE*-ε4

Following the format laid out in section 3.5, the second part of this chapter aims to investigate associations between performance and biomarkers of AD in cognitively-normal participants for whom complete biomarker data are available. The number of participants meeting these criteria who also had usable data from the Response Inhibition task was 405 (see section 3.3).

As explained in section 3.5.2.1, I wanted to derive some summary scores that capture the key aspects of performance on each task, to use for comparing results across the different cognitive tests in the Insight 46 battery (see Chapter 9). As a test of response inhibition, the outcome of primary interest in this task is how well participants can inhibit incorrect responses when presented with incongruent stimuli. This is the dimension of the task which my hypotheses relate to, rather than the other factors that were varied across the experiment, namely cue duration and cue type (arrow *vs.* word). The results reported above support the idea that congruency is the aspect on which interesting differences in RT between participants are observed, as the extent to which RT was slowed on incongruent stimuli varied by childhood cognitive ability, education and adult socioeconomic position, whereas the effects of cue duration and cue type on RT did not vary by any of the predictors tested (except a small interaction between sex and cue type (see section 7.3.2.2.2).

Based on this, I decided to calculate summary scores that quantify the extent to which RT and accuracy were compromised on incongruent trials, compared to congruent trials. This approach is commonly adopted on response inhibition tasks. For example on the Simon task, the 'Simon effect' is the difference between mean RT for congruent trials and mean RT for incongruent trials (e.g. Stoet, 2017), and the 'Stroop effect' is usually defined as the total time to complete the incongruent condition of the Stroop task minus the total time to complete the congruent condition (MacLeod, 1991). I calculated the following two summary outcomes for each participant:

i) **"Incongruent cost" to RT**, defined as the mean RT for incongruent trials minus the mean RT for congruent trials, divided by the mean RT for congruent trials. This is a measure of the relative increase in RT for incongruent trials. I chose to calculate the relative difference, rather than the absolute difference, because it may be more meaningful given the large range of RTs and because it more closely approximates the normal distribution.
ii) **"Incongruent cost" to error rate**, defined as the percentage of incorrect responses on incongruent trials minus the percentage of incorrect responses on

congruent trials. I considered whether to normalise this outcome to each participant's performance in the congruent condition (i.e. to calculate the relative difference in error rate rather than the absolute difference) but this was problematic because it would require dividing by zero for the vast majority of participants (83%) who made no errors on congruent trials. Because of this ceiling effect on accuracy for congruent trials, this outcome is essentially a measure of error rate on incongruent trials, but one that attempts to minimise the influence of anticipations or lapses of attention – the most likely explanation for making errors in the congruent condition.

For each of these two outcomes, I tested for associations with the same biomarkers as in the previous chapter (see section 3.5.2).

7.4.1. Statistical Analyses

"Incongruent cost" to RT and **"incongruent cost" to error rate** were both analysed using multivariable linear regression models. For the error rate variable, bootstrapping was used to produce bias-corrected and accelerated 95% CIs from 2000 replications, as it had a positive skew**.**

All models included predictors of amyloid status (positive *vs.* negative), whole brain volume, WMHV and *APOE* genotype (ε4-carrier *vs.* non-carrier). To adjust for the correlation between whole brain volume and head size, total intracranial volume (TIV) was included in all models, as were the demographic factors investigated in section 6.3 (sex, age at assessment, childhood cognitive ability, education and adult socioeconomic position).

The models were additionally rerun replacing dichotomised amyloid status with SUVR to test whether increasing Aβ deposition was associated with differences in performance. To check whether associations between SUVR and cognition were sensitive to the inclusion of the imputed SUVR values (see section 3.2.2), the analyses were rerun excluding the 26 participants with imputed SUVR data.

7.4.2. Results

Results of the regression models for the two outcomes are reported in Table 7-4. Results for the demographic and life-course factors (sex, age at assessment, childhood cognitive ability, education and adult socioeconomic position) are not reported as they are essentially unchanged from the first analysis section (7.3.2).

There was no evidence of any associations between any biomarkers and either of the two outcomes.

Table 7-4. Associations between biomarkers and Response Inhibition outcomes in cognitively-normal participants (n = 405)

Predictor	Coefficients and 95% CIs	
	"Incongruent cost" to RT [a]	**"Incongruent cost" to error rate [b]**
β-amyloid status (negative as reference)	-0.01 (-0.04, 0.02)	0.77 (-1.79, 3.99)
WMHV (per 10 ml)	0.01 (-0.01, 0.03)	0.90 (-0.61, 2.40)
Whole brain volume (per 10 ml)	-0.00 (-0.00, 0.00)	0.02 (-0.17, 0.24)
***APOE*-ε4** (non-carriers as reference)	0.00 (-0.02, 0.03)	-1.04 (-3.34, 1.38)

Multivariable regression models were used so each association is independent of all others. In addition to the predictors listed, models also included sex, age at assessment, childhood cognitive ability, adult socioeconomic position and total intracranial volume.

Coefficients in bold are significant at p < 0.05 and asterisks indicate significance at p < 0.01.
[a] "Incongruent cost" to RT is the relative increase in RT for incongruent trials, defined as the mean RT for incongruent trials minus the mean RT for congruent trials, divided by the mean RT for congruent trials. [b] "Incongruent cost" to error rate is defined as the percentage of incorrect responses on incongruent trials minus the percentage of incorrect responses on congruent trials. CI = confidence interval; WMHV = white matter hyperintensity volume.

7.5. Discussion

7.5.1. Summary

This study investigated inhibitory control on a task where incongruent stimuli were used to create a conflict between the automatic (but incorrect) response and the correct response. As hypothesised, there was evidence that higher childhood cognitive ability and education were associated with increased inhibitory control, as evidenced by a lower error rate and a reduction in the tendency to respond more slowly to incongruent stimuli. There was evidence of a similar effect of higher adult socioeconomic position. There was no evidence to support the hypothesis that Aβ+ participants would show reduced inhibitory control. There was no evidence of speed-accuracy trade-offs; rather there was a positive association between speed and accuracy. These results are discussed in greater detail in the following sub-sections.

7.5.2. Patterns of performance

As expected, responses were slower when the stimuli were incongruent and when participants had less time to prepare for their response (i.e. the cue was of short duration). Responses were also slightly slower when the cue instructed participants to respond to the arrow rather than the word. This may be explained by the fact that reading words is an automatic or "overlearned" process, and would naturally take precedence over attending to the direction of the arrow – this is the phenomenon behind the well-established Stroop effect (see 7.1). The results of the Choice RT experiment reported in the previous chapter support this explanation, because in the Choice RT experiment (where arrows and words were presented alone, not combined together as in the Response Inhibition experiment) responses were slightly faster to arrows than words, suggesting that in the Response Inhibition task it was indeed the presence of the irrelevant word that slowed down responses to arrows. In the experiment by Aron *et al.* (2004) on which this one was based, cue type was not considered as a factor in the analyses so it is not possible to compare my results with theirs in this respect.

The finding that the fastest-responding participants tended to make the fewest errors, is the opposite of the relationship observed in the Choice RT experiment reported in the previous chapter, where there was a trade-off between speed and accuracy. Similarly,

the observation that incorrect responses were typically slower than correct responses was the opposite of the Choice RT experiment. In my discussion of those results (section 6.5.2), I hypothesised that the complexity of the task is the key factor in determining the relationship between speed and accuracy, as errors in a more complex task may arise primarily from making an incorrect judgement, whereas on a more simple task errors may arise primarily from anticipatory or hasty responses.

7.5.3. Demographic and life-course predictors

7.5.3.1. Associations with age at assessment

As on the Choice RT experiment reported in the previous chapter, there was an association between older age and slower RT. However, there was no evidence of an age effect on inhibitory control, as the tendency to respond more slowly to incongruent stimuli was independent of age. This agrees with the conclusion of a recent meta-analysis which concluded that there is little evidence that inhibitory control declines with age (Rey-Mermet and Gade, 2018), although the very narrow age range of Insight 46 participants (2.6 years – reflecting the time it took to collect the data, since all participants were born in the same week) means that I was not expecting to observe age effects in general. As discussed previously (see sections 4.5.2.3 and 6.5.3.1) I considered the possibility of a recruitment bias whereby participants seen towards the beginning of the data collection period may have differed in some ways to those seen towards to the end, and this is discussed in greater detail in Chapter 10.

7.5.3.2. Associations with childhood cognitive ability, education and adult socioeconomic position

Error rate on this task showed strong independent associations with childhood cognitive ability, education and adult socioeconomic position. As error rates for congruent stimuli were extremely low, these associations are driven by errors in the incongruent condition, which occurred primarily when participants were required to ignore the word and respond to the direction of the arrow (see 6.5.2). Higher childhood cognitive ability, higher educational attainment and higher adult socioeconomic position also independently predicted a reduced tendency to slow down when the stimuli were incongruent, indicating better inhibitory control. These results are consistent with previous studies that have described an effect of education on response inhibition tasks (Puccioni and Vallesi, 2012;

Aschenbrenner *et al.*, 2015), but to my knowledge this is the first study to report independent effects of these factors on inhibitory control in older age.

Consistent with the results of the Choice RT experiment reported in the previous chapter, higher childhood cognitive ability was associated with faster RT overall, whereas there was no evidence for associations between RT and education or adult socioeconomic position. As discussed in the previous chapter, associations between RT and general cognitive ability are widely reported in the literature.

7.5.3.3. Sex differences

There was evidence of a sex difference in RT whereby males generally responded faster. In the previous chapter I discussed the fact that the lack of evidence for sex differences on the Choice RT experiment was contrary to some previous studies which had reported that males tended to respond faster on Choice RT tasks (Deary and Der, 2005; Der and Deary, 2006; Dykiert *et al.*, 2012b; Vincent *et al.*, 2012; Phillips *et al.*, 2013). While this task was primarily designed to measure inhibitory control, it can still be considered as a Choice RT task, albeit one with an extra layer of demands due to the congruent and incongruent stimuli. One theory of relevance that could explain this finding is the hypothesis that females may show greater post-error slowing (Thakkar *et al.*, 2014; Fischer *et al.*, 2016), so I plan to conduct further analyses to see if this was indeed the case.

This study did not find any evidence of a sex difference in inhibitory control, which is consistent with the literature on the Stroop task where sex differences in inhibitory control are not generally observed (MacLeod, 1991) but contrary to the results of a previous study which reported that males had reduced interference on the Simon task (Stoet, 2017). I would speculate that a possible reason for the discrepancy could be related to the design of the task, because the incongruity in the Simon task comes from the direction of an arrow and its location, whereas the incongruity in the Stroop task comes from a word and colour, and in the task used in Insight 46 it comes from a word and the direction of an arrow. As discussed in Chapter 5 in the context of the observation that males had slightly better memory for the location of objects in the “What was where?” task, there is evidence for sex differences in spatial abilities, so this could perhaps play a role in the result reported on the Simon task. However, it should also be noted that the literature on sex difference in the Simon task is somewhat mixed (Stoet, 2017).

7.5.4. Associations with biomarkers and APOE-ε4

On the two summary outcomes chosen as measures of inhibitory control on this task, there was no evidence of differences between cognitively-normal Aβ+ and Aβ- participants, nor of any associations with whole brain volume, WMHV or *APOE* genotype. Given that I have found evidence that β-amyloid deposition is associated with subtly poorer performance on a range of other cognitive measures covering a variety of cognitive domains (see Chapters 4 to 6), this may suggest that inhibitory control is not an area in which such early changes are seen. That would be contrary to the conclusion of a previous study which suggested that decline in inhibitory control may be evident before memory impairment, as they found that individuals with pathological β-amyloid$_{42}$/tau ratios showed greater interference on the Stroop task but did not differ from controls on any other cognitive measure (Harrington *et al.*, 2013). The participants in that study were slightly older and the sample size was small (mean age 76 for controls, n = 36; mean age 78 for those with pathological β-amyloid$_{42}$/tau ratios, n = 34). While the evidence for impaired inhibitory control in MCI and AD dementia is more well-established (e.g. Castel *et al.*, 2007; Bélanger, Belleville and Gauthier, 2010) further studies are needed to understand the timing of declines in this area and how this may relate to accumulation of AD pathology. It still remains to be understood how potential declines in inhibitory control may relate to changes in other executive functions such as processing speed, cognitive flexibility, attention and working memory, where there is accumulating evidence for subtle declines in the preclinical stage of AD (Grober *et al.*, 2008; Clark *et al.*, 2012; Hassenstab *et al.*, 2015; Baker *et al.*, 2017; Duke Han *et al.*, 2017; Mortamais *et al.*, 2017).

7.5.5. Strengths and limitations

This task worked well as a measure of inhibitory control, in that it detected substantial differences between participants in terms of their ability to deal with incongruent stimuli. However its design was complex, arising from the fact that it was initially based on a task-set shifting experiment, but I concluded that its suitability for measuring task-set shifting was too limited (see 7.2.3). In order to measure task-set shifting, the experiment would need a greater number of trials, and for trial types to be properly counterbalanced.

In the previous chapter I made three recommendations to improve the design of the Choice RT experiment (see section 6.5.5). The same three recommendations apply to the Response Inhibition experiment. The fact that the first trial of each block had to be dropped (see first recommendation) had an additional consequence on the Response Inhibition task because it meant that the number of trials was no longer perfectly

counterbalanced for congruency, cue type (arrow / word) and cue duration (long / short), so there is an additional motivation to address that issue.

The experiment contained two blocks of a long and short cue duration, but I did not have prior hypotheses about whether certain groups of participants would perform differently depending on the cue duration, and in fact there were no differences between participants in this regard. This feature of the design could be removed without affecting the ability of the task to measure inhibitory control. Similarly, I did not have prior hypotheses about whether certain groups of participants would perform differently depending on whether they were cued to respond to the word or the arrow, so it was difficult to interpret these effects (for example, I observed that the tendency to respond more slowly to arrow cues than words cues was exaggerated in female participants and in older participants), and it should be remembered that some of the findings may be due to chance. For a future study, I would recommend either simplifying the design of this experiment to focus purely on inhibition, or using a different inhibition task. The Stroop task and Simon task (see section 7.1) are both good candidates. If the Stroop task is administered in a computerised format with a microphone to record the verbal responses (e.g. Bélanger, Belleville and Gauthier, 2010), then it offers the same advantages as the Simon task in terms of allowing the RT and accuracy of each individual response to be recorded.

Strengths and limitations that apply to all the analyses presented in this thesis, such as considerations relating to the generalisability of the sample, are discussed in Chapter 10.

7.5.6. Conclusion

The results of this Response Inhibition task have provided evidence that childhood cognitive ability, education and adult socioeconomic position each have independent effects on the ability to resist interference from irrelevant stimuli in older age. This task did not appear to be sensitive to brain pathology in cognitively-normal 70-year-olds.

8. CIRCLE-TRACING AND SERIAL SUBTRACTION

8.1. Introduction

Many common daily activities such as eating, getting dressed or driving depend on the ability to combine visuoperceptual and motor skills. In particular, visual information often has to be continuously integrated with motor output – this is referred to as visuomotor integration. This is similar to the concept of hand-eye coordination which describes the ability to use visual information to guide hand movements, for example when reaching or grasping objects. Visuomotor integration encompasses situations where the visual feedback is either direct or indirect. Direct visual feedback means that the visual information required to carry out the movement can be gathered by observing the movement directly. For example, when someone is hand-writing or using a fork to pick up food, they watch their hand as they complete the action. Indirect visual feedback means that the visual information is derived from an indirect source. For example, when someone is using a computer mouse, they look at the screen rather than the mouse itself and continuously interpret the movement of the pointer on the screen to guide their hand movements, or when reversing a car, they may look in the mirrors to guide their movement of the steering wheel. Visuomotor integration is difficult to assess using paper-and-pencil tests but is a good candidate for novel computerised assessments that allow real-time capture of movement data.

Visuoperceptual difficulties can be a varied, complex and debilitating aspect of many types of dementia, for example causing problems with detecting movement, recognising objects, and judging distances (Alzheimer's Society, 2016b). The posterior parietal cortex is understood to be particularly important for visuomotor integration as it is involved in spatial perception and coordinating information about eye and hand movements through its connections with the frontal cortex (Tippett and Sergio, 2006), and it is vulnerable to early damage in AD (Hawkins and Sergio, 2014). Visuomotor integration has received relatively little attention in AD research but there is some evidence that it may become impaired early in the disease process. Several studies using a visuomotor integration task with nonstandard visual feedback (e.g. rotated with respect to the plane of movement) have reported impairments in AD patients, even those with otherwise mild cognitive deficits (Tippett and Sergio, 2006; Tippett, Krajewski and Sergio, 2007; Tippett, Sergio and Black, 2012) and in those at high risk of AD (strong family history or MCI) (Hawkins and Sergio, 2014). Other studies have observed that AD patients are slower to initiate and carry out goal-directed hand movements compared to controls (Verheij *et al.*, 2012; de Boer *et al.*, 2016) and less likely to move their eyes in an anticipatory manner when carrying out a sequential tapping task (Verheij *et al.*, 2012).

There has been little investigation to date of whether visuomotor integration tasks could be sensitive markers of subtle cognitive decline in preclinical AD, but one recent study with a small sample size found that cognitively-normal individuals with CSF evidence of elevated amyloid deposition (n=19) performed slower than controls (n=47) on a reaction time task with a visuomotor component (Mollica *et al.*, 2017). Interestingly, the presence of abnormal amyloid and tau biomarkers has been reported to predict decline in an annual on-road driving test in cognitively-normal individuals (Roe *et al.*, 2017, 2018).

A circle-tracing task has detected subtle impairments in visuomotor integration in presymptomatic carriers of the mutation for Huntingdon's disease – an autosomal dominant inherited disorder with progressive motor, cognitive and psychiatric symptoms – up to a decade before estimated symptom onset (Say *et al.*, 2011). Compared to healthy controls, presymptomatic gene carriers traced less accurately, particularly when the visual feedback was indirect (see Figure 8-1). I previously analysed the results of a study which administered this same circle-tracing task to presymptomatic FAD mutation carriers and found a similar result, with presymptomatic carriers tracing less accurately than controls (Macpherson *et al.*, 2017). This circle-tracing task was included in the Insight 46 battery to investigate whether it might be sensitive to subtle visuomotor integration deficits associated with amyloid deposition.

If it is to have potential as a useful marker of cognitive decline, it is important to be able to account for other factors which predict differences in performance between individuals, such as sex and education. To my knowledge, no studies have investigated the effects of these factors on the specific circle-tracing task used here, but sex differences were assessed on a different circle-tracing task which involves pursuing a moving target around a circle, with the finding that females were less accurate than males in some age groups but not in other age groups (Stirling *et al.*, 2013). Studies of associations between visuomotor integration and educational attainment have tended to focus on the development of handwriting skills in children, since handwriting requires visuomotor integration and is an important skill for performance at school (e.g. (Van Hoorn *et al.*, 2010)), but effects of educational attainment on visuomotor integration in later life have received little attention. Insight 46 offers a novel opportunity to assess potential effects of childhood cognitive ability, education and adult socioeconomic position on visuomotor integration at age 70.

Another factor which is important to consider when interpreting the results of visuomotor integration tasks is the impact of speed-accuracy trade-offs, which have already been discussed in the context of the Choice Reaction Time task (see Chapter 6). Speed-accuracy trade-offs are highly relevant to motor tasks and have been studied in the context of sport, for example to optimise bowling in cricket (Freeston and Rooney, 2014).

Speed-accuracy trade-offs have been observed on the circle-tracing task previously, with the participants who traced fastest being more error-prone (Vaportzis, Georgiou-Karistianis and Stout, 2014).

Following the structure of previous chapters, the aims of this study were to understand patterns of performance on the circle-tracing task, characterise associations between task performance and demographic and life-course predictors, and investigate associations between performance and biomarkers of brain pathologies among cognitively-normal participants.

8.2. Methods

8.2.1. Stimuli and Procedure

The circle-tracing task was presented on a Lenovo ThinkPad X61 tablet laptop, placed horizontally on the table in front of the participant, with an additional vertically placed monitor behind it for the indirect condition (Figure 8-1). The monitor was at a distance of about 60cm from the participant. On the laptop, the display showed a 90mm diameter circle and 5mm thick annulus; on the monitor, the display showed a 143mm diameter circle and 9mm thick annulus. Although the actual size of the circle was larger on the monitor than the laptop, the visual angle was comparable – about 13^{o} in each case – because the monitor was at a greater distance from the eyes.

Participants were instructed to trace clockwise round the circle using a stylus as quickly and accurately as possible (trying to stay within the annulus) without leaning their hand on the screen, starting from the vertical apex of the circle. A thin blue line appeared on the display to show their tracing path. In the direct condition, participants could see their hand and the path they were tracing on the tablet screen (Figure 8-1). In the indirect condition, the laptop was covered by an upturned box with the front open to allow the participant to put their hand inside. The participant wore a long cape which completely covered both their arms so they had no direct visual feedback while they were drawing, but they could view a copy of the circle and the tracing path on the monitor (Figure 8-2).

The length of each trial was 45 seconds. There were three trials of each feedback condition, administered in the order Direct, Indirect, Indirect, Direct, Direct, Indirect. Two practice trials – one direct and one indirect – were administered approximately one hour before the main experiment, so that participants could familiarise themselves with the procedure, but with a long enough delay to mitigate against immediate practice effects. Some previous studies have randomised or counterbalanced the order of presentation

of trials (Lemay *et al.*, 2005; Say *et al.*, 2011) but others have administered them in a fixed order to allow participants to start with an easier direct trial and build up experience for the more difficult indirect trial (Vaportzis, Georgiou-Karistianis and Stout, 2014; Vaportzis *et al.*, 2015b).

The task was administered in a dual-task format with a concurrent task of serial subtraction, as has been done in previous studies (Vaportzis, Georgiou-Karistianis and Stout, 2014; Vaportzis *et al.*, 2015b). Participants were asked to count backwards in threes from a given starting number while performing each circle-tracing trial. Starting numbers across the six trials were the same for all participants as follows: 99, 98, 97, 96, 95 and 94. Participants were instructed that, if they reached zero or near zero, they should begin again from the starting number. The tester wrote down the sequence of numbers called out by the participant.

In December 2016 (approximately half-way through the data collection period) the procedure was modified to add an additional two circle-tracing trials (Direct, Indirect) to the end of the experiment. These two trials were performed as a single task without serial subtraction, so that I could compare circle-tracing performance in the dual and single task conditions in a sub-sample of participants, to investigate how tracing speed and accuracy were influenced by having to attend to the concurrent subtraction task (the "dual-task cost"). Limitations of this design are discussed later (section 8.5.5).

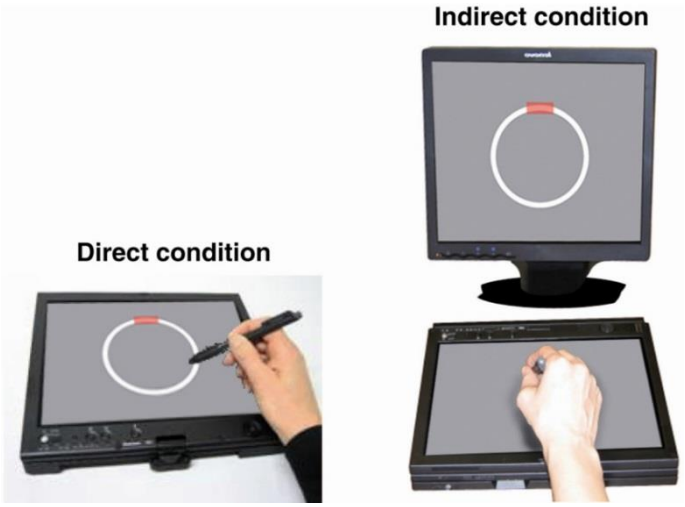


Figure 8-1. Circle-tracing apparatus

This figure is reprinted from Say et al. (2011) with permission from Elsevier. The additional box and cape for the indirect condition are not shown, but see Figure 8-2.

Figure 8-2. Box and cape to obscure the laptop in the indirect condition of the circle-tracing task

8.2.2. Outcome Variables

8.2.2.1. Circle-tracing

For each trial, the following outcome measures were derived:

i) Number of rotations completed in 45 seconds, as an index of **tracing speed**. This was recorded by the software to 14 decimal places.

ii) Number of errors per rotation, as an index of **tracing accuracy**. An error was recorded whenever the stylus deviated outside of the annulus (either beyond its inner or outer edge) for more than 100 ms.

iii) **Error time** per rotation (milliseconds) defined as the time spent tracing outside of the annulus. This variable was split by a factor of direction: when the stylus was moving away from the annulus this was classified as error detection, when the stylus was moving back towards the annulus this was classified as error correction.

iv) For the sub-sample of participants who completed the additional 'single task' trials (circle-tracing without concurrent serial subtraction), **dual-task cost** for tracing speed was calculated, which is a measure of the extent to which tracing speed was compromised by having to attend to the subtraction task. Dual-task cost was defined as: *(single speed – dual speed) / single speed*, where speed is the number of rotations completed in 45 seconds, as above.

8.2.2.2. Serial Subtraction

For each trial, the following outcome measures were derived:

i) **Subtraction rate** (responses per second), calculated by dividing the number of responses by the duration of the trial (45 seconds).

ii) Percentage of incorrect responses, as an index of **subtraction accuracy.**

8.2.3. Hypotheses

Based on the results of previous studies (Lemay *et al.*, 2005; Say *et al.*, 2011; Vaportzis, Georgiou-Karistianis and Stout, 2014; Vaportzis *et al.*, 2015b), I expected that the indirect condition (indirect visual feedback) would be associated with slower tracing, reduced accuracy, and a disproportionately longer time spent on error detection relative to error correction. I also expected to observe speed-accuracy trade-offs whereby participants who traced more quickly would tend to deviate outside the circle more often.

Given that education is associated with better performance on most cognitive tasks across a wide range of cognitive domains, and analyses from the NSHD have consistently reported associations between childhood cognitive ability, education and cognition during adulthood (Richards and Sacker, 2003; Richards *et al.*, 2019), I anticipated that higher childhood cognitive ability and educational attainment would be associated with better performance on this task. However, because the two main outcomes are in competition with each other (speed vs. accuracy), "better performance" is difficult to define and I did not have specific hypotheses about the effects of these predictors on individual outcome measures.

I aimed to test the hypothesis that cognitively-normal Aβ+ participants would show evidence of subtle impairments along the same lines as pre-symptomatic Huntington's gene carriers i.e. disproportionately poorer tracing accuracy in the indirect condition (Say *et al.*, 2011).

Although the experimental design was not ideal for measuring "dual-task cost" (see discussion in section 8.5), I hypothesised that Aβ+ participants would have a greater dual-task cost than Aβ- participants.

8.2.4. Participants

483 participants completed the dual-task circle-tracing (see section 3.3), of whom 209 also completed the single-task circle-tracing. Participant characteristics are reported in section 3.6. The proportion of participants who were right-handed was 91%.

8.2.5. Data processing

8.2.5.1. Circle-tracing

The data were generated in xml format (one file per participant) and I processed them using the HD-CAB Data Analyzer software application PxAnalyze, which produced a single spreadsheet of outcome variables for all participants.

Certain trials were excluded from analysis according to the following criteria:

i) A 'tracing time' variable was generated as part of the software output; it should be 45 seconds (i.e. the duration of the trial) but could be less if the trial was not completed perfectly, for example if the participant lifted the pen off the screen during the trial, they pressed too hard with the stylus such that no trace was

recorded temporarily, or they failed to begin tracing immediately upon initiating the trial. The maximum time in this dataset was exactly 44 seconds rather than 45 seconds as expected. I corresponded with the research group in Monash University who wrote the software and they did not have an explanation for this discrepancy, as the author of the source code was not available. I chose a threshold of three standard deviations below the mean (34.7 seconds) as a minimum acceptable tracing time. Thirty-two participants had one trial excluded for being below this threshold, an additional three participants had two trials excluded, and an additional one participant had four trials excluded (Figure 8-3). After these exclusions, the number of participants with usable single-task data dropped to 208 and the number with usable dual-task data remained at 483.

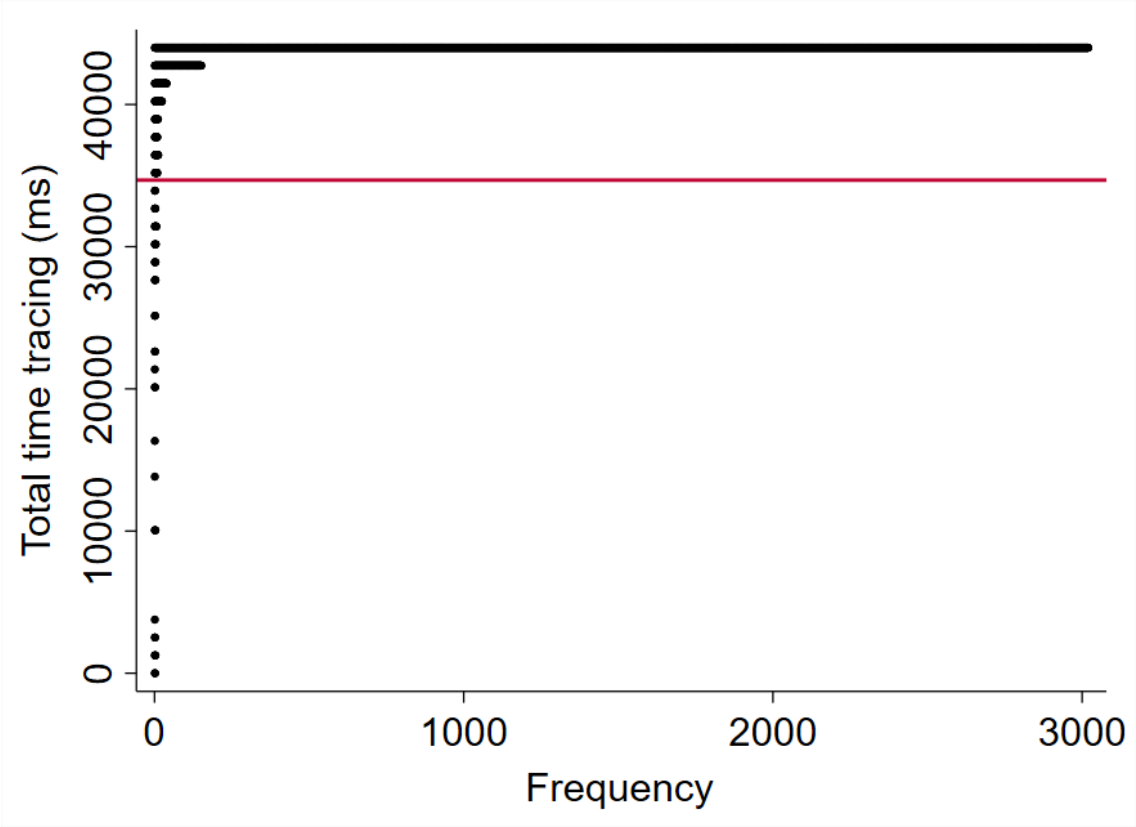


Figure 8-3. Dot plot showing total tracing time on the circle-tracing task

Red line shows the threshold below which trials were excluded (3 standard deviations below the mean).

ii) An additional two participants had one indirect trial excluded because the output showed a negative number of rotations (-0.07 and -0.08). I corresponded with the research group in Monash University who wrote the software and they did not have an explanation for these odd results. I considered if this could be caused by initially tracing anti-clockwise out of the red box (see Figure 8-1) but when I tested this myself, the results were normal. However, I identified that if all rotations are

completed anti-clockwise (contrary to instructions) this outputs a negative number. I highly doubt that any testers would have failed to correct a participant who was tracing anti-clockwise, but it is possible. It is slightly concerning not to know how these results arose, but it only affects 0.07% of trials.

iii) Two previous studies using the circle-tracing task excluded trials where the speed or error rate was more than 3.5 standard deviations from the individual's mean (Vaportzis, Georgiou-Karistianis and Stout, 2014; Vaportzis *et al.*, 2015b). However, they used 18 trials lasting 20 seconds each, whereas this task uses 6 trials lasting 45 seconds each (plus two additional trials for those who completed the single-task). Given the smaller number of trials and the fact that speed and accuracy vary between the direct and indirect conditions, it would be inappropriate to consider each participant's six trials as a distribution. Therefore, I decided not to follow such a method for data cleaning, but simply to check that none of the speed or error rates looked impossible. The number of rotations ranged from 0.37 to 50.62 and the number of errors per rotation ranged from 0 to 49, which are all plausible.

iv) An additional issue was considered, which could affect the validity of analyses comparing error detection and error correction. Sixteen participants had one trial where they were recorded as having traced outside of the annulus only in one direction, e.g. they recorded an 'error detection time' for tracing away from the annulus, but they have no corresponding 'error correction time'. The most likely explanations for this are either that the participants lifted the pen off the screen and placed it back inside the annulus, or that they had a temporary issue with the tracing pressure such that there was a break in the line due to them pressing too hard or too lightly with the stylus. I decided not to exclude these trials because the data probably reflect the participant's true performance, rather than being due to a software error. Also, other trials could have been affected by the same irregularities in tracing but within the annulus, which is not apparent from the software output.

In line with the policy of keeping the sample as representative as possible, the only reason to exclude participants with outlying performance was deemed to be a clear indication that they deviated from the protocol e.g. a fundamental misunderstanding of the instructions. To identify any participants whose performance was outlying overall, I calculated each participant's mean speed and error rate, then applied transformations to reduce the skew of the distributions (log-transform for speed, square-root for accuracy).

As there are trade-offs between speed and accuracy, outlying values of either could represent a valid extreme of prioritising speed or accuracy, so the most valid way to identify outliers is to consider speed and accuracy together. I plotted speed and error rate against each other, doing this separately for the dual and single tasks because I expected that the relationship between speed and accuracy might differ between the two, and because the single task was only performed on a sub-sample.

In the dual task, the two participants with the highest error rates appear to be outlying from the distribution, whereas the participants with the fastest and slowest speeds appear to conform to the normal pattern of speed-accuracy trade-offs (Figure 8-4). Both outlying participants have a diagnosis of Parkinson's disease so it is reasonable to suppose that their high error rates are due to their motor symptoms. This is not grounds to exclude these participants from analyses because they performed the task according to the instructions and they are not fundamentally different to some other participants with milder motor symptoms (e.g. other forms of tremor) whose performance was not outlying but was affected by their symptoms nonetheless.

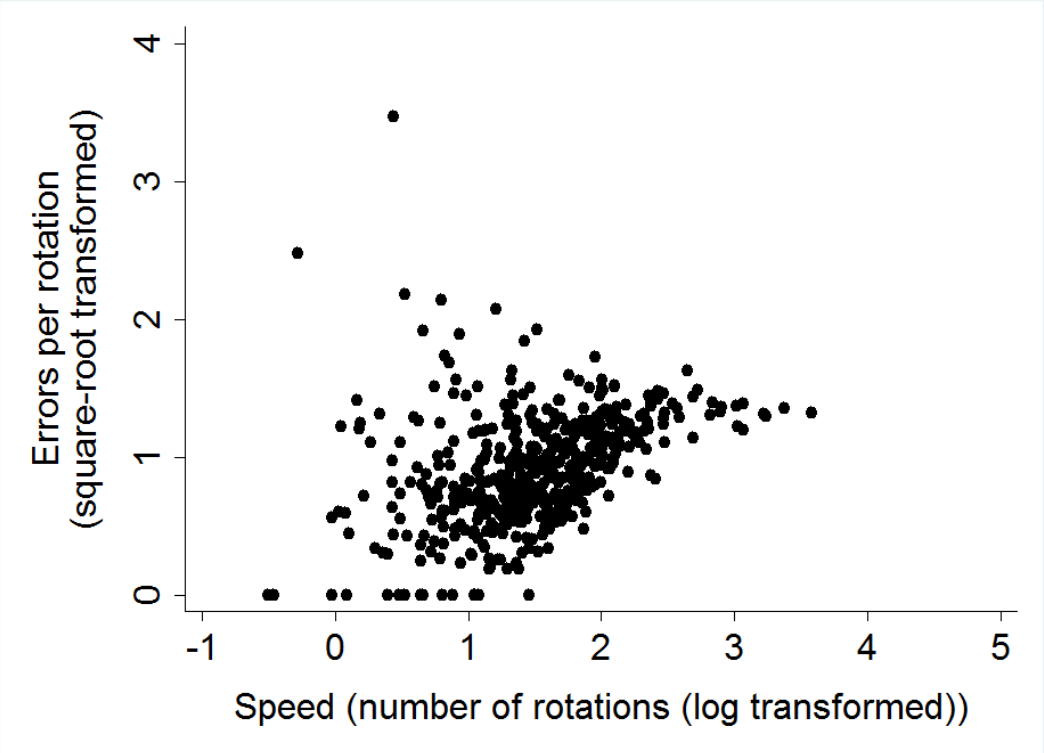


Figure 8-4. Speed-accuracy trade-offs on the circle-tracing dual task

Each participant's mean speed is plotted against their mean error rate. This graph is presented on the same scale as Figure 8-5 to facilitate visual comparison.

In the single task, two participants appeared to be outlying from the distribution in terms of high performance, as they achieved unusually fast tracing with an error rate of around

average (Figure 8-5). Their results are plausible and there were no grounds to exclude them.

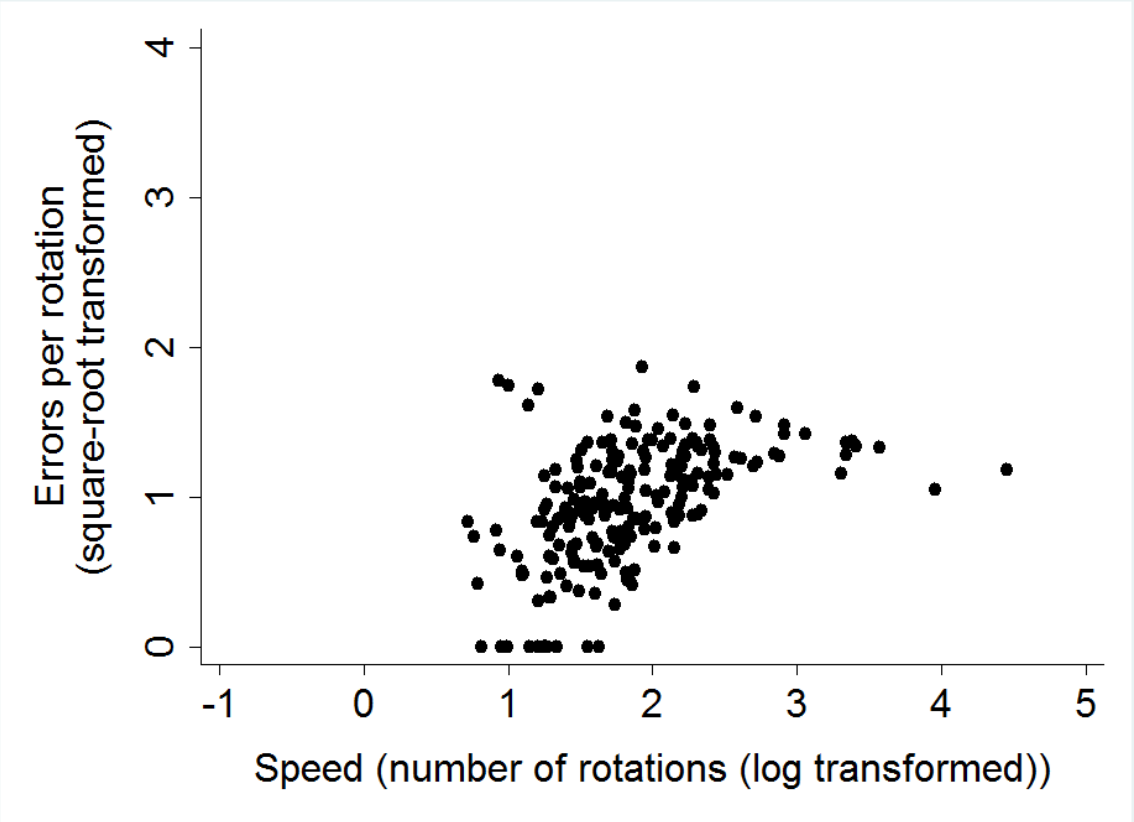


Figure 8-5. Speed-accuracy trade-offs on the circle-tracing single task

Each participant's mean speed is plotted against their mean error rate. This graph is presented on the same scale as Figure 8-4 to facilitate visual comparison.

8.2.5.1.1.Comparison of dual and single task

An ideal experimental design for comparing single vs. dual task would have equal numbers of single and dual trials, equally divided between the direct and indirect conditions, and the order of trials would be counterbalanced to control for practice effects. As our design did not achieve this (see Table 8-1), a simple average of speed or accuracy on dual vs. single trials might not be the fairest comparison because there were unequal numbers of trials (6 dual and 2 single) and the single-task trials always came at the end of the experiment where they may benefit more from practice effects. Therefore, I examined the pattern of responses across each trial before deciding which would be the best trials to use for comparison of single and dual task.

Figure 8-6 shows the mean speed and accuracy for each trial, for the 208 participants with usable data for both the dual and single tasks. Tracing speed improved on

successive trials within both the direct and indirect conditions of the dual-task, but the gains were relatively small in magnitude so an average of speed across the dual-task might be a fair reflection of tracing speed ability (Figure 8-6A). However, error rate was much more variable from one trial to another, specifically in the indirect condition. The highest error rates on the dual task were recorded on the two indirect trials that immediately followed a direct one (trials 2 and 6), suggesting that switching to the indirect visual feedback was initially difficult and required some adjustment each time, whereas performance was much improved when completing a second consecutive indirect trial (trial 3) (Figure 8-6B). Therefore, I decided that a fair comparison between the dual and single task should use trials which are matched for these factors as much as possible.

Table 8-1 Order and description of circle-tracing trials

Trial Number	Condition	Task
1	Direct	Dual
2	Indirect	Dual
3	Indirect	Dual
4	Direct	Dual
5	Direct	Dual
6	Indirect	Dual
7	Direct	Single
8	Indirect	Single

I considered the following options for comparing the dual and single tasks.

Option 1: compare trials 1 and 2 against 7 and 8

- Advantage: none of these trials are preceded by another trial in the same condition, so there would not be any confounding practice effects due to performing consecutive trials in one condition.
- Disadvantage: Practice effects across the whole experiment may depress the dual task results (trials 1 and 2) – especially in terms of accuracy as trial 2 (the first indirect trial) had a particularly high error rate.

Option 2: compare trials 5 and 6 against 7 and 8

- Advantage: Theoretically this minimises the influence of general practice effects as it uses trials from the end of the experiment.
- Disadvantage: Trial 5 (dual task, direct condition) is preceded by another trial in the same condition, whereas none of the others are. This means it may have benefitted more from practice effects, which could explain why trial 5 had faster tracing and a slightly lower error rate than trial 4 (although the difference in error rate was not statistically significant) (Figure 8-6B). This may inflate the dual task results.

Option 3: compare trials 4 and 6 against 7 and 8

- Advantage: All trials are alike in that they are preceded by a trial in the opposite condition (direct or indirect).
- Advantage: Theoretically this almost minimises the influence of general practice effects as it uses trials from towards the end of the experiment.

Option 4: Compare the average of all direct dual task trials (1/4/5) and all indirect dual task trials (2/3/6) against trials 7 and 8.

- Advantage: This might improve the reliability of the dual-task results because it would minimise the impact of any individual outliers.
- Disadvantage: By including trials from the beginning of the experiment, this fails to minimise the impact of practice effects so may depress the dual-task results. Conversely, as it includes trials 3 and 5 which benefit from being in the same condition their preceding trials, this may inflate the dual-task results.

I decided to choose Option 3 as it gives the best match between the circumstances of the dual and single-task trials, so I only included trials 4, 6, 7 and 8 in analyses comparing the dual and single tasks. Of the 208 participants who completed the dual and single tasks, nine did not have usable data for all four of these trials because their tracing time had been less than 34.53 seconds on at least one trial – see point (1) in section 8.2.5.1 – so these participants were excluded from analyses involving the single task, leaving a sample of n=199.

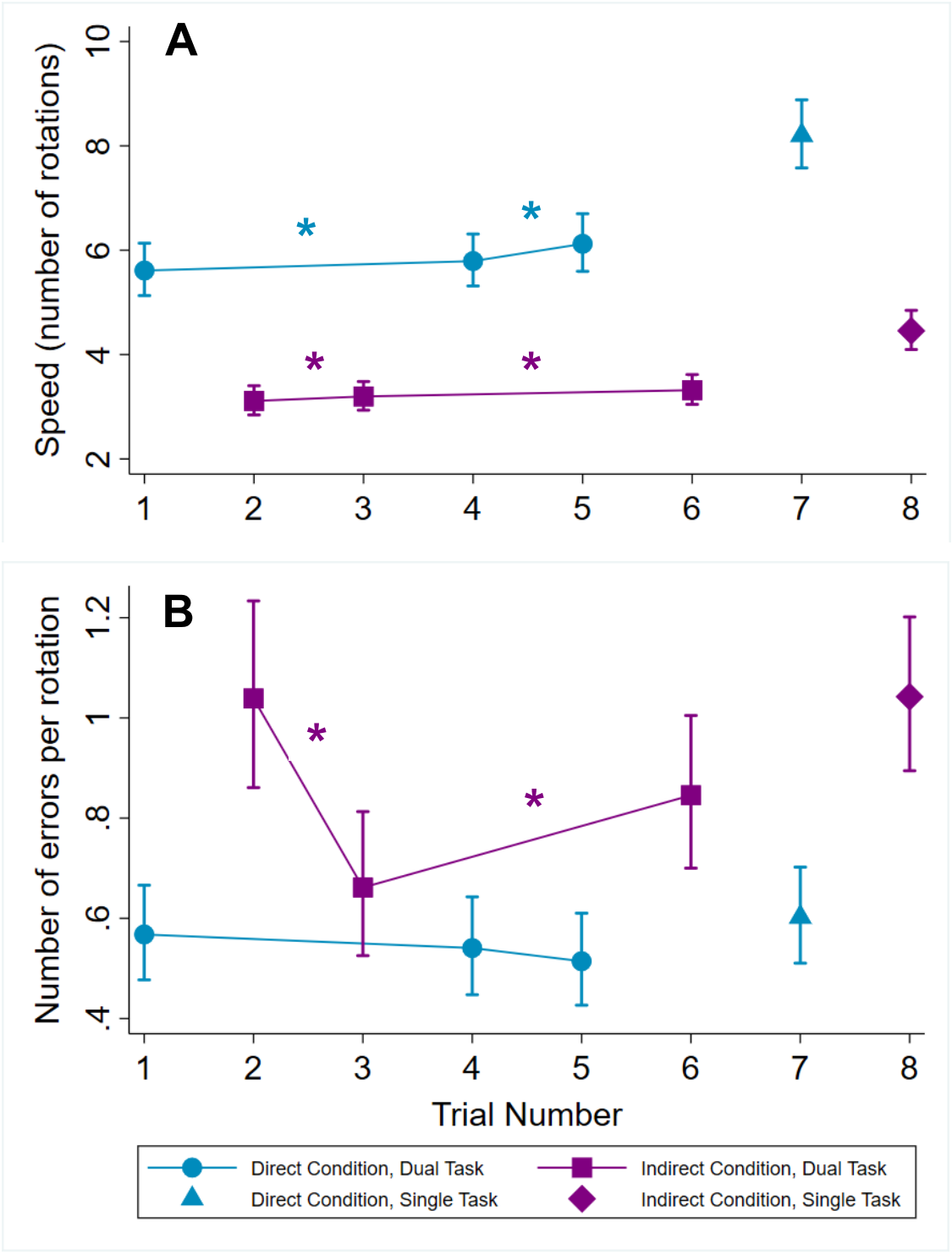


Figure 8-6. Means and 95% confidence intervals for circle-tracing speed and accuracy on each trial. A = speed. B = accuracy.

Asterisks indicate statistically significant difference in pairwise comparisons between successive dual-task trials within the direct or indirect condition ($p < 0.05$). The means and confidence intervals were generated from the transformed data (log transformed for speed, square-root transformed for error rate) as these are the variables that were used in the regression models (see 8.3.1.1.2), but the values were back-transformed before plotting on the graphs for ease of interpretation.

8.2.5.2. Serial Subtraction

When processing the serial subtraction task, I decided to exclude trials where the concurrent circle-tracing data had been excluded, as these exclusions were mostly due to the tracing time being significantly less than expected (see section 8.2.5.1) indicating a problem with performing the circle-tracing. I observed that when such problems arose, the participant invariably paused their subtraction while trying to address it. This resulted in 34 participants having one trial excluded and an additional two participants having two trials excluded. I examined the distributions of subtraction rates and error rates to check for implausible outliers. Some participants found this task very challenging, as evidenced by a minimum subtraction rate of 0.12 responses per second and a maximum error rate of 49% (to calculate these values, each participant's performance was averaged across the six subtraction trials, see section 8.3.1.2.1). This is not unexpected, especially as the dual task design meant that participants could not give the subtraction task their full attention. Notes from the testing sessions reported that some participants struggled with this task but there was no indication that they did not understand the instructions or deviated from the protocol. Therefore, no outlying responses were excluded.

8.3. Patterns and predictors of performance

Following the format laid out in section 3.5, the first part of this chapter aims to describe patterns of performance across the various outcomes and conditions of the task, and to investigate the effects of demographic and life-course predictors on performance in the full Insight 46 sample. The demographic and life-course predictors (sex, age at assessment, childhood cognitive ability, educational attainment, adult socioeconomic position and presence of a neurological or psychiatric condition) are defined in sections 3.2.4 and 3.2.3 respectively.

8.3.1. Statistical Analyses

8.3.1.1. Circle-tracing

8.3.1.1.1. Speed and accuracy

All main circle-tracing analyses were carried out using the data from the six dual-task trials completed by 483 participants.

To improve data distribution, a log transformation was applied to the speed variable (number of rotations) and square-root transformations were applied to the accuracy and error time measures. The transformed data were used in all statistical analyses.

Regression models were fitted for the transformed speed (number of rotations) and accuracy (number of errors per rotation) variables using GEE, assuming a normal distribution for the dependent variable and an identity link (as with standard linear regression), but including an exchangeable correlation structure and robust standard errors to allow for the correlation between repeated measures of the same participant. Predictors in the models were condition (direct / indirect), age at assessment, sex, childhood cognitive ability, adult socioeconomic position, education and presence of a neurological or psychiatric condition (yes / no). Handedness (left / right) was included as an additional covariate because anecdotally it can be more difficult to trace clockwise for left-handed people.

As I was interested in identifying factors that predict disproportionately poorer performance in the indirect condition, I tested for interactions between these predictors and circle-tracing condition.

8.3.1.1.2. Practice Effects

The order of trials (D-I-I-D-D-I) is such that the indirect trials have a higher average position in the order than the direct trials. Therefore, comparisons of circle-tracing and serial subtraction outcomes between the direct and indirect conditions may be confounded by practice effects. While this cannot be tested explicitly, exploring practice effects within the direct and indirect trials separately could give an indication of whether practice effects are generally observed on this test. Practice effects on speed and accuracy across the dual-task trials have already been described for the sub-sample of participants who also completed the single task, using pairwise comparisons of successive trials within the direct and indirect conditions (section 8.2.5.1.1). For each pair of trials, these comparisons were conducted by fitting the GEE regression models

for speed and accuracy with trial number as a predictor, also including the usual predictors of sex, age at assessment, childhood cognitive ability, education, adult socioeconomic position, handedness and presence of a neurological or psychiatric condition. These comparisons were repeated for the full sample of 483 participants.

8.3.1.1.3.Error detection time versus error correction time

A similar GEE model was used for the error time outcome, including an additional predictor of tracing direction (away from the circle / towards the circle) in order to compare detection time with correction time. The software output gives the total error time per trial, split into detection and correction, but does not provide any information about the duration of individual errors. The total error time is sufficient for making comparisons between error detection time versus error correction time, but not for drawing conclusions about whether some participants detected and corrected errors more quickly than others – to answer that question the number of errors should be taken into account. This was not done in the original paper (Say *et al.*, 2011) and therefore their conclusion that "premanifest and early Huntington's disease groups required longer to detect and correct errors (than the control group)" could alternatively be explained by the fact that the Huntington's disease groups made a greater number of errors, so naturally their total error time would be greater even if each individual error was of a comparable duration to the errors made by the control group. To exclude this alternative explanation, I adjusted for the number of errors per rotation (square-root transformed). In summary, the outcome was error time (ms) and the predictors were tracing direction (away / towards), number of errors per rotation (square-root transformed), condition (direct / indirect), age at assessment, sex, childhood cognitive ability, adult socioeconomic position, education, handedness, and presence of a neurological or psychiatric condition.

As for the speed and accuracy variables above, I tested for interactions between these predictors and circle-tracing condition (direct vs. indirect). I also tested for interactions between the predictors and error direction (detection / correction), to identify predictors of spending a relatively longer time on either detecting or correcting.

8.3.1.1.4.Speed-accuracy trade-offs

As discussed in the context of the Choice Reaction Time task (see Chapter 6) speed-accuracy trade-offs can be investigated between-subject or within-subject. Between-subject analyses address questions such as: "Which participants are most likely to trade speed for accuracy?" or "Which participants tend to perform well on both speed and

accuracy?". These questions are relevant to my hypotheses about whether this task can identify participants who may show evidence of subtle cognitive decline, as differences may become apparent when considering speed and accuracy together, rather than as separate outcomes. Within-subjects analyses address questions such as "Are quick responses more quickly to be inaccurate?" which was of interest in the Choice Reaction Time task, but are not appropriate in the context of these circle-tracing and serial subtraction tasks, since we do not have information on the timing of individual errors within the circle-tracing task, nor the latency of individual subtraction responses. Therefore within-subjects speed-accuracy trade-offs were not investigated.

To address the between-subjects questions for the circle-tracing task, each participant's mean number of rotations was compared to their mean number of errors per rotation using Pearson's correlation. Initially I did this separately for the direct and indirect conditions, to examine whether the pattern of speed-accuracy trade-offs differed between the two. As patterns of trade-offs were similar across both (see section 8.3.2.1.4) I calculated each participant's average speed and accuracy combined across both conditions, to more easily assess overall differences between individuals. I did this by averaging the means for the direct and indirect conditions, rather than simply averaging across all the trials – this was to avoid an unfair outcome for the participants who had one or more trials excluded (n=36, see section 8.2.5.1). For example, if a participant was missing an indirect trial, a simple average of their scores would be weighed towards direct trials (known to be easier) and the average score would look better than it should.

As speed logically precedes accuracy, I wanted to investigate whether any of the effects on accuracy investigated in the primary analyses (section 8.3.1.1.1) could be explained by speed-accuracy trade-offs. It is possible that having a higher-than-expected error rate (higher than predicted at a given speed) may be a more sensitive indicator of visuomotor integration problems than having a high error rate *per se*. A linear regression model was fitted with an outcome of mean number of errors per rotation (square-root transformed) and predictors of mean number of rotations (log-transformed), age at assessment, sex, childhood cognitive ability, adult socioeconomic position, education and presence of neurological or psychiatric condition (yes / no). Robust standard errors were used because of heteroscedasticity in the data (the variance in error rate is not constant across the speed distribution (Figure 8-4)).

8.3.1.1.5. Comparison of Dual and Single Task

In order to compare tracing speed in the dual and single tasks, I calculated each participant's "dual-task cost" in terms of tracing speed (number of rotations). Dual-task cost was defined as:

(*mean single-task speed – mean dual-task speed) / mean single-task speed*

For example, a dual-task cost of 0.4 would mean that tracing speed in the dual task was 40% slower than in the single task. As explained in section 8.2.5.1.1, only the data from trials 4, 6, 7 and 8 were used. A regression model was fitted with an outcome of dual-task cost and predictors of age at assessment, sex, childhood cognitive ability, adult socioeconomic position, education, handedness and presence of a neurological or psychiatric condition.

Dual-task cost for error rate was not calculated, since accuracy was actually higher on the dual task due to a speed-accuracy trade-off (see section 8.3.2.1.5).

8.3.1.2. Serial Subtraction

8.3.1.2.1. Speed and accuracy

A regression model was fitted for subtraction rate using GEE assuming a normal distribution for the dependent variable and an identity link (as with standard linear regression), but including an exchangeable correlation structure and robust standard errors to allow for the correlation between repeated measures of the same participant.

Subtraction error rate was analysed using a GEE logistic regression model with an independent correlation structure and robust standard errors. The outcome was the number of incorrect subtractions responses on each trial, which was treated as a proportion of the total number of subtraction responses on each trial. Results are expressed as odds ratios for ease of interpretation.

Predictors in the models were condition (direct / indirect), age at assessment, sex, childhood cognitive ability, adult socioeconomic position, education, handedness and presence of a neurological or psychiatric condition (yes / no).

8.3.1.2.2. Relationship between speed and accuracy

Each participant's mean subtraction rate was compared to their overall error rate using Spearman's correlation. Mean subtraction rate was calculated by averaging the

subtraction rate across all trials. If the same approach were used for calculating mean error rate, this would mean that errors committed on trials where subtraction was slower would be weighted more heavily than errors committed on trials where subtraction was faster; to avoid this, error rate was calculated as the total number of incorrect responses summed across all trials, divided by the total number of responses summed across all trials.

8.3.1.2.3.Practice effects

Practice effects on subtraction rate and accuracy were investigated by rerunning the models (see section 8.3.1.2) with an additional factor of trial number (1 to 6).

8.3.2. Results

Descriptive statistics for the circle-tracing and serial subtraction outcome variables are provided in Table 8-2.

Table 8-2. Descriptive statistics for circle-tracing and serial subtraction outcomes in the dual task

	Median	Interquartile range	Range
Tracing speed (number of rotations in 45 seconds)			
Direct	5.73	3.93 – 8.25	0.59 – 50.63
Indirect	3.09	2.09 – 4.46	0.37 – 36.52
Tracing accuracy (number of errors per rotation)			
Direct	0.59	0.20 – 1.16	0 – 10.27
Indirect	0.84	0.27 – 1.80	0 – 24.04
Error time per rotation (ms)			
Error detection			
Direct	85	19 - 201	0 – 5682
Indirect	149	23 - 413	0 – 13174
Error correction			
Direct	94	22 - 205	0 – 5556
Indirect	149	25 - 399	0 – 9654
Subtraction rate (responses per second)			
Direct	0.51	0.38 – 0.64	0.04 – 1.22
Indirect	0.47	0.36 – 0.58	0.07 – 1.04
Subtraction error rate (%)			
Direct	0	0 – 3.13	0 – 77.78
Indirect	0	0 – 3.70	0 – 57.14

8.3.2.1. Circle-tracing

8.3.2.1.1. Speed and accuracy

Results of the multivariable regression models for speed and accuracy are reported in Table 8-3, along with results of interaction tests between each predictor and tracing condition (direct vs. indirect).

Table 8-3. Associations between demographic and life-course predictors and circle-tracing speed and accuracy (n = 483)

Predictor	Speed (number of rotations) [a]: coefficient and 95% CIs		Accuracy (number of errors per rotation) (square-root transformed) [b]: coefficient and 95% CIs	
	Main effect of predictor	**Interaction between predictor and condition (direct vs. indirect)**	**Main effect of predictor**	**Interaction between predictor and condition (direct vs. indirect)**
Indirect condition (compared to direct as reference category)	**0.54*** (0.53, 0.56)	N/A	**0.17*** (0.12, 0.21)	N/A
Sex (female as reference category)	1.08 (0.97, 1.21)	1.01 (0.97, 1.05)	0.03 (-0.04, 0.11)	-0.04, (-0.13, 0.05)
Age at assessment (per year)	**1.15*** (1.06, 1.24)	1.01 (0.99, 1.04)	**0.07*** (0.02, 0.13)	-0.002 (0.070, 0.065)
Childhood cognitive ability (per z-score)	0.97 0.90, 1.05)	1.00 (0.98, 1.03)	-0.04 (-0.09, 0.02)	-0.05 (-0.11, 0.02)
Education (per category) [c]	0.97 (0.92, 1.02)	1.01 (1.00, 1.03)	**-0.04** (-0.07, -0.01)	0.02 (0.05 to 0.02)
Adult SEP (per category) [c]	1.02 (0.96, 1.08)	1.00 (0.98, 1.02)	-0.02 (-0.06, 0.02)	-0.04 (-0.09, 0.00)
Neurological or psychiatric condition [d] (cognitively-normal as reference category)	0.94 (0.77, 1.14)	0.98 (0.92, 1.05)	**0.22** (0.04, 0.39)	0.01 (-0.17, 0.20)
Handedness (right-handed as reference category)	0.91 (0.77, 1.08)	1.04 (0.97,1.11)	0.07 (-0.07, 0.21)	**0.42*** (0.26, 0.58)

Coefficients in bold are significant at p < 0.05 and asterisks indicate significance at p < 0.01. Multivariable regression models were used so each association is independent of all others.
[a] For the circle-tracing speed outcome, as the data were log-transformed, the coefficients are quoted in exponentiated form for ease of interpretation; for example, a coefficient of 1.5 would mean that the factor was associated with 50% faster tracing. [b] The coefficients for the square-root transformed error rate are not easily interpretable as a back-transformation would not be meaningful. [c] See section 3.2.4 for definition of categories. [d] See section 3.2.3 for definitions.

CI = confidence interval; SEP = socioeconomic position

As expected, the indirect condition was associated with slower tracing than the direct condition (54% fewer rotations in the adjusted model) and poorer accuracy (Table 8-2, Table 8-3).

On average, males traced slightly faster and made slightly more errors, but these differences were not statistically significant (Table 8-3).

There was a significant main effect of age on speed and accuracy, with older participants tracing faster (15% more rotations per year of older age) (Table 8-3) but making more errors (Table 8-3).

Higher educational attainment was not associated with speed but was associated with fewer errors per rotation (Table 8-3).

Childhood cognitive ability and adult socioeconomic position were not associated with speed or accuracy (Table 8-3).

Participants with neurological and psychiatric conditions did not differ from cognitively-normal participants in terms of tracing speed but they made more errors per rotation (Table 8-3) (unadjusted untransformed means: cognitively-normal participants = 0.79 errors per rotation, participants with neurological and psychiatric conditions = 1.02 errors per rotation).

There was no evidence of a statistically significant difference between left and right-handed participants in terms of their overall speed and accuracy, but there was an interaction between handedness and circle-tracing condition such that left-handed participants made significantly more errors per rotation in the indirect condition (Figure 8-7).

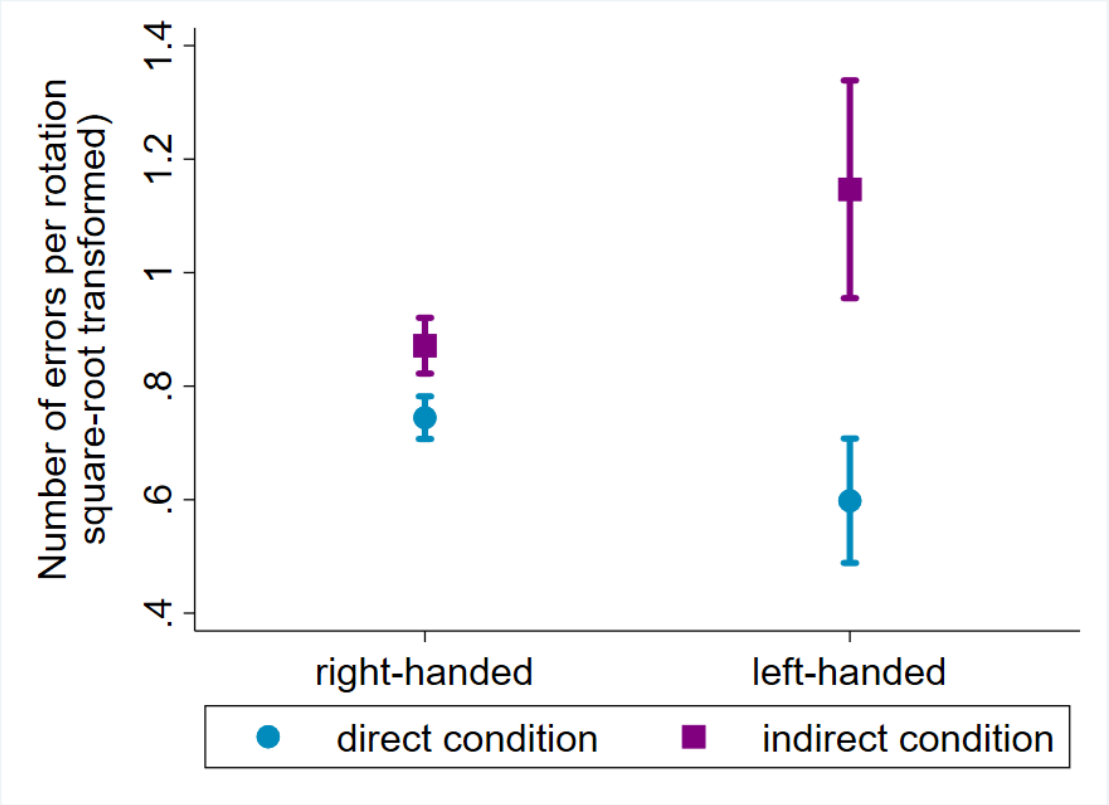


Figure 8-7. Number of errors in the direct and indirect conditions of the circle-tracing task, by handedness

There were 438 right-handed and 45 left-handed participants. Markers show the predicted means from the multivariate regression model and error bars show the 95% confidence intervals. The model included adjustment for tracing direction (away / towards), condition (direct / indirect), number of errors per rotation, age at assessment, sex, childhood cognitive ability, adult socioeconomic position, education, handedness, and presence of a neurological or psychiatric condition.

Apart from handedness, there was no evidence of interactions between any other predictors and circle-tracing condition, so the detrimental effect of the indirect condition on speed and accuracy appeared to be fairly universal.

8.3.2.1.2.Practice Effects

Practice effects on speed and accuracy across the dual-task trials have already been described for the sub-sample of participants who completed the single task (section 8.2.5.1.1). The analyses were repeated for the full sample of 483 participants and the same pattern of effects was observed, so results are not reported again but their implications are briefly discussed below.

The finding of practice effects on speed (speed improved in successive trials within the direct and indirect conditions) could affect the comparison between the direct and indirect conditions. Although the indirect condition was associated with substantially slower tracing (section 8.3.2.1.1), the true effect of indirect visual feedback could be underestimated in this study as the indirect trials had a higher average position in the trial order so would theoretically benefit more from practice effects.

8.3.2.1.3.Error detection time versus error correction time

As all analyses of error time controlled for the number of errors per rotation, the outcome is an index of how quickly participants detected and corrected errors, rather than just a correlate of how many errors were made.

As expected, error time was longer in the indirect condition than the direct condition (regression coefficient = 1.22 (ms, square-root transformed), 95% CIs 0.90 to 1.55, $p < 0.0001$). While there was no overall difference in time spent on error detection compared to correction (regression coefficient = 0.03, 95% CIs -0.06 to 0.12, $p = 0.52$), there was an interaction between condition and error direction (regression coefficient = 0.29, 95% CIs 0.12 to 0.47, $p = 0.001$). This interaction arose from opposite profiles being observed in the direct and indirect conditions: in the indirect condition participants spent a little more time on detection than correction, which is consistent with a previous study (Say *et al.*, 2011), whereas in the direct condition there was a small difference in the opposite direction (error correction taking longer than detection) (Table 8-2, Table 8-4). Previous studies have reported no difference between detection and correction time in the direct condition (Lemay *et al.*, 2005; Say *et al.*, 2011).

In light of this, the effects of each predictor on error time were analysed in the direct and indirect conditions separately and are reported in Table 8-4, along with results of interaction tests between each predictor and error direction (detection vs. correction).

Table 8-4. Associations between demographic and life-course predictors and circle-tracing error time (n = 483)

Predictor	Coefficients and 95% confidence intervals (milliseconds, square-root transformed)			
	Direct condition		Indirect condition	
	Main effect of predictor	Interaction between predictor and error direction (detection vs. correction)	Main effect of predictor	Interaction between predictor and error direction (detection vs. correction)
Error direction (detection, compared to correction as a reference)	**-0.12** (-0.21, -0.03)	N/A	**0.17** (0.02, 0.33)	N/A
Sex (female as reference category)	-0.23 (-0.67, 0.21)	-0.08 (-0.26, 0.10)	**-0.67** (-1.27, -0.08)	0.15 (-0.15, 0.46)
Age at assessment (per year)	0.10 (-0.11, 0.31)	-0.04 (-0.18, 0.09)	0.23 (-0.11, 0.58)	-0.12 (-0.34, 0.09)
Childhood cognitive ability (per z-score)	0.07 (-0.26, 0.41)	0.02 (-0.09, 0.14)	**-0.73*** (-1.19, -0.26)	**0.38*** (0.14, 0.62)
Education (per category) [a]	0.05 (-0.14, 0.23)	-0.03 (-0.11, 0.04)	0.03 (-0.22, 0.29)	-0.13 (-0.26, 0.00)
Adult socioeconomic position (per category) [a]	-0.03 (-0.20, 0.14)	0.01 (-0.09, 0.12)	0.04 (-0.27, 0.35)	-0.07 (-0.24, 0.09)
Neurological or psychiatric condition [b] (cognitively-normal as reference category)	1.66 (-0.11, 3.42)	0.03 (-0.28, 0.34)	1.64 (-0.13, 3.42)	0.38 (-0.31, 1.06)
Handedness (right-handed as reference category)	-0.16 (-0.68, 0.36)	0.04 (-0.29, 0.37)	0.60 (-0.90, 2.10)	0.05 (-0.45, 0.54)

Coefficients in bold are significant at p < 0.05 and asterisks indicate significance at p < 0.01. Multivariable regression models were used so each association is independent of all others. In addition to the predictors listed, models were additionally adjusted for number of errors per rotation (square-root transformed). The units of the coefficients are not easily interpretable as a back-transformation would not be meaningful. [a] See section 3.2.4 for definition of categories. [b] See section 3.2.3 for definitions.

There was no evidence of sex differences in error time in the direct condition but males had a shorter error time per rotation in the indirect condition (Table 8-4, Figure 8-8) suggesting that they were able to detect and correct errors slightly more quickly. There was no evidence of sex differences in the relative times spent on detection and correction in either condition (Table 8-4, Figure 8-8).

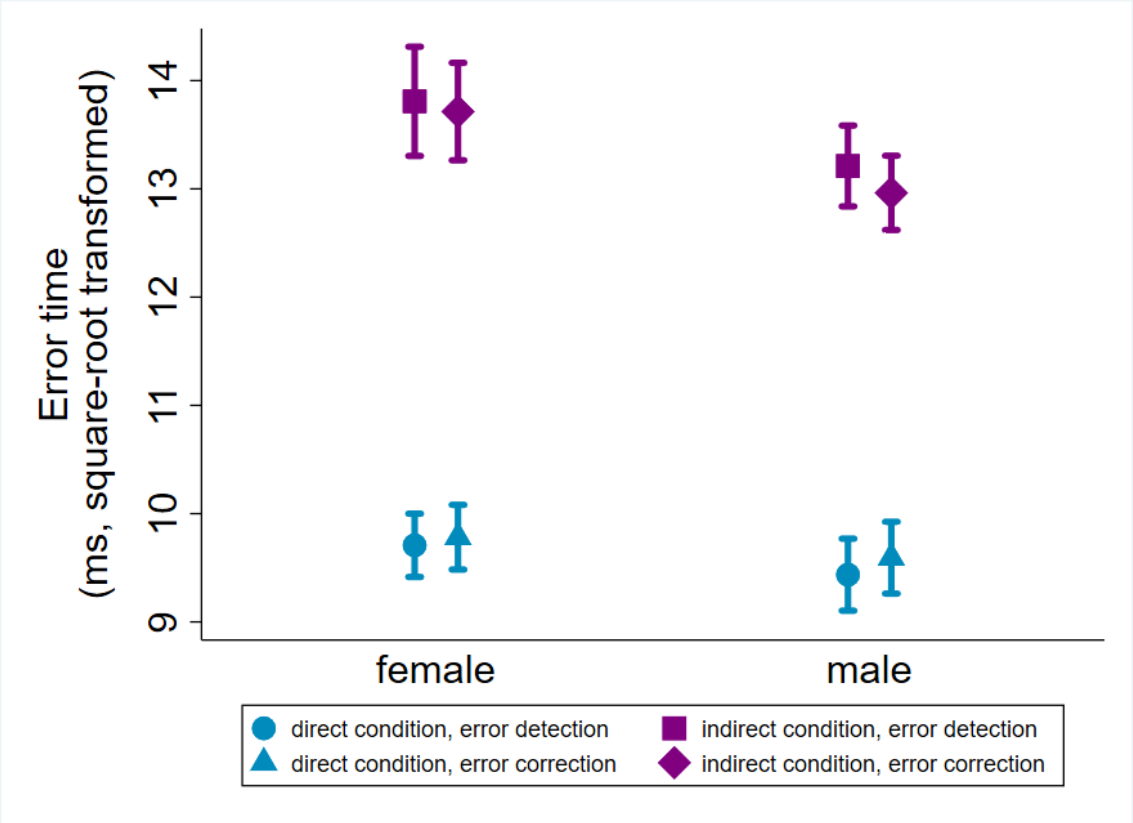


Figure 8-8. Error detection and error correction time of males and females on the direct and indirect condition of the circle-tracing task

Markers show the predicted means from the multivariate regression model and error bars show the 95% confidence intervals. The model included adjustment for tracing direction (away / towards), condition (direct / indirect), number of errors per rotation, age at assessment, sex, childhood cognitive ability, adult socioeconomic position, education, handedness, and presence of a neurological or psychiatric condition. Error time is the total time spent tracing outside the annulus during a 45 second trial, split into error detection (tracing away from the annulus) and error correction (tracing back towards the annulus).

Age was not associated with error time in either condition and there was no evidence of age effects on the relative times spent on error detection and correction in either condition (Table 8-4). This indicates that, although older participants tended to make more errors per rotation (see section 8.3.2.1.1), they did not differ in terms of the time required to detect and correct errors on average.

Childhood cognitive ability was not associated with error time in the direct condition but higher childhood cognitive ability predicted shorter error time in the indirect condition (Table 8-4). In the direct condition, there was no evidence of an interaction between childhood cognitive ability and error direction, so the tendency to spend slightly longer on error correction in the direct condition applied across the range of childhood cognitive abilities (Table 8-4, Figure 8-9). However, in the indirect condition, there was an interaction between childhood cognitive ability and error direction such that with increasing childhood cognitive ability, error detection and correction times were more similar i.e. the tendency to spend longer on error detection was reduced (Table 8-4, Figure 8-9). Overall these results suggest that, although participants with higher childhood cognitive ability did not make fewer errors overall (see 8.3.2.1.1), they were able to detect and correct their errors more quickly in the indirect condition, with a particular advantage in the time taken to detect errors.

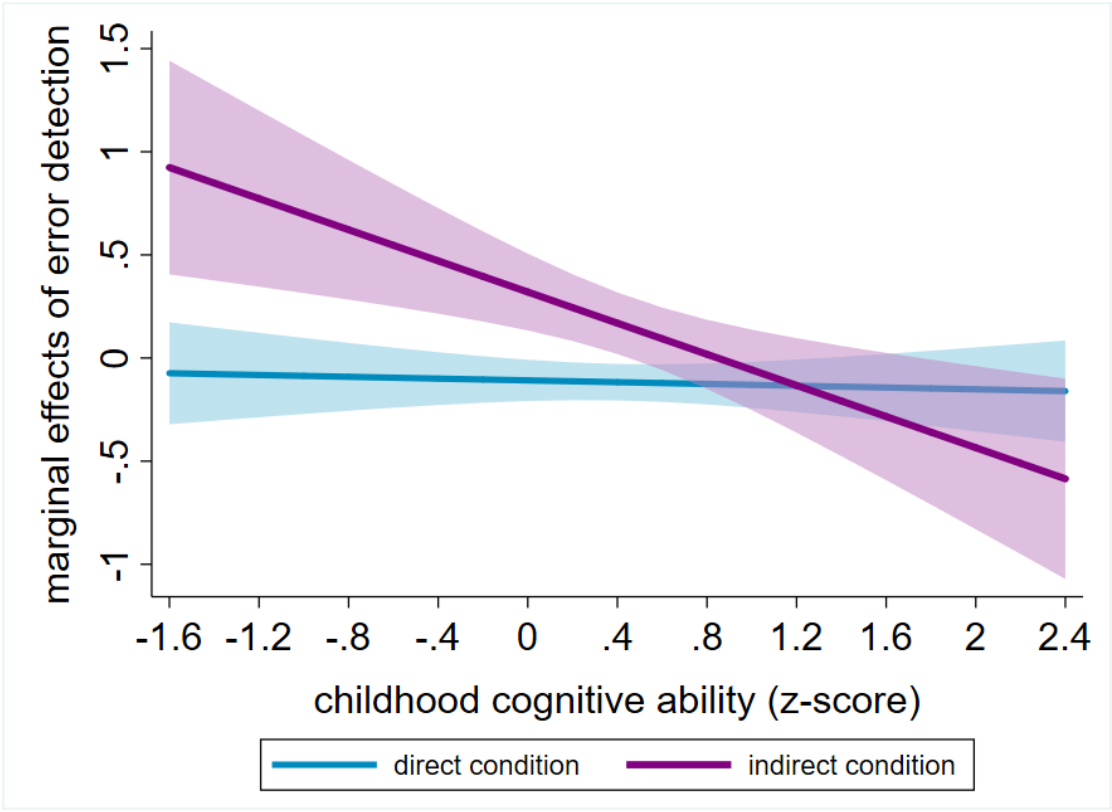


Figure 8-9. Difference between error detection time and error correction time across the range of childhood cognitive abilities, for the direct and indirect conditions of the circle-tracing task

Solid lines represent predictions from the multivariate regression models, holding all other predictors at average values, and shaded areas represent 95% confidence intervals. y=0 represents no difference between error detection and correction. Positive values of y indicate spending longer on detection than correction. Negative values of y indicated spending longer on correction than detection. The model included adjustment for tracing direction (away / towards), condition (direct / indirect), number of errors per rotation, age at assessment, sex, adult socioeconomic position, education, handedness, and presence of a neurological or psychiatric condition. For an explanation of the childhood cognitive ability variable, see section 3.2.4.

Education and adult socioeconomic position were not associated with error time in either condition and there was no evidence of these variables being associated with differences in the relative times spent on error detection and correction (Table 8-4).

Compared to cognitively-normal participants, participants with neurological and psychiatric conditions had a greater error time in both conditions, but these differences were not statistically significant (Table 8-4). They followed the same pattern as cognitively-normal participants in terms of the relative times spent on detection and correction in each condition, as there was no evidence of interactions between error direction and presence of a neurological or psychiatric condition (Table 8-4).

Handedness was not associated with error time in either condition and there was no evidence of differences between right-handed and left-handed participants in terms of the relative times spent on error detection and correction in either condition (Table 8-4). This indicates that, although left-handed participants tended to make more errors in the indirect condition (see 8.3.2.1.1), they did not differ in terms of the time required to detect and correct errors on average.

8.3.2.1.4.Speed-accuracy trade-offs

Speed-accuracy trade-offs were initially examined in the direct and indirect conditions separately. Figure 8-10 shows that a trade-off was observed in both conditions i.e. faster tracing was associated with a higher number of errors per rotation (Direct: Pearson's $r = 0.40$, $p < 0.0001$; Indirect: Pearson's $r = 0.30$, $p < 0.0001$).

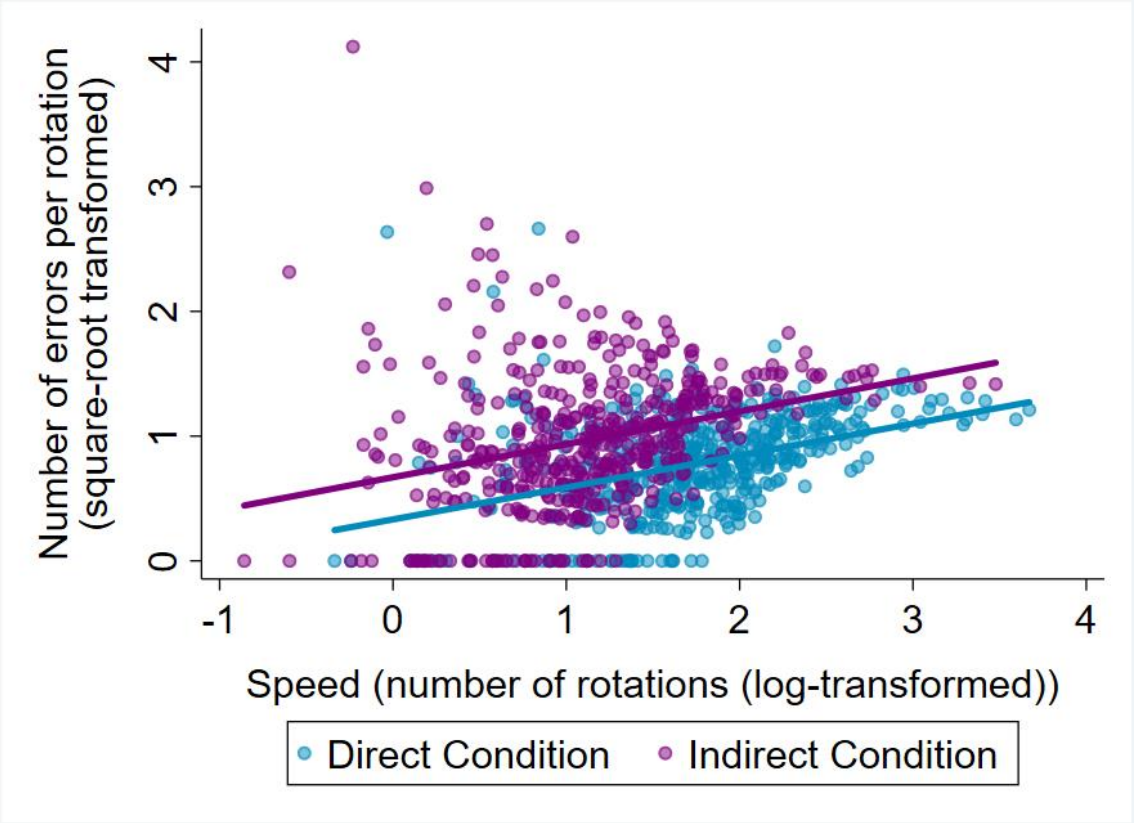


Figure 8-10. Speed-accuracy trade-offs in the direct and indirect conditions of the circle-tracing task

Each participant has one point plotted for each condition, showing their mean speed and mean error rate. Transparent markers have been used so that the density is visible where there are markers on top of each other. Solid lines indicate lines of best fit.

Examination of the residual error rates (vertical distance to the line of best fit in Figure 8-10) revealed that participants were fairly consistent between the two conditions in terms of how far their error rate deviated from that which would be predicted given their speed (Figure 8-11). This is consistent with results reported above, where there was no evidence that any particular groups of participants (according to sex, age, education etc.) performed disproportionately worse in the indirect condition (section 8.3.2.1.1). Therefore, I decided that averaging each participant's performance across the two conditions was justified as a way of examining overall between-subject differences in the tendency to prioritise speed or accuracy.

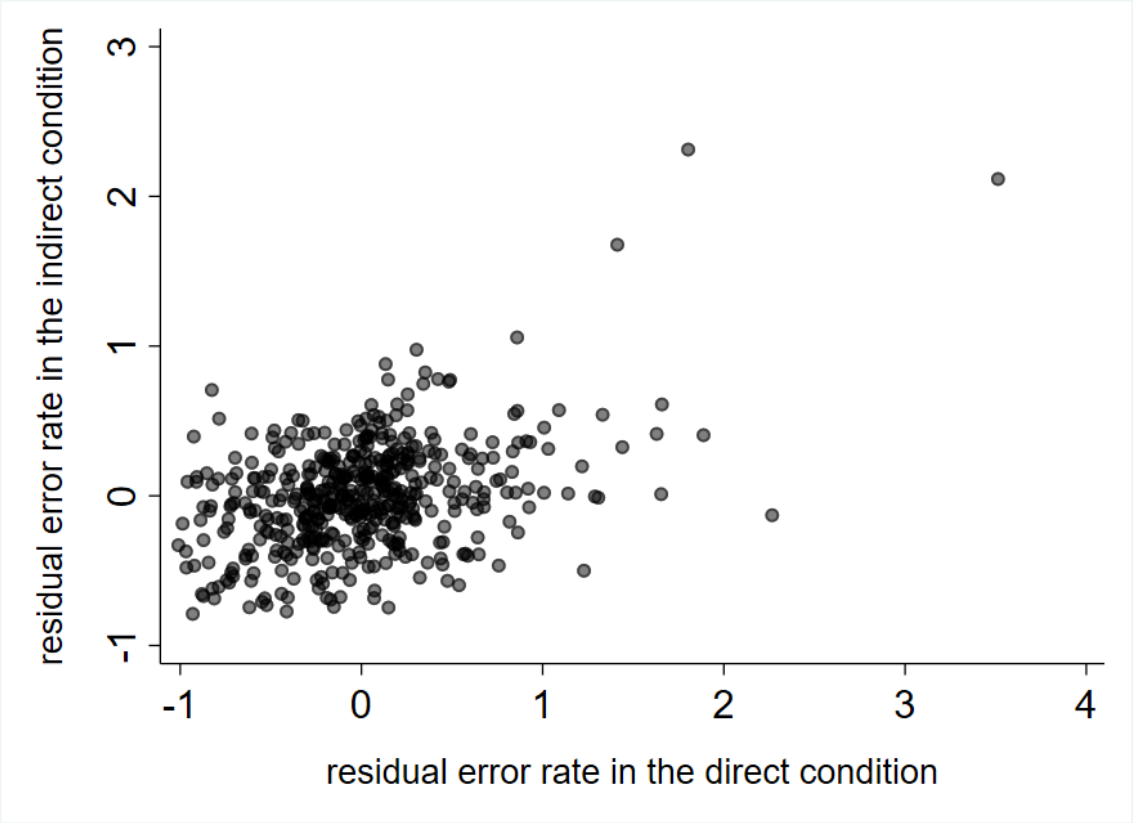


Figure 8-11. Residual error rate in the direct condition against residual error rate in the indirect condition of the circle-tracing task

Residual error rate was calculated from a linear regression of mean speed against mean error rate (see Figure 8-10). A positive residual indicates that the participant made more errors than would be predicted from their mean speed. Residual error rates in the direct and indirect conditions were correlated (Pearson's $r = 0.41$, $p < 0.0001$). Transparent markers have been used so that the density is visible where there are markers on top of each other.

Figure 8-12 shows the speed-accuracy trade-off across the direct and indirect conditions combined. The correlation appears relatively weak among participants who traced more slowly, but it is evident that none of the fastest-tracing participants managed to maintain a low error rate.

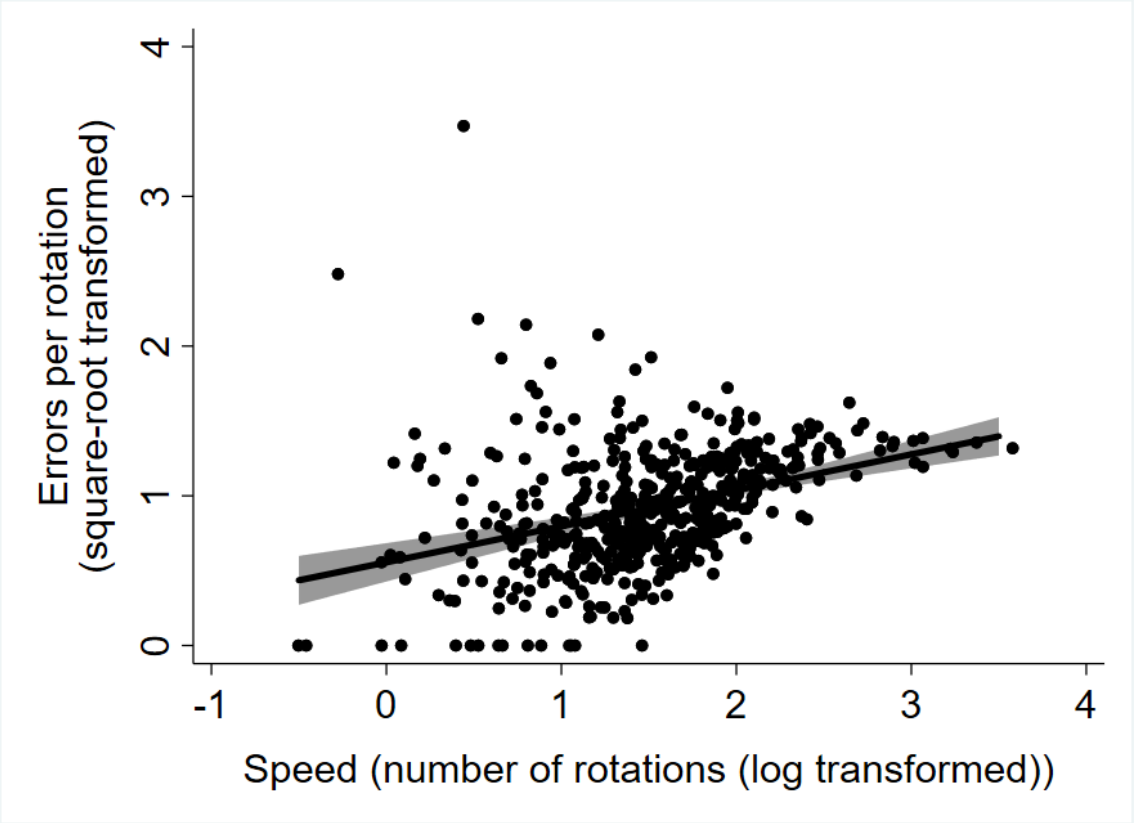


Figure 8-12. Speed-accuracy trade-offs combined across the direct and indirect conditions of the circle-tracing task

Each participant has one point plotted, showing their mean speed and mean error rate. Solid line indicates predictions from the multivariate model adjusted for sex, age at assessment, childhood cognitive ability, education, adult socioeconomic position, handedness, and presence of a neurological or psychiatric condition. Shaded areas represent 95% confidence intervals.

In the earlier analysis of circle-tracing accuracy, which did not account for tracing speed, predictors of higher error rate were lower education, older age at assessment, and presence of a neurological or psychiatric condition (section 8.3.2.1.1). Results of the regression model for error rate *accounting for speed* revealed that participants with neurological or psychiatric conditions still had higher error rates at a given speed (regression coefficient = 0.22 (number of errors per rotation, square-root transformed), 95% CIs 0.04 to 0.41, p = 0.019) and higher educational attainment was still associated with lower error rates at a given speed (regression coefficient = -0.03, 95% CIs -0.06 to -0.00, p = 0.035). However, the effect of age was much reduced and was no longer statistically significant, suggesting that the higher error rates of older participants can be explained by their faster tracing speed (regression coefficient = 0.03, 95% CIs -0.01 to 0.08, p = 0.15).

Sex, childhood cognitive ability, adult socioeconomic position and handedness were not significant predictors of error rate at a given speed.

8.3.2.1.5.Comparison of Dual and Single Task

In the sub-sample of 199 participants who completed both the dual and single tasks, the mean dual-task cost to circle-tracing speed was 0.24 (i.e. 24% fewer rotations on average in the dual task, compared to the single task). 13% of participants had a negative "dual-task cost", meaning that they traced more slowly in the single task than the dual task (Figure 8-13).

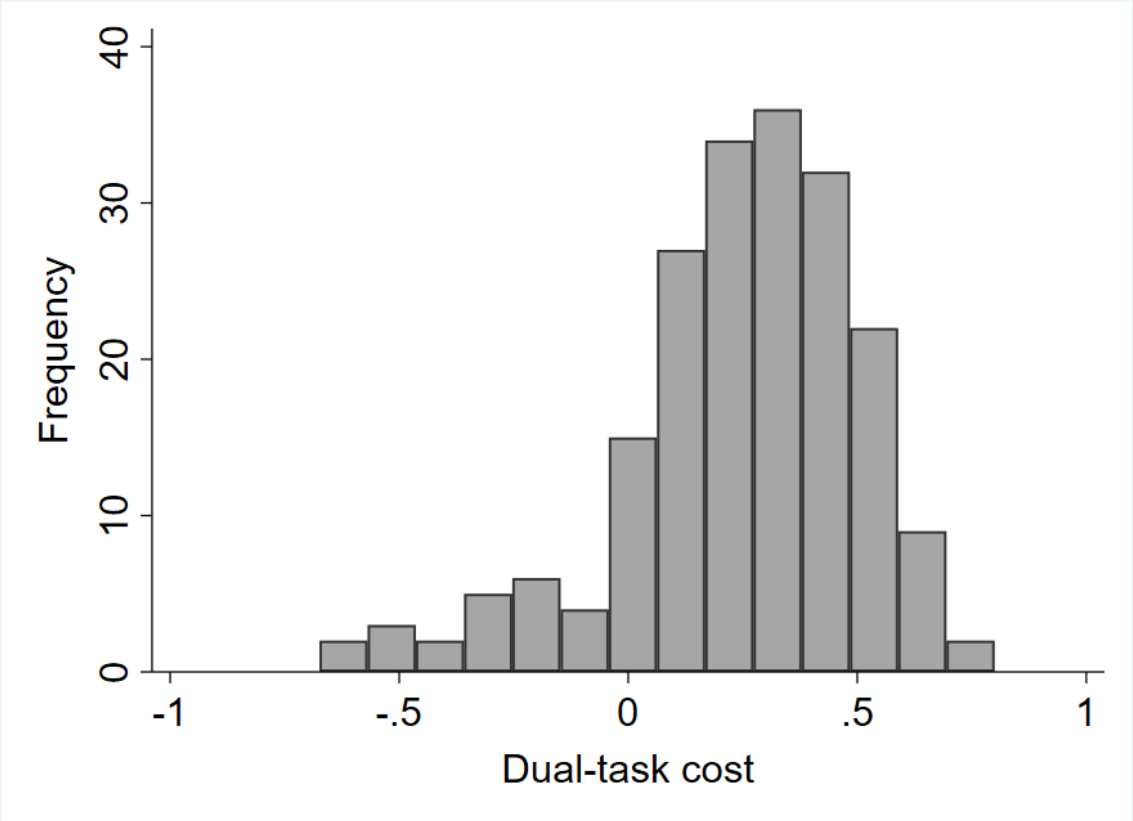


Figure 8-13. Dual-task cost to circle-tracing speed

Dual-task cost is a measure of the extent to which tracing was slower in the dual task compared to the single task (for full definition see section 8.3.2.1.5).

Dual-task cost to circle-tracing speed did not show any statistically significant associations with sex, age at assessment, education, adult socioeconomic position, handedness or presence of a neurological or psychiatric condition (Table 8-5). There was a trend towards an association between higher childhood cognitive ability and smaller dual-task cost, suggesting that participants with higher childhood cognitive ability might be less impacted by having to perform concurrent serial subtraction (Table 8-5).

Table 8-5. Associations between demographic and life-course predictors and dual-task cost (n =208)

Predictor	Coefficient (95% confidence interval)	*p*
Sex (female as reference category)	0.03 (-0.04, 0.11)	0.41
Age at assessment (per year)	0.05 (-0.08, 0.18)	0.44
Childhood cognitive ability (per z-score)	-0.06 (-0.12, 0.00)	0.06
Education (per category) [a]	0.02 (-0.02, 0.05)	0.33
Adult socioeconomic position (per category) [a]	0.01 (-0.03, 0.05)	0.65
Neurological or psychiatric condition [b] (cognitively-normal as reference category)	-0.08 (-0.22, 0.06)	0.27
Handedness (right-handed as reference category)	0.05 (-0.07, 0.17)	0.43

Multivariable regression models were used so each association is independent of all others.
[a] See section 3.2.4 for definition of categories. [b] See section 3.2.3 for definitions.

Dual-task cost to circle-tracing accuracy was not calculated, since error rate was actually slightly higher on the single task on average (mean errors per rotation: single = 1.08; dual = 0.99), with 53% of participants having a higher mean error rate in the single task than the dual task.

8.3.2.2. Serial subtraction

8.3.2.2.1. Speed and accuracy

Results of the multivariable regression models for speed and accuracy are reported in Table 8-6, along with results of interaction tests between each predictor and circle-tracing condition (direct vs. indirect).

Table 8-6. Associations between demographic and life-course predictors and performance on the serial subtraction task (n = 483)

Predictor	Subtraction rate: coefficient (responses per second) and 95% CIs		Odds ratio for making a subtraction error and 95% CIs	
	Main effect of predictor	**Interaction between predictor and condition (direct vs. indirect)**	**Main effect of predictor**	**Interaction between predictor and condition (direct vs. indirect)**
Indirect condition (direct condition as reference category)	**-0.04*** (-0.05, -0.04)	N/A	1.12 (0.99, 1.27)	N/A
Sex (female as reference category)	**0.10*** (0.07, 0.12)	**-0.01** (-0.02, -0.00)	**0.68** (0.51 0.92)	0.94 (0.73, 1.22)
Age at assessment (per year)	0.00 (-0.02, 0.02)	0.00 (-0.00, 0.01)	1.11 (0.90, 1.37)	0.93 (0.78, 1.10)
Childhood cognitive ability (per z-score)	**0.05*** (0.02, 0.07)	**-0.01*** (-0.01, -0.00)	**0.69*** (0.58, 0.83)	0.98 (0.85, 1.14)
Education (per category) [a]	**0.02*** (0.00, 0.03)	**-0.00** (-0.01, -0.00)	**0.84** (0.73, 0.98)	0.98 (0.90, 1.07)
Adult socioeconomic position (per category) [a]	(0.00 (-0.01, 0.02)	-0.00 (-0.01, 0.00)	1.02 (0.88, 1.18)	1.02 (0.92, 1.14)
Neurological or psychiatric condition [b] (cognitively-normal as reference category)	**-0.08*** (-0.12, -0.03)	0.009 (-0.007, 0.025)	**1.75** (1.01, 3.03)	1.08 (0.69, 1.69)
Handedness (right-handed as reference category)	-0.04 (-0.09, 0.00)	**-0.03*** -0.04, -0.01)	1.34 (0.79, 2.27)	0.72 (0.47, 1.12)

Coefficients and Odds Ratios in bold are significant at p < 0.05 and asterisks indicate significance at p < 0.01. Multivariable regression models were used so each association is independent of all others. [a] See section 3.2.4 for definition of categories. [b] See section 3.2.3 for definitions

CI = confidence interval

Participants subtracted more slowly when the concurrent circle-tracing task was in the indirect condition (indirect visual feedback) compared to direct condition (Table 8-2, Table 8-6). The effect size was small, equivalent to 1.8 fewer responses on average over the 45-second period. On average, participants made slightly more subtraction errors during the indirect condition of the circle-tracing task, compared to the direct condition, but this difference was not statistically significant (p = 0.08) (Table 8-2, Table 8-6).

Predictors of faster and more accurate subtraction were male sex, higher childhood cognitive ability, higher educational attainment, and absence of a neurological or psychiatric condition (Table 8-6). The mean values for males and females were as follows: females = 0.45 responses per second, 3.5% error rate; males = 0.55 responses per second, 2.2% error rate.

For subtraction accuracy, these effects appeared to apply equally in both the direct and indirect conditions of the concurrent circle-tracing task, as there was no evidence of interactions between the predictors and circle-tracing condition (Table 8-6). For subtraction rate, there were interactions with circle-tracing condition whereby the effects of sex, childhood cognitive ability and education were slightly reduced in the indirect condition compared to the direct condition (Table 8-6). This suggests that participants were on a more equal footing in terms of their subtraction speed when they were doing the more difficult concurrent task (indirect circle-tracing). However the detrimental effect of neurological or psychiatric conditions on subtraction rate (equivalent to 3.6 fewer responses over the 45-second period) was not reduced in the indirect condition (Table 8-6).

There was an interaction between handedness and circle-tracing condition such that left-handed participants subtracted disproportionately slowly in the indirect condition (Table 8-6, Figure 8-14). This mirrors the earlier finding that left-handed participants traced disproportionately less accurately in the indirect condition, which I believe was due to the set-up of the equipment being less convenient for them (see section 8.3.2.1.1). This provides further evidence that subtraction tended to be slower when the circle-tracing was more difficult, whether because the visual feedback was indirect or because of some other problem.

Age at assessment and adult socioeconomic position were not associated with performance on the subtraction task (Table 8-6).

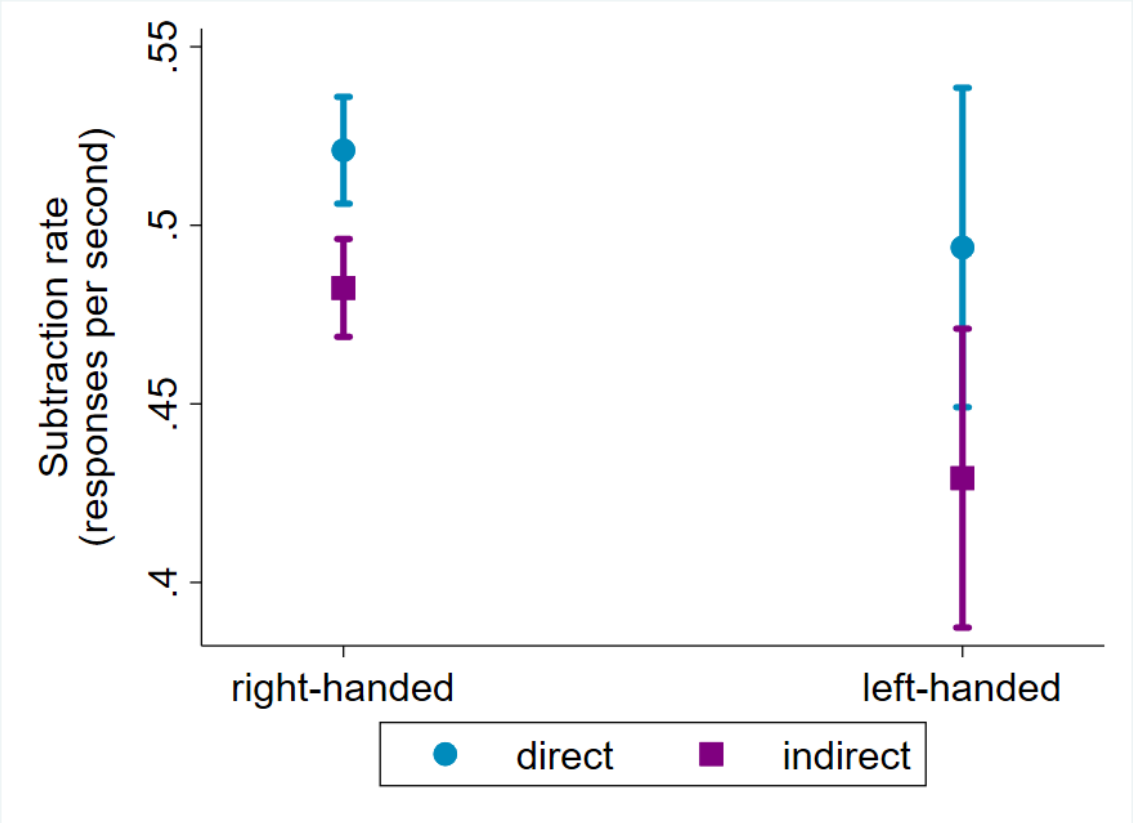


Figure 8-14. Subtraction rate of right-handed (n = 438) and left-handed (n = 45) participants in the direct and indirect conditions of the concurrent circle-tracing task

Markers show the predicted means from the multivariate regression model and error bars show the 95% confidence intervals. The model included adjustment for age at assessment, sex, childhood cognitive ability, adult socioeconomic position, education, and presence of a neurological or psychiatric condition.

8.3.2.2.2.Relationship between speed and accuracy

Faster subtraction rate was associated with lower error rate (Spearman's ρ = -0.51, p < 0.0001) i.e. the opposite of a speed-accuracy trade-off (Figure 8-15). The relationship is non-linear due to a ceiling effect on subtraction accuracy. It is evident that all of the fastest-subtracting participants also achieved low error rates, whereas participants who subtracted slowly had a wide range of error rates.

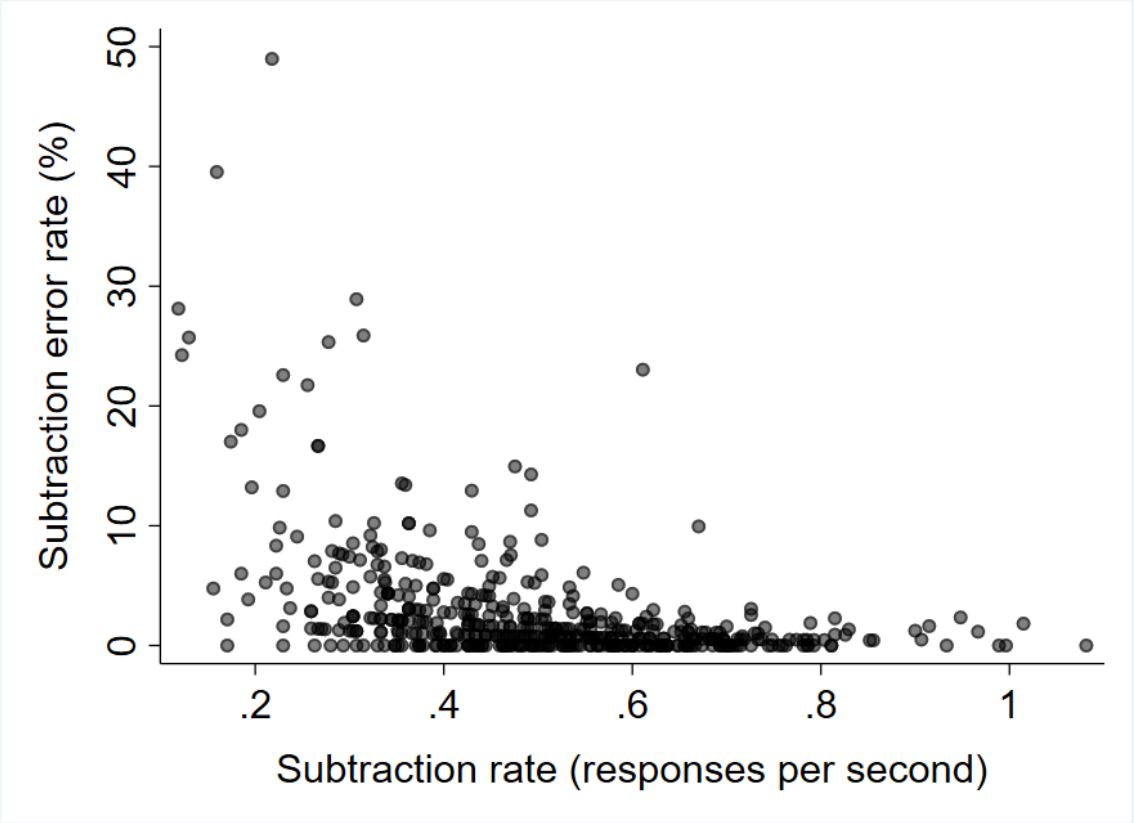


Figure 8-15. Speed against accuracy on the serial subtraction task

Each participant has one point plotted, showing their mean subtraction rate (an index of speed) and mean error rate across the six subtraction trials. Transparent markers have been used so that the density is visible where there are markers on top of each other.

8.3.2.2.3.Practice effects

Trial-by-trial performance on serial subtraction is shown in Figure 8-16. There was evidence of a small practice effect on subtraction rate, equivalent to an improvement of 0.006 responses per second on each successive trial (*95% CIs* 0.005 to 0.008, $p < 0.0001$). Therefore the finding that subtraction was slower when the concurrent circle-tracing task had indirect visual feedback 8.3.2.2.1 could be an underestimate of the true effect, as the indirect trials had a higher average position in the trial order so would theoretically benefit more from practice effects.

There was a slight improvement in subtraction accuracy on successive trials but this was not statistically significant (*OR* for making an error per successive trial = 0.97, *95% CIs* 0.93 to 1.00, $p = 0.065$).

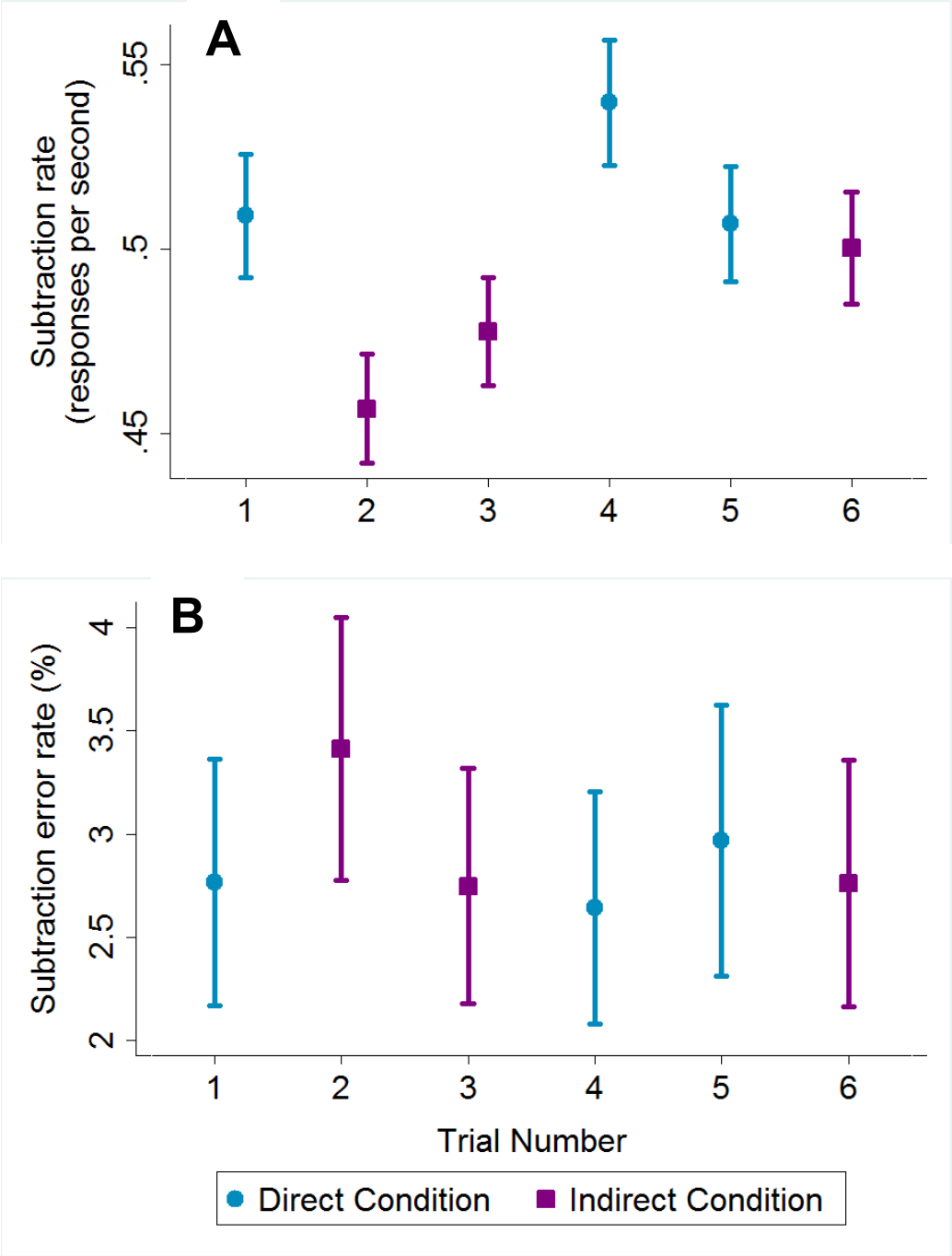


Figure 8-16. Means and 95% confidence intervals for subtraction speed and accuracy on each trial

A = subtraction rate (an index of speed). B = subtraction error rate (an index of accuracy).

8.4. Associations with biomarkers and *APOE*-ε4

Following the format laid out in section 3.5, the second part of this chapter aims to investigate associations between performance and biomarkers of AD in cognitively-normal participants for whom complete biomarker data are available. The number meeting these criteria who also had usable data from the circle-tracing task was 392 for the dual task and 165 for the single task.

As explained in section 3.5.2.1, I wanted to derive some summary scores to capture the key aspects of performance on this task. In the Choice RT, Response Inhibition and ""What was where?" tasks (see Chapters 5 to 7), it was appropriate to create summary outcomes by averaging responses across all trials (mean RT, overall identification rate etc.) because, although performance levels did vary according to the conditions of the task (e.g. in the Choice RT task, RT tended to be faster for words rather than arrows), the magnitude of these differences were small. By contrast, in the circle-tracing task, the magnitude of differences in performance between the direct and indirect conditions was large, with speed differing by approximately a factor of two. Therefore, while a simple average would still be valid as a way of comparing participants' tracing speed, it would have limited meaning as a representation of typical performance. Also, there is a theoretical reason to focus on the difference in performance between the two conditions (which would be masked by a simple average), because the outcome that was sensitive to early changes in Huntington's disease was the interaction between tracing condition and clinical group (Huntington's gene carriers were disproportionately disadvantaged in the indirect condition (Say *et al.*, 2011)). Therefore, I decided to investigate the associations between biomarkers and circle-tracing outcomes using the trial-by-trial data as I did in the previous section (8.3), and then to use these results to inform the selection of the most promising summary scores.

I focused on the speed and accuracy variables, rather than the error time variable, because speed and accuracy are the main dimension of the task and exploring the error time variable was more a way of understanding the nature of errors. I also investigated dual-task cost because deficits in dual-tasking have been reported in presymptomatic FAD mutation carriers at a stage when their episodic memory was unimpaired (MacPherson *et al.*, 2012), so it has potential as a marker of subtle cognitive change associated with preclinical AD.

8.4.1. Statistical Analysis

8.4.1.1. Circle-tracing

8.4.1.1.1. Speed and accuracy

GEE regression models were fitted as above for tracing speed and accuracy (see section 8.3.1.1.1) including predictors of amyloid status, whole brain volume, WMHV and *APOE* genotype. I tested for interactions between these factors and circle-tracing condition. As in previous chapters, total intracranial volume was included to adjust for the correlation between whole brain volume and head size, as were the demographic factors investigated in section 8.3 (sex, age at assessment, childhood cognitive ability, education and adult socioeconomic position).

The models were additionally rerun replacing dichotomised amyloid status with SUVR to test whether increasing Aβ deposition was associated with differences in performance. To check whether associations between SUVR and cognition were sensitive to the inclusion of the imputed SUVR values (see section 3.2.2), the analyses were rerun excluding the 26 participants with imputed SUVR data.

8.4.1.1.2. Comparison of Dual and Single Task

A regression model was fitted with an outcome of "dual-task cost" as above (see section 8.3.1.1.5) including predictors of amyloid status, whole brain volume, WMHV and *APOE* genotype. As above, total intracranial volume was included to adjust for the correlation between whole brain volume and head size, as were the demographic factors investigated in section 8.3 (sex, age at assessment, childhood cognitive ability, education and adult socioeconomic position). The model was additionally rerun replacing dichotomised amyloid status with SUVR, and a sensitivity analysis was run excluding the 26 participants with imputed SUVR data.

8.4.1.2. Serial subtraction

GEE regression models were fitted as above for tracing speed and accuracy (see section 8.3.1.2.1) including additional factors of amyloid status, whole brain volume, WMHV and *APOE* genotype, and tested for interactions between these factors and circle-tracing

condition. As in previous chapters, total intracranial volume was included as to adjust for the correlation between whole brain volume and head size, as were the demographic factors investigated in section 8.3 (sex, age at assessment, childhood cognitive ability, education and adult socioeconomic position).

The model was additionally rerun replacing dichotomised amyloid status with SUVR, and a sensitivity analysis was run excluding the 26 participants with imputed SUVR data.

8.4.2. Results

8.4.2.1. Circle-tracing

8.4.2.1.1. Speed and accuracy

Results of the multivariable regression models for speed and accuracy are reported in Table 8-7, along with results of interaction tests between each predictor and tracing condition (direct vs. indirect). As these regression models used log-transformed and square-root transformed data, descriptive statistics on the untransformed data for Aβ+ and Aβ- participants are provided in Table 8-8 to aid interpretation of the results. Results for the demographic and life-course factors (sex, age at assessment, childhood cognitive ability, education and adult socioeconomic position) are not reported as they are essentially unchanged from the first analysis section (8.3.2.1.1).

Table 8-7. Associations between biomarkers and circle-tracing speed and accuracy in cognitively-normal participants (n = 392)

	Speed (number of rotations) [a]: coefficient and 95% CIs		**Accuracy (number of errors per rotation)** (square-root transformed) **[b]:** coefficient and 95% CIs	
Predictor	**Main effect of predictor**	**Interaction between predictor and condition (direct vs. indirect)**	**Main effect of predictor**	**Interaction between predictor and condition (direct vs. indirect)**
β-amyloid status (negative as reference)	1.05 (0.90, 1.22)	1.05 (1.00, 1.10)	0.09 (-0.01, 0.19)	**0.16** (0.03, 0.29)
WMHV (per 10 ml)	0.91 (0.82, 1.01)	1.02 (0.98, 1.06)	0.03 (-0.05, 0.10)	**0.10** (0.00, 0.20)
Whole brain volume (per 10 ml)	**1.02*** (1.01, 1.04)	1.00 (1.00, 1.00)	-0.00 (-0.01, 0.01)	-0.00 (-0.01, 0.00)
***APOE*-ε4** (non-carriers as reference)	1.01 (0.88, 1.16)	1.04 (0.99, 1.09)	-0.03 (-0.12, 0.06)	0.03 (-0.07, 0.13)

Multivariable regression models were used so each association is independent of all others. In addition to the predictors listed, models also included sex, age at assessment, childhood cognitive ability, adult socioeconomic position, handedness and total intracranial volume.

Coefficients in bold are significant at p < 0.05 and asterisks indicate significance at p < 0.01.

[a] For the circle-tracing speed outcome, as the data were log-transformed, the coefficients are quoted in exponentiated form for ease of interpretation; for example, a coefficient of 1.5 would mean that the factor was associated with 50% faster tracing. [b] The coefficients for the square-root transformed error rate are not easily interpretable as a back-transformation would not be meaningful.

CI = confidence interval; WMHV = white matter hyperintensity volume

Table 8-8. Descriptive statistics for circle-tracing speed and accuracy, by amyloid status

Outcome variable	mean, *median*, (range)	
	amyloid positive (n = 72)	**amyloid negative (n = 320)**
Speed (number of rotations)		
Direct condition	6.8, *5.7*, (1.0 to 31.7)	6.7, 5.6, (0.6 to 44.1)
Indirect condition	3.8, *3.1*, (0.8 to 15.7)	3.6, *3.1*, (0.4 to 33.1)
Accuracy (number of errors per rotation)		
Direct condition	0.7, *0.5*, (0 to 5.5)	0.7, *0.6*, (0 to 9.6)
Indirect condition	1.4, *1.2*, (0 to 9.2)	1.1, *0.7*, (0 to 12.2)

There were no statistically significant differences between Aβ+ and Aβ- participants in terms of their overall accuracy, but the detrimental effect of the indirect condition was greater in Aβ+ participants (Table 8-7, Figure 8-17). As a post-hoc exploration of this interaction, the models were rerun in the direct and indirect conditions separately. In the direct condition alone, the difference between Aβ+ and Aβ- participants was minimal and not statistically significant (regression coefficient = 0.01 errors per rotation (square-root transformed), 95% CIs -0.10 to 0.11, p = 0.92), but in the indirect condition Aβ+ participants made significantly more errors (regression coefficient = 0.18, 95% CIs 0.05 to 0.31, p = 0.008).

When rerunning the models replacing dichotomous amyloid status with continuous SUVR, the same pattern emerged, as SUVR did not have a statistically significant association with error rate (regression coefficient = 0.50 errors per rotation (square-root transformed), 95% CIs –0.06 to 1.06, p = 0.08) but there was evidence of an interaction between higher SUVR and circle-tracing condition, with higher SUVR predicting disproportionately poorer accuracy in the indirect condition (regression coefficient = 0.88, 95% CIs 0.22 to 1.54, p = 0.009). These results were unchanged in a sensitivity analysis excluding the individuals with imputed SUVR values (see section 3.2.2).

Overall these results suggest that Aβ deposition was associated with greater difficulty in tracing accurately with indirect visual feedback, consistent with subtle deficits in visuomotor integration.

In terms of tracing speed, there was no statistically significant difference between Aβ+ and Aβ- participants, nor a statistically significant interaction between Aβ and circle-tracing condition (Table 8-7).

When rerunning the models replacing dichotomous amyloid status with continuous SUVR, there was no evidence of an association between SUVR and tracing speed (regression coefficient = 1.25, 95% CIs 0.50 to 3.09, *p* = 0.63) and there was no statistically significant interaction between SUVR and circle-tracing condition (regression coefficient = 1.26, 95% CIs 0.96 to 1.67, *p* = 0.10). These results were unchanged in a sensitivity analysis excluding the individuals with imputed SUVR values (see section 3.2.2).

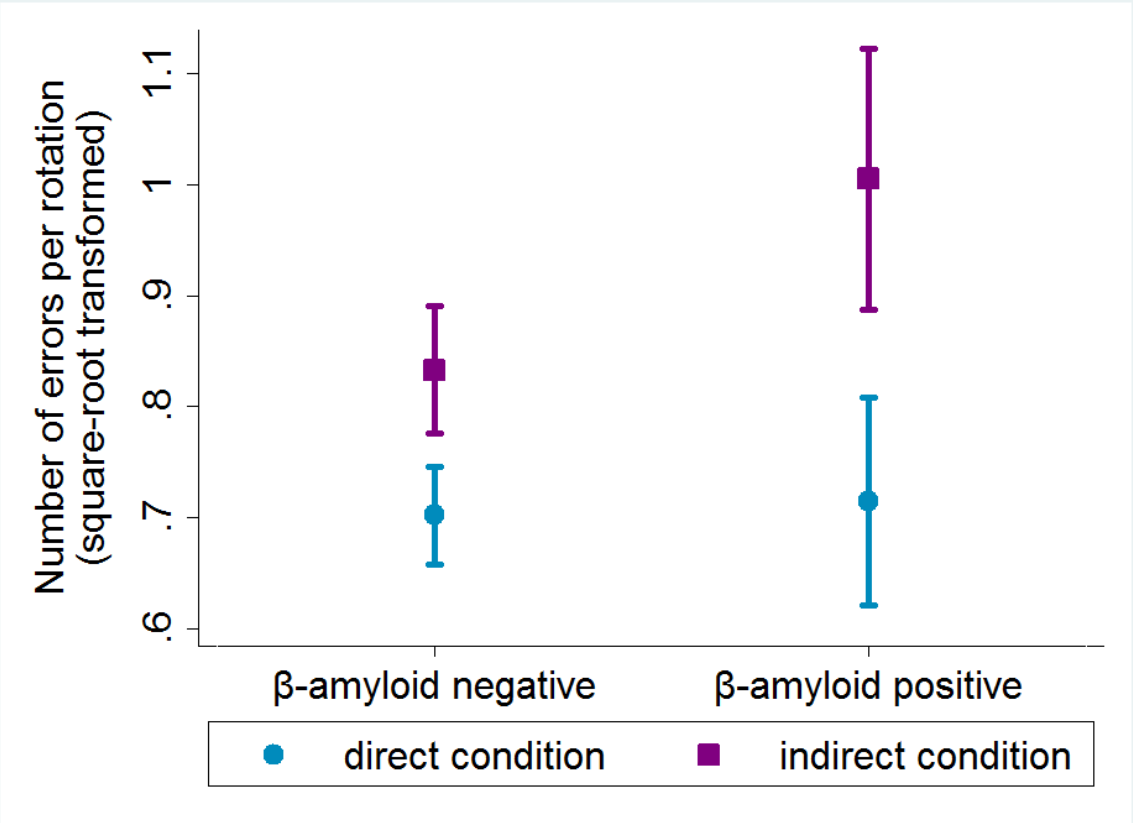


Figure 8-17. Tracing accuracy in the direct and indirect conditions of the circle-tracing task, by amyloid status

Markers show the predicted means from the multivariate regression model and error bars show the 95% confidence intervals. The model included adjustment for sex, age at assessment, childhood cognitive ability, adult socioeconomic position, education, handedness, white matter hyperintensity volume, whole brain volume, total intracranial volume and APOE genotype.

Larger whole brain volume was associated with faster tracing, equivalent to 2% more rotations per 10ml (Table 8-7, Figure 8-18). There was no evidence that this effect differed between the direct and indirect conditions (Table 8-7). There was no evidence of any relationship between whole brain volume and tracing accuracy (Table 8-7).

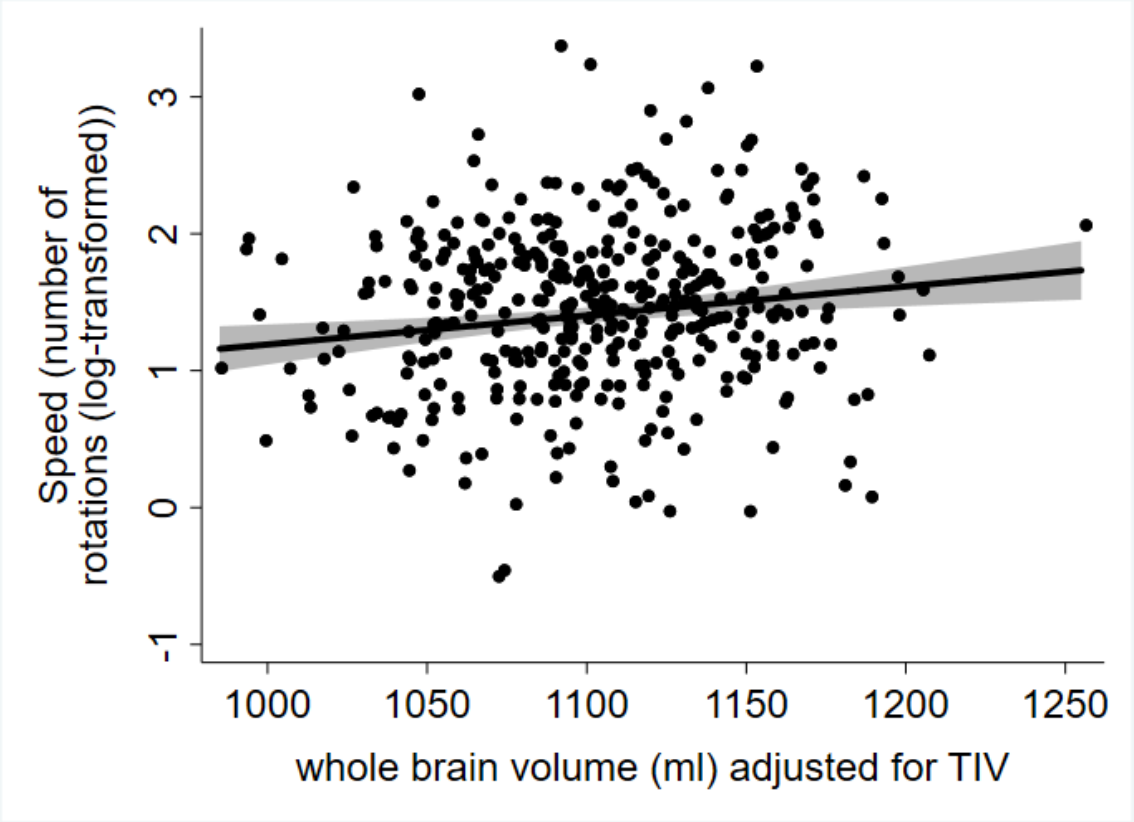


Figure 8-18. Association between whole brain volume and circle-tracing speed

Solid line represents prediction from the multivariate regression model, holding all other predictors at average values, and shaded area represents 95% confidence intervals. The model included adjustment for age at assessment, sex, adult socioeconomic position, education, handedness, amyloid status, white matter hyperintensity volume, APOE-ε4 genotype and total intracranial volume. Markers show each participant's mean speed combined across the direct and indirect conditions. TIV = total intracranial volume.

There was a trend towards an association between greater WMHV and slower tracing (p = 0.07), but no evidence that this effect differed between the direct and indirect conditions (Table 8-7). For tracing accuracy, there was no overall association with WMHV but there was evidence of an interaction between WMHV and circle-tracing condition, such that greater WMHV was associated with disproportionately poorer accuracy in the indirect condition (Table 8-7). As a post hoc exploration of this interaction, the models were rerun in the direct and indirect conditions separately, and WMHV was not associated with tracing accuracy in either condition alone (Direct: regression coefficient = -0.02 errors

per rotation (square-root transformed) per 10 ml, 95% CIs -0.10 to 0.06, p = 0.68; Indirect: regression coefficient = 0.07, 95% CIs -0.02 to 0.17, p = 0.13).

APOE-ε4 was not associated with any circle-tracing outcomes (adjusting for amyloid status and all other covariates) (Table 8-7).

8.4.2.1.2.Comparison of Dual and Single Task

Larger whole brain volume was associated with smaller dual-task cost (regression coefficient = -0.01 per 10 ml, 95% CIs -0.02 to -0.00, p = 0.023), meaning that participants with larger brain volumes tended to be less influenced (slowed) in their tracing speed by the concurrent subtraction task (Figure 8-19). The units represent proportional decrease in speed in the dual task compared to the single task (see 8.3.1.1.5) so the regression coefficient is equivalent to a reduction of 1 percentage point for every additional 10ml of brain volume.

Amyloid status, WMHV and *APOE*-ε4 genotype showed no evidence of associations with dual-task cost. When the models were rerun replacing amyloid status with SUVR, it also showed no association with dual-task cost, and this was unchanged when excluding the participants with imputed SUVR values.

Results for the demographic and life-course factors (sex, age at assessment, childhood cognitive ability, education and adult socioeconomic position) are not reported as they are essentially unchanged from the first analysis section (8.3.2.1.5).

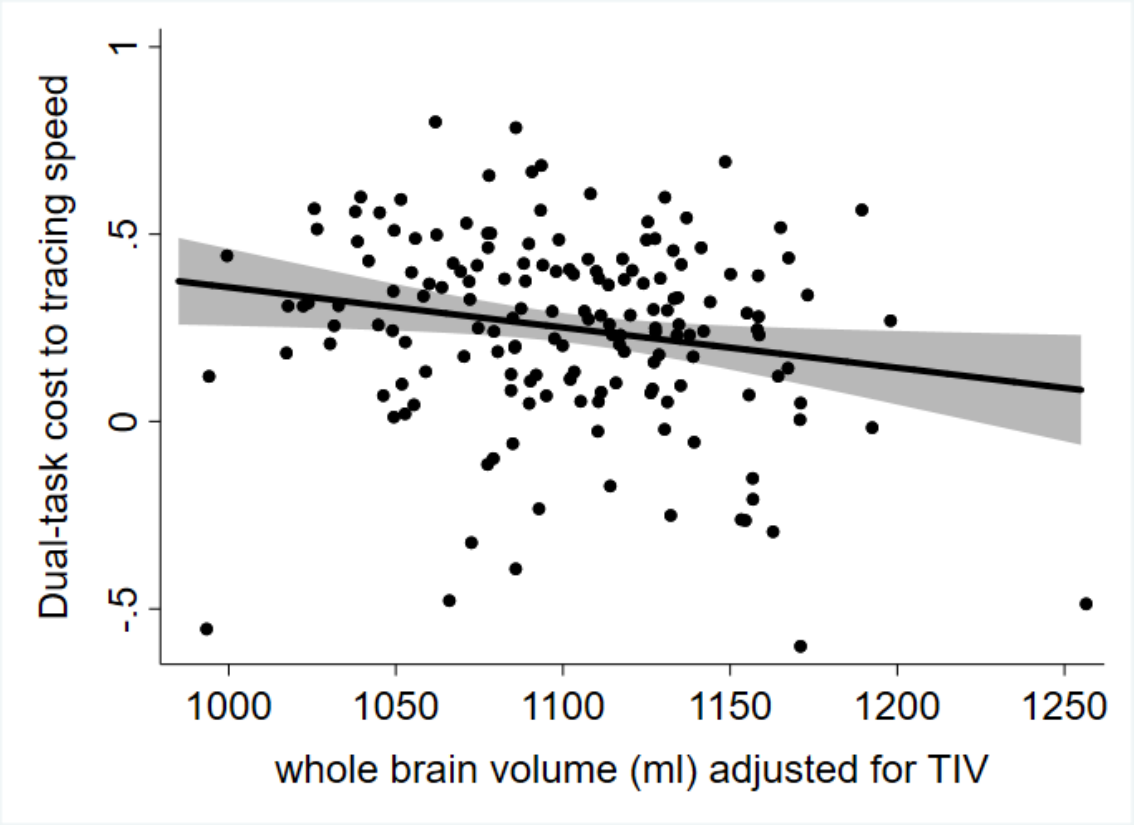


Figure 8-19. Association between whole brain volume and dual-task cost to circle-tracing speed

Solid line represents prediction from the multivariate regression model, holding all other predictors at average values, and shaded area represents 95% confidence intervals. The model included adjustment for age at assessment, sex, adult socioeconomic position, education, handedness, amyloid status, white matter hyperintensity volume, APOE-ε4 genotype and total intracranial volume. Markers show each participant's mean dual-task cost against whole brain volume adjusted for total intracranial volume.

8.4.2.2. Serial subtraction

Results of the multivariable regression models for subtraction speed and accuracy are given in Table 8-9. Results for the demographic and life-course factors (sex, age at assessment, childhood cognitive ability, education and adult socioeconomic position) are not reported as they are essentially unchanged from the first analysis section (8.3.2.2).

Table 8-9. Associations between biomarkers and subtraction speed and accuracy in cognitively-normal participants (n = 392)

	Subtraction rate: coefficient (responses per second) and 95% CIs		**Odds ratio for making a subtraction error and 95% CIs**	
Predictor	**Main effect of predictor**	**Interaction between predictor and condition (direct vs. indirect)**	**Main effect of predictor**	**Interaction between predictor and condition (direct vs. indirect)**
β-amyloid status (negative as reference)	**-0.05*** (-0.09, -0.02)	0.01 (-0.00, 0.02)	1.23 (0.86, 1.77)	0.88 (0.60, 1.30)
WMHV (per 10 ml)	-0.02 (-0.04, 0.01)	0.00 (-0.01, 0.01)	1.24 (0.95, 1.64)	**0.73** (0.57, 0.93)
Whole brain volume (per 10 ml)	**0.004** (0.001, 0.007)	-0.000 (-0.001, 0.000)	0.99 (0.97, 1.02)	1.00 (0.99, 1.02)
***APOE*-ε4** (non-carriers as reference)	0.02 (-0.02, 0.05)	0.01 (-0.00, 0.02)	0.81 (0.60, 1.09)	0.83 (0.61, 1.14)

Multivariable regression models were used so each association is independent of all others. In addition to the predictors listed, models also included sex, age at assessment, childhood cognitive ability, adult socioeconomic position, handedness and total intracranial volume.

Coefficients and Odds Ratios in bold are significant at p < 0.05 and asterisks indicate significance at p < 0.01.

CI = confidence interval; WMHV = white matter hyperintensity volume

On average, Aβ+ participants subtracted more slowly than Aβ- participants (Table 8-9). The effect size is equivalent to 2.3 fewer responses in the 45 second period (predicted means: Aβ+ = 20.6 responses in 45 seconds, Aβ- = 22.9 responses in 45 seconds). There was no evidence that this effect differed according to whether the concurrent circle-tracing task was in the direct or indirect condition (Table 8-9).

When rerunning the models replacing dichotomous amyloid status with continuous SUVR, the results followed the same pattern: there was an association between higher SUVR and slower subtraction rate (regression coefficient = -0.23, 95% CIs -0.42 to -0.04, p = 0.018) and no statistically significant interaction between SUVR and circle-tracing condition (regression coefficient = 0.03, 95% CIs -0.03 to 0.09, p = 0.38). These results were unchanged in a sensitivity analysis excluding the individuals with imputed SUVR values (see section 3.2.2).

There was no evidence of a statistically significant difference between Aβ+ and Aβ- participants in subtraction error rate (unadjusted means: Aβ+ = 3.1%; Aβ- = 2.6%) (Table 8-9).

Larger whole brain volume was associated with slightly faster subtraction rate but was not associated with subtraction accuracy (Table 8-9).

WMHV was not associated with subtraction rate or accuracy overall (Table 8-9). However, for subtraction accuracy there was an interaction between WMHV and circle-tracing condition such that the tendency to make more subtraction errors in the indirect condition (compared to the direct condition) was reduced in participants with greater WMHV. As a post-hoc exploration of this interaction, the models were rerun in the direct and indirect conditions separately, which revealed that greater WMHV predicted a greater likelihood of making subtraction errors in the direct condition (*OR* = 1.45 per 10 ml, 95% CIs 1.07 to 1.96, p = 0.02), but WMHV was not associated with subtraction accuracy in the indirect condition (*OR* = 1.03 per 10 ml, 95% CIs 0.78 to 1.37, p = 0.81).

APOE-ε4 was not associated with any subtraction outcomes (adjusting for amyloid status and all other covariates) (Table 8-9).

8.4.2.3. Selection of summary outcome variables

I decided that the following two summary outcomes would be the most meaningful representation of the circle-tracing effects described above. They are based on the six dual-task trials completed by all participants.

Difference in error rate between indirect and direct conditions. This was defined by calculating each participant's mean error rate in the direct and indirect conditions separately (the mean number of errors per rotation across the three trials within each condition) and subtracting the former from the latter**.** This is an index of the extent to which participants' tracing accuracy was affected by the indirect visual feedback. As the circle-tracing task was designed to assess visuomotor integration ability, this outcome reflects the core demands of the task better than overall error rate, and captures the aspect of performance that showed a difference between Aβ+ and Aβ- participants. I considered whether to normalise this outcome to each participant's performance in the direct condition i.e. to calculate the relative difference in error rate rather than the absolute difference. However, this was problematic because it would require dividing by zero for some participants who made no errors in the direct condition. As the error rate variable is already expressed per rotation, large differences between participants have already been eliminated and its range is relatively small (-1.58 to 9.9 errors per rotation), therefore I decided that the absolute difference was appropriate.

Overall tracing speed. This was defined as the mean number of rotations across all six dual-task trials (combined across direct and indirect conditions). This is the same outcome that was used to examine speed-accuracy trade-offs (see sections 8.3.1.1.4 and 8.3.2.1.4). I chose this outcome, rather than the difference in speed between the direct and indirect conditions, as it captures a wide and meaningful variation between individuals (range 0.6 to 35.8 rotations in 45 seconds) and is the aspect of performance which was associated with whole brain volume.

For the serial subtraction task, the summary outcomes are obvious: **mean subtraction rate** (responses per second) and **overall error rate**, as defined earlier 8.3.1.2.2.

These summary outcomes will be used when comparing performance across different cognitive tests (see Chapter 9).

8.5. Discussion

8.5.1. Summary

This study reported the results of a circle-tracing and serial subtraction task which were administered concurrently. The serial subtraction task showed evidence of sex differences (males subtracting faster and more accurately) and relatively strong associations with childhood cognitive ability and educational attainment, whereas circle-tracing performance was generally less associated with these factors. As hypothesised,

cognitively-normal Aβ+ individuals showed evidence of subtle deficits in visuomotor integration, indicated by disproportionately poorer tracing accuracy when visual feedback was indirect. There was also evidence of an association between Aβ pathology and slower subtraction rate. Speed-accuracy trade-offs were an important feature of circle-tracing performance.

8.5.2. Patterns of performance

8.5.2.1. Patterns of performance on the circle-tracing task

Performance of Insight 46 participants on the circle-tracing task was consistent with previous studies in that tracing was slower and less accurate in the condition of indirect visual feedback compared to the direct condition. A trade-off between speed and accuracy was observed, with the participants who traced more quickly tending to make more errors per rotation. As discussed earlier in the context of the Choice Reaction Time task (see section 6.5.2), it is worth noting that a person may alter their speed-accuracy strategy during the task, which is not captured by a between-subjects analysis.

The speed-accuracy trade-off showed a surprising relationship with age at assessment whereby older participants traced faster and less accurately, despite the age range being so narrow (2.6 years – reflecting the time taken to collect the data as all participants were born in the same week). There is no obvious explanation for this finding. The nature of the trade-off means that performance was not better or worse with age – just different. Age effects on this circle-tracing task have been investigated in one study which compared older and younger adults and found that they did not differ in their tracing accuracy but the older adults traced more slowly (Vaportzis, Georgiou-Karistianis and Stout, 2014). In another study which recruited adults aged from 21 to 95, strong associations were observed between older age and decreasing accuracy on a different circle-tracing task which involved pursuing a moving target around a circle, but speed was not reported (Stirling *et al.*, 2013). As mentioned in previous chapters (4, 6 and 7) where “age effects” were observed, I considered the possibility of a recruitment bias whereby participants seen towards the beginning of the data collection period may have differed in some ways to those seen towards to the end. This is discussed in greater detail in Chapter 10. One factor of potential relevance to differences in speed-accuracy trade-offs is risk sensitivity, as one study has reported that risk-seeking individuals are more likely to prioritise speed whereas risk-averse individuals are more likely to prioritise accuracy on a task of motor control (Nagengast, Braun and Wolpert, 2011).

In the sub-sample of participants who completed two additional single-task circle-tracing trials (without the concurrent subtraction task), the majority traced more quickly in the single task compared to the dual. According to the influential multiple-component working memory system proposed by Baddeley (Baddeley and Hitch, 1974; Baddeley, 2000), healthy individuals show little interference when performing dual tasks that rely on different components of working memory, such as processing verbal and visuospatial information (e.g. (MacPherson *et al.*, 2012)), but these results suggest that the circle-tracing and serial subtraction tasks were competing for the same attentional and working memory resources. Anecdotally, I often observed participants slowing down or even pausing their tracing on the dual task when they were struggling to work out the next subtraction response. However, tracing accuracy was poorer in the single task, which is consistent with a previous study using the same experiment (Vaportzis, Georgiou-Karistianis and Stout, 2014) and can be explained by speed-accuracy trade-offs. It appears that the influence of the subtraction task in slowing down participants' tracing prevented them from adopting a fast and inaccurate tracing style, whereas when free to focus entirely on the circle-tracing task they could prioritise speed at the expense of accuracy. Limitations of this experiment for measuring dual-task abilities are discussed in section 8.5.5.

8.5.2.2. Patterns of performance on the serial subtraction task

There was wide variability between participants in terms of their subtraction rates, with the fastest participant subtracting 10 times faster than the slowest, while error rates were generally low. There was a correlation, rather than a trade-off, between speed and accuracy, with the highest error rates recorded by participants with the slowest subtraction rates. A possible explanation for this is that arithmetic is a task where speed has to be adjusted to maintain accuracy, because each response is either correct or incorrect, so – unless one is prepared to simply guess the answer – one has to take as much time as necessary to work out each answer. This explanation is consistent with the fact that subtraction rate – but not subtraction accuracy – was affected by the demands of the concurrent circle-tracing task, with slower subtraction rates while tracing with indirect visual feedback compared to direct visual feedback. Another piece of evidence in support of this idea that only subtraction rate (but not accuracy) is influenced by external demands is the observation that left-handed participants subtracted more slowly than right-handed participants in the indirect condition – where circle-tracing was disproportionately difficult for them – but did not perform worse than right-handed participants on subtraction accuracy.

The finding that subtraction rate was slower while tracing with indirect visual feedback is somewhat analogous to the concept of “dual-task cost” discussed earlier, as it could be explained by the fact that the indirect condition placed a greater demand on cognitive resources, leaving less resources available to allocate to the subtraction task.

8.5.3. Demographic and life-course predictors

8.5.3.1. Predictors of performance on the circle-tracing task

Participants with neurological or psychiatric conditions traced less accurately than cognitively-normal participants, but the fact that this effect was observed with both direct and indirect visual feedback suggests that motor problems or more distributed cognitive deficits may have been the main contributing factors rather than a difficulty with integrating non-standard visual feedback.

On the whole, the circle-tracing task was relatively free from associations with childhood cognitive ability, educational attainment and adult socioeconomic position, with the exception that higher educational attainment predicted more accurate tracing. This is perhaps surprising given the importance of these predictors for most cognitive tests including tests with a speed component such as Digit-Symbol Substitution (see Chapter 4). One speculative explanation for this could be that the circle-tracing task relies on fine motor control, which is more affected by physical health factors at this age, such as tremors and arthritis. Although childhood cognitive ability was not associated with tracing speed or accuracy, analysis of error times revealed a specific effect whereby participants with higher childhood cognitive ability were able to detect and correct their errors more quickly in the condition of indirect visual feedback, with a particular advantage for error detection. To my knowledge, this is the first study to report associations between childhood cognitive ability, educational attainment and performance on a visuomotor integration task in older age, and this evidence provides important context for interpreting any decline in visuomotor integration that may be associated with accumulating brain pathologies.

The finding that left-handed participants traced disproportionately less accurately in the indirect condition is likely to reflect a difficulty with drawing left-handed under the constraints of the equipment that was used to cover the tablet screen, rather than that left-handed participants were disproportionately disadvantaged by the indirect visual feedback. The fact that left-handed participants also subtracted disproportionately slowly in the indirect condition supports this view. The box used to cover the tablet screen was

rather small, so I believe that left-handed participants were more restricted in how they could angle their arm to draw clockwise.

8.5.3.2. Predictors of performance on the serial subtraction task

Sex differences of a moderate magnitude were observed on the subtraction task, with males subtracting more quickly and more accurately. A male advantage in mathematics has been widely reported and its basis is the subject of some controversy. While some argue that it may be underpinned by biological differences in visuospatial abilities, others have pointed to environmental and cultural factors (Halpern *et al.*, 2007). For example, an analysis of quarter of a million students from 40 countries found that sex differences in mathematical test performance were eliminated in countries with high levels of gender equality (Guiso *et al.*, 2008). Other predictors of faster and more accurate subtraction were higher childhood cognitive ability, higher educational attainment and absence of a neurological or psychiatric condition. As the subtraction task was not administered alone (without concurrent circle-tracing), it is not possible to determine whether these predictors apply to subtraction in general, or were specific to this dual-task format. This issue is discussed in greater detail in section 8.5.5.

8.5.4. Associations with biomarkers and APOE-ε4

8.5.4.1. Circle-tracing

To my knowledge, this is the first study to evaluate the effects of β-amyloid pathology on visuomotor integration in cognitively-normal older adults. These results support the hypothesis that β-amyloid deposition is associated with subtle deficits in visuomotor integration, as Aβ+ participants had disproportionately poorer circle-tracing accuracy with indirect visual feedback. This mirrors the result reported in presymptomatic Huntington's mutation carriers (Say *et al.*, 2011) and builds on my previous analysis of the same circle-tracing task in presymptomatic FAD mutation carriers, who were an average of 7 years before estimated age of symptom onset (Macpherson *et al.*, 2017). The mutation carriers (n = 19) traced less accurately than controls (n = 12) across the experiment as a whole and there was a suggestion that this effect was exaggerated in the indirect condition but the interaction was not statistically significant. The results are also consistent with previous studies that have described visuomotor integration deficits

in patients with AD dementia (Tippett and Sergio, 2006; Tippett, Krajewski and Sergio, 2007; Tippett, Sergio and Black, 2012; Verheij *et al.*, 2012; de Boer *et al.*, 2016).

While loss of motor control is a major symptom of Huntington's disease, it is not a prominent feature of AD, particularly in the early stages where memory deficits tend to dominate. This result is of considerable interest in understanding more about the subtle cognitive decline that may accompany the accumulation of Alzheimer's pathology during the preclinical stage of the disease. However, one must of course remain mindful of the fact that Aβ positivity does not necessarily predict progression to AD on an individual basis (Brookmeyer and Abdalla, 2018).

It is interesting that the outcome that was sensitive to Aβ positivity (disproportionate decrease in accuracy in the indirect condition) also showed evidence of being sensitive to white matter disease burden. This is consistent with the sensitivity of this outcome to presymptomatic Huntington's disease, as white matter damage has been documented as an early feature of Huntington's disease (McColgan *et al.*, 2017). This outcome was not associated with any of the demographic and life-course predictors (sex, age at assessment, childhood cognitive ability, education, adult socioeconomic position – with the exception of left-handedness which I believe was an artefact of the experimental design (see section 8.5.2.1)), suggesting that it may be specifically sensitive to brain pathology rather than general cognitive ability.

This study did not find any evidence of an association between β-amyloid deposition and dual-task cost. However the measurement of dual-task cost was subject to limitations in the experimental design, discussed in section 8.5.5.

The association between greater whole brain volume and faster tracing speed is consistent with my earlier finding of a similar association on the Digit-Symbol processing speed test (see Chapter 4) and the association with subtraction speed, discussed below.

8.5.4.2. Serial subtraction

I did not have prior hypotheses about associations between amyloid deposition and performance on the serial subtraction task, as the subtraction task was initially chosen as a way to examine the impact of a dual task on circle-tracing, rather than as a primary outcome of interest in its own right. The finding that Aβ+ participants subtracted more slowly is interesting but not easy to interpret because the experiment did not contain a measure of "pure" subtraction ability (i.e. subtraction as a single task without the concurrent circle-tracing task). Impaired calculation ability, or dyscalculia, is known to occur in AD and to have significant functional consequences (Girelli and Delazer, 2001).

Like visuomotor integration, it is understood to be associated with parietal damage, specifically the left inferior parietal lobule (Hirono *et al.*, 1998). Research studies that assess calculation tend to focus on errors rather than speed (e.g. (Martin *et al.*, 2003), although assessments usually have time limits so an error would be recorded if a response is too slow (e.g. the arithmetic sub-test of the Wechsler Adult Intelligence Scale – Revised (Wechsler, 1981)).

As subtraction rate was dependent on the demands of the concurrent circle-tracing task (see 8.3.2.2.1 and 8.5.3.2), the slower subtraction rate of Aβ+ participants could potentially stem from greater interference from the concurrent circle-tracing task, rather than a difficulty with subtraction *per se*. Unfortunately, this theory cannot be tested with these data. Also, this subtraction task was not designed to probe differences in calculation ability, as subtracting in 3s is not challenging for the majority of people and there was a strong ceiling effect, so a more difficult test would be required to properly probe differences in calculation ability between Aβ+ and Aβ- individuals.

The finding of an association between whole brain volume and subtraction rate mirrors its association with circle-tracing speed and information processing speed as assessed by the Digit-Symbol test (see Chapter 4), indicating that the effect of brain volume on speed applies across a range of cognitive domains and task demands – this is discussed further in Chapter 10.

8.5.5. Limitations

There were a number of limitations to the experimental design. The most significant was regarding its suitability for assessing "dual-task cost" – the extent to which performance on a task is compromised by having to perform a second task concurrently. There was no measure of subtraction ability as a single task, which made it difficult to interpret the effects of predictors that were associated with subtraction rate and accuracy, since it was impossible to know the extent to which the dual-task aspect was driving performance, rather than the core demands of the subtraction task. Less than half of participants had a measure of circle-tracing ability as a single task, as this part of the protocol was added later. However, the size of this sub-sample (n=208) was still large in comparison to most neuropsychological studies. The more potentially limiting issue was the fixed order and unequal numbers of the trials (6 dual-task trials, followed by 2 single-task trials), coupled with the complex practice effects, which made it difficult to decide how best to compare the dual and single tasks. Ideally there would be five conditions with equal numbers of trials administered in a randomised order: circle-tracing with direct visual feedback;

circle-tracing with indirect visual feedback; serial subtraction; circle-tracing with direct visual feedback plus serial subtraction; circle-tracing with indirect visual feedback plus serial subtraction.

Although the circle-tracing task has been administered in a dual-task format before (Vaportzis, Georgiou-Karistianis and Stout, 2014; Vaportzis *et al.*, 2015b), it is not a straight-forward choice for measuring dual-tasking ability as it is hard to define optimum performance on the task, since speed and accuracy are in competition with each other. If participants adopt a different balance of speed and accuracy on the dual and single tasks, this can result in a negative dual-task cost on either outcome alone, as observed in this study. To probe dual-tasking ability more easily, it may be better to use a task less susceptible to speed-accuracy trade-offs, such as a tracking task (Della Sala *et al.*, 2010; MacPherson *et al.*, 2012) or a simple reaction time task (Logie *et al.*, 2007).

Another limitation of the analyses is that mild motor symptoms such as action tremor were not controlled for. While I do not think this was a major issue in this study, this potential confounder could be accounted for using scores from the Unified Parkinson's Disease Rating Scale (UPDRS), which was administered to all Insight 46 participants.

There were a number of other limitations to the experimental design. Below is a list of recommendations (in no particular order) for how I think the task should ideally be altered to address these limitations.

i) If a participant traces quickly and completes many rotations, the blue line showing their tracing path can begin to obscure the annulus. I suspect this makes further errors more likely, so ideally the tracing path should fade during each rotation so that by the time the participant reaches the top of the circle, they have a clean circle for the next rotation.

ii) As the speed-accuracy trade-off is an important feature of this task, it would be useful to check its influence more explicitly. This could be done by adding a condition where participants are asked to prioritise accuracy, tracing as slowly as is necessary to achieve their best accuracy, as was done in one previous study (Lemay *et al.*, 2005).

iii) The set-up for the indirect condition should be improved so that it does not disadvantage left-handed people. This could be done by using a bigger box to obscure the screen, or possibly allowing left-handed people to trace anticlockwise if the software was modified to allow for this.

iv) The software output for error time was sufficient for comparing error detection with error correction, but for a more accurate analysis of error time – for example, to investigate whether certain groups of were particularly slow to detect or correct errors – it would be better to have fine-grained data on the duration of each error. I suspect that this information is recorded in the raw data-files but do not know how to extract it.

v) In light of the interesting association between amyloid deposition and subtraction rate, it would be informative to compare this with a more challenging calculation task, as well as conducting the subtraction task on its own without circle-tracing. It is possible that the slower subtraction observed in this study could be indicative of general decline in calculation ability, which might be detectable as errors on a more challenging task.

Strengths and limitations that apply to all the analyses presented in this thesis, such as considerations relating to the generalisability of the sample, are discussed in Chapter 10.

8.5.6. Conclusions

In summary, this study has provided novel evidence that Aβ+ cognitively-normal older adults may have subtle deficits in visuomotor integration. The finding of an association between Aβ pathology and slower subtraction rate merits further investigation and should be explored in future studies that address some of the limitations of this experimental design.

9. INTEGRATION OF RESULTS ACROSS THE COGNITIVE BATTERY, AND POTENTIAL APPLICATIONS

So far in this thesis each cognitive test has been considered separately. In order to view the results as a whole and draw broader conclusions from them, I aimed to 1) summarise the results reported in previous chapters; 2) evaluate which combination of cognitive measures may be most sensitive to amyloid status; 3) reflect on potential applications of these results. The rationale, methods and results for each of these three aims are reported in the following three sub-sections.

9.1. Summary of results across cognitive tests

9.1.1. Rationale

In order to draw conclusions about patterns of performance across all the tests, I aimed to summarise and compare the effects of the various predictors (demographic characteristics, biomarker measures etc.) across the different cognitive measures and to examine correlations between the cognitive measures.

9.1.2. Methods

While the paper-and-pencil tests all had simple outcomes (total number of items correct), the computerised tests had a variety of outcome measures that could be analysed on a trial-by-trial basis or summarised across the various different conditions of the task. The selection of two or three key summary outcomes for each computerised test was explained in previous chapters (see section 3.5.2.1). For the purposes of comparison across tests, all outcomes were framed such that a higher score indicates better performance and converted into z-scores based on all participants. Transformations were applied to skewed outcomes where helpful to enable the data to more closely approximate the normal distribution. Table 9-1 summarizes the 19 cognitive outcomes that are used in this chapter and how they were derived.

Table 9-1. List of cognitive outcome measures from the Insight 46 battery

Name of cognitive measure	Brief descriptor	How score was derived
MMSE (see Chapter 4)	Screening test for cognitive impairment, covering multiple cognitive domains	Total score (max. 30) → standardised into z-score.
Matrix Reasoning (see Chapter 4)	Non-verbal reasoning	Total score (max. 32) → standardised into z-score.
FNAME-total (see Chapter 4)	Associative memory for face-name and face-occupation pairs	Total score of names and occupations (max. 96) → standardised into z-score.
Logical Memory Immediate (see Chapter 4)	Immediate recall for a short story	Total score (max. 25) → standardised into z-score.
Logical Memory Delayed (see Chapter 4)	Recall for a short story after approximately 20 minutes	Total score (max. 25) → standardised into z-score.
Digit-Symbol (see Chapter 4)	Processing speed	Total score (max. 93) → standardised into z-score.
PACC (see Chapter 4)	Composite measure of MMSE, FNAME-total, Logical Memory Delayed and Digit Symbol	Mean of the following four z-scores: MMSE, Logical Memory Delayed Recall, Digit-Symbol, FNAME-total
"What was where?" Identification (see Chapter 5)	Memory for the identity of objects in the "What was where?" task	Proportion of correct identifications → standardised into z-score
"What was where?" localisation (see Chapter 5)	Memory for the location of objects in the "What was where?" task	Mean localisation error for correctly identified items → log-transformed → multiplied by -1 → standardised into z-score
"What was where?" binding (see Chapter 5)	Measure of "swap errors" where objects and locations are "mis-bound" in the "What was where?" task	Proportion of swap errors within correctly identified items → multiplied by -1 → standardised into z-score
Choice RT (see Chapter 6)	Mean response time in the 2-choice RT task	Mean RT (ms) for correct responses → multiplied by -1 → standardised into z-score
Choice RT accuracy (see Chapter 6)	Accuracy in the 2-choice RT task	Percentage of correct responses → standardised into z-score
Choice RT consistency (see Chapter 6)	Consistency of response times in the 2-choice RT task	Intra-individual variability in RT for correct responses (*SD* / mean) → multiplied by -1 → standardised into z-score

Name of cognitive measure	Brief descriptor	How score was derived
Response Inhibition "incongruent cost" to RT (see Chapter 7)	Proportional difference in response time between conditions of congruent and incongruent stimuli	(Mean RT for incongruent trials) minus (mean RT for congruent trials) → divided by mean RT for congruent trials → multiplied by -1 → standardised into z-score
Response Inhibition "incongruent cost" to accuracy (see Chapter 7)	Difference in accuracy between conditions of congruent and incongruent stimuli	(Error rate for incongruent trials) minus (error rate for congruent trials) → multiplied by -1 → standardised into z-score.
Circle-tracing speed (see Chapter 8)	Speed of circle-tracing while concurrently completing serial subtraction	Mean number of rotations → log-transformed → standardised into z-score
Circle-tracing "indirect cost" to accuracy (see Chapter 8)	Difference in circle-tracing accuracy between conditions of indirect and direct visual feedback, while concurrently completing serial subtraction	(Mean number of errors per rotation in indirect trials) minus (mean number of errors per rotation in direct trials) → standardised into z-score
Subtraction speed (see Chapter 8)	Speed of subtracting in 3s while concurrently complete circle-tracing	Number of responses per second → standardised into z-score
Subtraction accuracy (see Chapter 8)	Accuracy of subtracting in 3s while concurrently complete circle-tracing	Percentage of incorrect responses → square-root transformed → multiplied by -1 → standardised into z-score

FNAME = Face-name associative memory exam; MMSE = Mini Mental State Examination; PACC = Preclinical Alzheimer Cognitive Composite; RT = reaction time

Multivariable regression models were fitted for each cognitive outcome, with sex, age at assessment, education, childhood cognitive ability, adult socioeconomic position, amyloid status, WMHV, whole brain volume and *APOE* genotype (ε4-carrier or non-carrier) included as predictors. As in previous analyses, bootstrapping was used to produce bias-corrected and accelerated 95% CIs from 2000 replications for outcomes with skewed distributions (MMSE, Matrix Reasoning, "What was where?" Identification, Choice RT, Choice RT accuracy, Serial subtraction accuracy, Response Inhibition "incongruent cost" to RT, and Response Inhibition "incongruent cost" to accuracy).

In chapters 4 to 8, associations between cognitive outcomes and the demographic and life-course predictors (sex, age at assessment, education, childhood cognitive ability,

adult socioeconomic position, presence of a neurological or psychiatric condition) were reported based on all 502 participants, whereas associations between cognitive outcomes and the biomarker predictors (β-amyloid, WMHV, whole brain volume, *APOE*) were reported based on the sub-sample of cognitively-normal participants with complete biomarker data (n = 406). For the purpose of summarising all the predictors together in the current chapter, only the sub-sample of cognitively-normal participants with complete biomarker data were included. This explains why the regression coefficients for the demographic and life-course predictors (and the R-squared values for the models) are not identical to those reported in previous chapters.

Correlations between cognitive outcomes were measured using Pearson's correlation coefficient, or Spearman's correlation coefficient if either or both of the outcomes had a skewed distribution (as per the list above). The PACC was not included in these correlation analyses because its sub-tests were already included as outcomes in their own right, so correlations between the PACC and its own sub-tests are present by definition and are not informative.

9.1.3. Results and commentary

Results of the multivariable regressions for each cognitive outcome are shown in Table 9-2. Some observations on the patterns of results are presented below, but a fuller discussion is provided in Chapter 10 which summarizes the findings of this thesis and interprets them in the context of the literature.

Overall it appears that sex differences were more evident on the standard paper-and-pencil tests and the serial subtraction task (with a consistent advantage for females, especially on memory-based tests, and a substantial advantage for males on subtraction speed), whereas sex differences on the computerised tests were much smaller in magnitude and not statistically significant (Table 9-2). (In Chapter 5 I reported sex differences on the "What was where?" computerised test when the full sample of 502 participants was included, which are now in the same direction but not statistically significant).

Childhood cognitive ability influenced performance on at least one outcome of every cognitive test (Table 9-2). (For the circle-tracing task, although it did not have a significant effect on the summary outcomes reported in this chapter, it was associated with differences in the time taken to detect and correct errors – see section 8.3.2.1.3.)

Education also had a consistent effect on many cognitive outcomes, whereas socioeconomic position was a much less notable predictor (Table 9-2).

Most cognitive outcomes did not have any association with age at assessment, which was as expected given the extremely narrow age range of Insight 46 participants (2.6 years – see section 3.6). However unexpected age effects were observed on the Matrix Reasoning, Choice RT and Circle-tracing tasks, as detailed in the relevant chapters (4, 6, 8). Possible reasons for these findings are discussed in Chapter 10.

Aβ+ participants consistently performed worse than Aβ- participants across a range of cognitive domains, with the greatest differences in Matrix Reasoning and Choice RT consistency (Table 9-2, Figure 9-1). On the eight measures where the difference was statistically significant, the effect size was reasonably consistent at around -0.3 *SD*.

Larger whole brain volume (adjusted for total intracranial volume) was associated with better performance on three cognitive measures which all have a speed component – Digit-Symbol, Circle-tracing speed, and Subtraction speed – with the largest effect on Digit-Symbol (Table 9-2, Figure 9-2). It is interesting that this relationship between whole brain volume and speed was shown on three tasks that all have very different demands in terms of what the participant is asked to do. WMHV was only associated with Digit-Symbol processing speed, and *APOE*-ε4 was associated with superior short-term memory on the Logical Memory and "What was where?" tasks. These findings are discussed in greater detail in Chapter 10.

Table 9-2. Predictors of performance for cognitively-normal participants on each cognitive outcome measure in Insight 46

Cognitive Measure	Coefficients (95% confidence intervals)									R^2
	Sex (female as reference)	**Age at assessment** (per year)	**Childhood cognitive ability** (per z-score)	**Education** (per category) [a]	**Adult SEP** (per category) [a]	**Aβ positivity** (Aβ negative as reference)	***APOE*-ε4** (non-carrier as reference)	**Whole brain volume** (per 10ml)	**WMHV** (per 10ml)	
MMSE	**-0.43* (-0.63, -0.18)**	-0.09 (-0.23, 0.03)	**0.17* (0.04, 0.31)**	**0.13* (0.05, 0.21)**	0.03 (-0.06, 0.12)	**-0.24 (-0.46, -0.05)**	0.10 (-0.06, 0.29)	-0.01 (-0.03, 0.01)	0.00 (-0.21, 0.14)	0.14
Matrix Reasoning	**-0.33 (-0.56, -0.09)**	**-0.17* (-0.30, -0.04)**	**0.13 (0.00, 0.27)**	**0.23* (0.14, 0.32)**	0.11 (-0.00, 0.23)	**-0.40* (-0.69, -0.15)**	0.13 (-0.09, 0.35)	0.01 (-0.01, 0.03)	-0.02 (-0.19, 0.12)	0.23
FNAME-total	**-0.54* (-0.77, -0.32)**	-0.05 (-0.18, 0.07)	**0.37* (0.24, 0.50)**	0.07 (-0.01, 0.15)	**0.13* (-0.04, 0.23)**	-0.10 (-0.33, 0.13)	0.02 (-0.17, 0.21)	0.01 (-0.01, 0.02)	-0.09 (-0.24, 0.07)	0.23
LM Immediate	**-0.58* (-0.83, -0.34)**	-0.03 (-0.16, 0.11)	**0.25* (0.11, 0.39)**	**0.11 (0.02, 0.19)**	0.02 (-0.08, 0.12)	**-0.30 (-0.54, -0.05)**	**0.22 (0.02, 0.43)**	-0.01 (-0.03, 0.01)	-0.10 (-0.26, 0.06)	0.16
LM Delayed	**-0.62* (-0.86, -0.38)**	-0.00 (-0.13, 0.13)	**0.29* (0.15, 0.42)**	0.06 (-0.03, 0.14)	0.08 (-0.02, 0.18)	-0.19 (-0.43, 0.05)	0.19 (-0.01, 0.40)	-0.01 (-0.03, 0.01)	-0.12 (-0.28, 0.04)	0.17
Digit-Symbol	**-0.33* (-0.56, -0.10)**	0.01 (-0.11, 0.14)	**0.23* (0.09, 0.36)**	**0.18* (0.10, 0.26)**	0.03 (-0.07, 0.12)	-0.17 (-0.41, 0.06)	-0.01 (-0.21, 0.19)	**0.05* (0.03, 0.07)**	**-0.21* (-0.37, -0.06)**	0.21
PACC	**-0.48* (-0.63, -0.33)**	-0.03 (-0.11, 0.04)	**0.26* (0.18, 0.35)**	**0.11* (0.05, 0.16)**	**0.07 (0.01, 0.13)**	**-0.17 (-0.32, -0.03)**	0.08 (-0.05, 0.20)	0.01 (-0.01, 0.02)	**-0.10 (-0.20, -0.01)**	0.34
WWW identification	-0.14 (-0.40, 0.12)	-0.03 (-0.17, 0.13)	**0.39* (0.25, 0.53)**	-0.04 (-0.14, 0.04)	0.09 (-0.02, 0.21)	**-0.30 (-0.55, -0.05)**	**0.24 (0.03, 0.43)**	0.01 (-0.01, 0.03)	0.02 (-0.13, 0.18)	0.11
WWW localisation	0.19 (-0.07, 0.45)	0.05 (-0.09, 0.20)	-0.01 (-0.16, 0.14)	**0.11 (0.02, 0.21)**	0.01 (-0.10, 0.12)	-0.11 (-0.37, 0.16)	**0.29 (0.07, 0.51)**	0.02 (-0.01, 0.04)	-0.01 (-0.17, 0.18)	0.09
WWW binding	-0.18 (-0.45, 0.09)	-0.00 (-0.10, 0.09)	**0.16 (0.00, 0.32)**	-0.00 (-0.10, 0.09)	-0.05 (-0.16, 0.07)	0.09 (-0.18, 0.37)	0.00 (-0.23, 0.23)	0.00 (-0.02, 0.02)	0.05 (-0.13, 0.23)	0.03

Cognitive Measure	Coefficients (95% confidence intervals)									R^2
	Sex (female as reference)	**Age at assessment** (per year)	**Childhood cognitive ability** (per z-score)	**Education** (per category) [a]	**Adult SEP** (per category) [a]	**Aβ positivity** (Aβ negative as reference)	***APOE*-ε4** (non-carrier as reference)	**Whole brain volume** (per 10ml)	**WMHV** (per 10ml)	
Choice RT	0.22 (-0.06, 0.46)	**-0.25* (-0.38, -0.12)**	0.15 (-0.01, 0.31)	0.07 (-0.02, 0.16)	0.07 (-0.05, 0.24)	-0.15 (-0.38, 0.10)	0.03 (-0.18, 0.22)	0.02 (-0.00, 0.04)	0.05 (-0.09, 0.19)	0.10
Choice RT accuracy	-0.23 (-0.49, 0.02)	**0.15 (0.02, 0.29)**	**0.25* (0.10, 0.39)**	-0.05 (-0.16, 0.04)	-0.05 (-0.14, 0.06)	-0.20 (-0.52, 0.04)	0.04 (-0.15, 0.24)	-0.00 (-0.02, 0.02)	0.06 (-0.09, 0.20)	0.06
Choice RT consistency	0.16 (-0.10, 0.41)	**-0.24* (-0.38, -0.10)**	0.14 (-0.00, 0.29)	0.09 (-0.01, 0.18)	-0.08 (-0.18, 0.03)	**-0.37* (-0.63, -0.11)**	0.12 (-0.10, 0.33)	0.01 (-0.01, 0.03)	0.08 (-0.09, 0.25)	0.08
RI "incongruent cost" to RT	0.18 (-0.07, 0.44)	0.09 (-0.03, 0.22)	0.13 (-0.01, 0.29)	0.05 (-0.05, 0.16)	0.08 (-0.02, 0.25)	0.06 (-0.20, 0.30)	-0.01 (-0.24, 0.19)	0.00 (-0.02, 0.03)	-0.05 (-0.20, 0.10)	0.05
RI "incongruent cost" to accuracy	-0.18 (-0.42, 0.04)	-0.03 (-0.16, 0.11)	**0.30* (0.16, 0.47)**	**0.13* (0.03, 0.23)**	**0.11 (0.00, 0.23)**	-0.07 (-0.35, 0.16)	0.09 (-0.12, 0.29)	-0.00 (-0.02, 0.02)	-0.08 (-0.21, 0.05)	0.17
CT speed	0.00 (-0.27, 0.27)	**0.23* (0.08, 0.37)**	-0.05 (-0.20, 0.10)	-0.06 (-0.15, 0.04)	0.04 (-0.08, 0.15)	0.07 (-0.20, 0.34)	0.01 (-0.22, 0.23)	**0.04* (0.01, 0.06)**	-0.16 (-0.34, 0.02)	0.06
CT "indirect cost" to accuracy	-0.09 (-0.33, 0.15)	0.02 (-0.10, 0.15)	0.03 (-0.10, 0.17)	-0.01 (-0.10, 0.08)	0.09 (-0.01, 0.19)	**-0.29 (-0.52, -0.05)**	0.03 (-0.17, 0.23)	-0.00 (-0.02, 0.02)	-0.14 (-0.30, 0.02)	0.04
Subtraction speed	**0.50* (0.24, 0.72)**	0.09 (-0.04, 0.22)	**0.24* (0.10, 0.37)**	**0.10 (0.01, 0.18)**	0.01 (-0.09, 0.11)	**-0.32* (-0.60, -0.08)**	0.09 (-0.11, 0.29)	**0.02 (0.00, 0.04)**	-0.12 (-0.28, 0.04)	0.21
Subtraction accuracy	-0.04 (-0.29, 0.17)	0.02 (-0.11, 0.16)	**0.23* (0.09, 0.38)**	**0.12* (0.03, 0.21)**	-0.04 (-0.15, 0.08)	-0.11 (-0.38, 0.13)	0.14 (-0.06, 0.33)	0.01 (-0.01, 0.02)	-0.11 (-0.31, 0.05)	0.11

Green text indicates positive association at p<0.05. ***Red text indicates negative association at p<0.05.*** *Asterisks indicate p<0.01. For each cognitive outcome, multivariable regression models were used so each association is independent of all others. R^2 is a measure of the proportion of variance explained by the model. The model for Logical Memory Delayed included an additional factor of delay duration (time elapsed between immediate and delayed recalls).* [a] *See section 3.2.4 for definition of categories.*

CT = circle-tracing; LM = Logical Memory; PACC = Preclinical Alzheimer Cognitive Composite; RI = Response Inhibition; RT = reaction time; SEP = socioeconomic position; WMHV = white matter hyperintensity volume; WWW = What was where?"

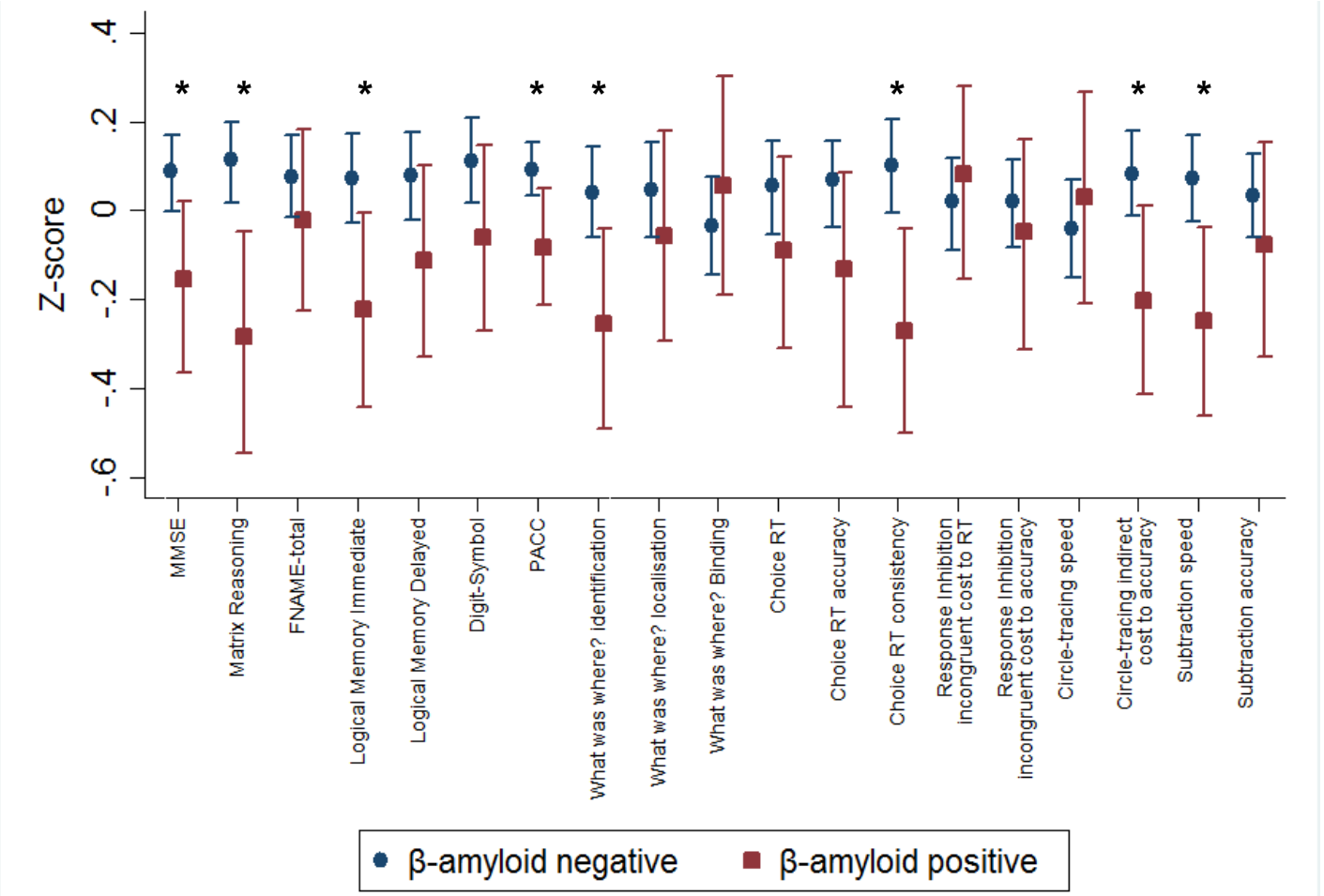


Figure 9-1. Performance of β-amyloid negative and β-amyloid positive participants on each cognitive test

Means and 95% confidence intervals are plotted from the multivariable regression models, adjusted for sex, age at assessment, childhood cognitive ability, education, adult socioeconomic position, APOE-ε4, whole brain volume, total intracranial volume, and white matter hyperintensity volume (WMHV). Asterisks indicate statistically significant differences.

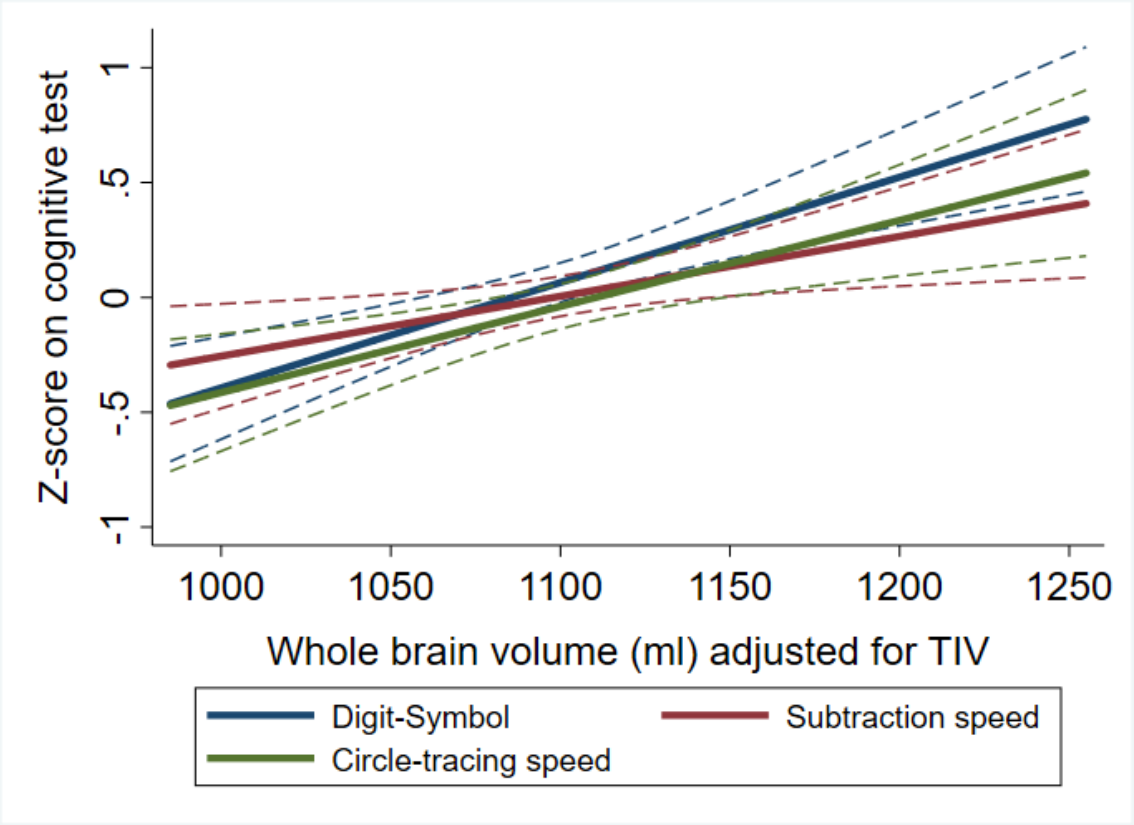


Figure 9-2. Associations between whole brain volume and cognition

Solid lines indicate predictions from the multivariate models adjusted for total intracranial volume, sex, age at assessment, childhood cognitive ability, education, adult socioeconomic position, APOE-ε4, amyloid status and white matter hyperintensity volume (WMHV). Dashed lines represent 95% confidence intervals. TIV = total intracranial volume.

Correlations between the cognitive outcomes are shown in Table 9-3. As a general rule, the standard paper-and-pencil tests were moderately correlated with each other, but the computerised tests were less so. This could be because, by having multiple metrics on the computerised tests, these metrics tend to hone in on more specific aspects of cognition, and hence reflect less of task-general demands. It could also be related to other properties of the tests, for example the standard tests all involved the participant responding verbally to the tester (with the exception of Digit-Symbol) and I would argue that the potential for adopting different approaches was relatively limited, whereas my impression from administering the computerised tests is that they offered more scope for adopting different strategies.

Within the six memory measures (FNAME-total, Logical Memory Immediate, Logical Memory Delayed, “What was where?” identification, “What was where?” localisation, and “What was where?” binding), correlations were only moderate at best (apart from the expected strong correlation between Logical Memory Immediate and Delayed, which not

only assessed similar cognitive processes but were based on identical test stimuli). Clearly these various memory tests encompassed a range of demands (visual, verbal, differing delay times etc.), so the specific set of cognitive processes that underlie performance on each test are different, and all are relevant to different aspects of memory in daily living. It is also likely that the tests were affected to differing degrees by other factors such as attention and confidence; anecdotally I observed that some participants commented that they experienced mind-wandering during the “What was where?” task, and some commented that they expected to perform poorly on the FNAME task. This highlights the value of including multiple measures of memory in a cognitive battery and illustrates the need for cautious interpretation of results, for example when using limited memory measures to define “amnestic” cognitive impairment or when equating results from studies that have used different memory tests.

Table 9-3. Correlation matrix for the cognitive outcome measures

	MMSE	Matrix Reasoning	FNAME-total	Logical Memory Immediate	Logical Memory Delayed	Digit-Symbol	"What was where?" identification	"What was where?" localisation	"What was where?" binding	Choice RT	Choice RT accuracy	Choice RT consistency	Response Inhibition "incongruent cost" to RT	Response Inhibition "incongruent cost" to accuracy	Circle-tracing speed	Circle-tracing "indirect cost" to accuracy	Subtraction speed	Subtraction accuracy
MMSE		0.26	0.30	0.30	0.31	0.29	0.11	0.15	0.10	n.s.	n.s.	n.s.	n.s.	0.24	n.s.	n.s.	0.23	0.21
Matrix Reasoning			0.27	0.23	0.17	0.41	0.28	0.22	n.s.	0.21	n.s.	0.23	0.16	0.43	0.09	0.09	0.31	0.28
FNAME-total				0.46	0.52	0.37	0.24	n.s.	0.12	n.s.	0.12	n.s.	n.s.	0.30	n.s.	n.s.	0.16	0.15
Logical Memory Immediate					0.85	0.31	0.17	0.14	0.10	n.s.	n.s.	n.s.	n.s.	0.19	n.s.	n.s.	0.13	0.10
Logical Memory Delayed						0.28	0.15	0.12	n.s.	n.s.	n.s.	n.s.	n.s.	0.18	n.s.	n.s.	0.14	n.s.
Digit-Symbol							0.22	0.20	0.11	0.37	n.s.	0.18	0.09	0.26	0.13	0.15	0.43	0.19
WWW identification								0.16	0.18	0.13	n.s.	0.13	n.s.	0.19	n.s.	n.s.	0.12	0.21

WWW localisation									-0.22	0.18	n.s.	0.11	n.s.	0.21	n.s.	n.s.	0.15	n.s.
WWW binding										n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	0.10	0.11
Choice RT											-0.28	0.19	n.s.	n.s.	0.13	n.s.	0.22	n.s.
Choice RT accuracy												0.18	n.s.	0.13	n.s.	n.s.	0.11	0.15
Choice RT consistency													n.s.	0.11	n.s.	n.s.	0.21	0.14
RI “incongruent cost” to RT														0.13	n.s.	n.s.	n.s.	n.s.
RI “incongruent cost” to accuracy															n.s.	n.s.	0.21	0.26
Circle-tracing speed																n.s.	0.18	0.13
Circle-tracing “indirect cost” to accuracy																	0.18	0.13
Subtraction speed																		0.51
Subtraction accuracy																		

Green colours indicate positive correlations, with darker shades denoting stronger correlations. Orange colour indicates negative correlations. Correlations were tested using Spearman’s or Pearson’s coefficient (see Methods). n.s. = no statistically significant correlation; RI = response inhibition; RT = reaction time; WWW = “What was where?”

9.2. Which combination of tests is most sensitive to amyloid status?

9.2.1. Rationale

As discussed in section 2.5.1, cognitive composites are increasingly being used as outcome measures in research studies in preclinical AD, based on evidence that a combination of cognitive measures covering different cognitive domains may be more sensitive to subtle decline than any single measure. Given that I have identified group differences between Aβ+ and Aβ- participants on a range of different cognitive measures, this leads to the question: which combination of these cognitive measures is most sensitive to amyloid status? It is possible that there is an optimum combination of measures sufficient to detect differences in cognition between the groups, beyond which the addition of further measures is redundant. I aimed to investigate this by modelling amyloid status as a function of different combinations of cognitive measures from the Insight 46 battery, plus relevant demographic and genetic predictors.

As age and *APOE*-ε4 are by far the biggest risk factors for Aβ positivity (*ARUK Dementia Statistics Hub*; Prince *et al.*, 2014), the strongest predictor in a model of amyloid status among Insight 46 participants, all approximately the same age, will clearly be *APOE*-ε4. As I have reported that the differences in cognition between Aβ+ and Aβ- participants were very subtle with overlapping distributions, and by definition all participants included in these analyses were cognitively-normal, it is unrealistic to expect that the results of a cognitive assessment alone could be a good predictor of dichotomous amyloid status on an individual level. However, it is still reasonable to hypothesise that a model which includes sensitive cognitive measures could fit the distribution of amyloid status better than a model based purely on *APOE* genotype and demographic factors, and to use this approach to evaluate which combination of cognitive measures adds the most value. One motivation for developing this approach as part of my PhD work is that once the Insight 46 follow-up assessments are completed and the cognitive measures can be expressed in terms of change over a ~2-year interval, this type of model may be useful for determining the combination of cognitive tests that is most sensitive to Aβ-related cognitive decline. There is consistent evidence that accelerated decline in cognition can be detected in Aβ+ individuals over time, on the PACC and other measures, (see sections 2.3 and 4.1) and the question of which combination of cognitive tests is most sensitive to such decline is a pertinent one for clinical trials.

As discussed in section 2.5.1, cognitive composites are generally constructed using one of two approaches: measures can either be selected based on their face validity and prior evidence for their sensitivity – as in the case of the PACC (Donohue *et al.*, 2014), or they can be selected using a data-driven approach (e.g. Ayutyanont *et al.*, 2014; Langbaum *et al.*, 2014; Tariot *et al.*, 2018). Both approaches require informed decisions to be made about which cognitive measures should be considered in the first place, but the first approach allows a higher priority to be placed on selecting tests that cover a range of cognitive domains and on accounting for practical considerations such as the time, equipment and expertise required to administer the assessments, depending on the intended application. Given that I have already used my judgment to define the most relevant summary outcomes for each cognitive test in the Insight 46 battery, and I have already drawn conclusions about the sensitivity of each cognitive measure to amyloid status, I decided to use this information to propose several combinations of cognitive measures and to compare them against each other, rather than comparing every possible combination of the measures. (With 19 summary outcome measures, the number of possible combinations of any number of them is over 500,000!)

As the PACC has emerged as a front-running outcome measure in therapeutic trials in preclinical AD (Weintraub *et al.*, 2018), I decided to frame my analyses in terms of starting with the PACC and testing whether the step-wise addition of other promising cognitive measures into the model yields improvements. After evaluating this and selecting the best combination of cognitive measures, I aimed to quantify how much this combination of cognitive measures could improve the prediction of an individual's amyloid status compared to a prediction based simply on their *APOE* genotype and demographic characteristics.

9.2.2. Methods

The first step in choosing which measures to add, and in which order, was to exclude those tests which form a part of the PACC composite, since performance on those tests is already captured by inclusion of the PACC in the model. Note that although the PACC contains only the delayed recall score from the logical memory test (not the immediate recall score), I excluded the immediate recall score from consideration as well, as it overlaps very closely with the delayed in terms of what it measures (Abikoff *et al.*, 1987). From the remaining outcomes, I considered only those for which there was a statistically significant difference between Aβ+ and Aβ- participants and ordered these by the effect size of the association with amyloid status (see Table 9-2, Figure 9-1). The outcome with

the largest effect size would be added to the model first, followed by step-wise addition of the other outcomes in decreasing order of effect size. I decided to make one alteration to this order based on the pragmatic fact that the serial subtraction and circle-tracing tasks are administered concurrently, so once subtraction speed is added, the circle-tracing outcome could be added with no extra time or effort for the tester or participant. Therefore, I moved the circle-tracing outcome up one place in the list, which seemed justified since its effect size was virtually identical to the next largest effect size ("What was where?" identification).

This yielded the following list, (see Table 9-1 for definitions of each outcome):

1. Matrix Reasoning (*d* = -0.40)
2. Choice RT consistency (*d* = -0.37)
3. Subtraction speed (*d* = -0.32)
4. Circle-tracing "indirect cost" to accuracy (*d* = -0.29)
5. "What was where?" identification (*d* = -0.30)

I also considered the cognitive domains that are covered by these measures and concluded that there is no reason to assume that any measures would be redundant since they all tap into different abilities and processes.

This gave me 7 models to test, where the outcome is amyloid status and the predictors are as follows:

Model A: sex, age at assessment, childhood cognitive ability, education, adult socioeconomic position, *APOE*-ε4 (carrier or non-carrier)

Model B: model A + PACC

Model C: model B + Matrix Reasoning

Model D: model C + Choice RT consistency

Model E: model D + Subtraction speed

Model F: model E + Circle-tracing "indirect cost" to accuracy

Model G: model F + "What was where?" identification

To compare how well these models predict amyloid status, I adopted a cross-validation approach. Cross-validation is a technique often used in machine-learning to evaluate the quality of a model in a relatively unbiased way by splitting the dataset into groups, called 'folds', and using these to generate multiple estimates of the model's accuracy. This has

advantages over simply evaluating a model's goodness-of-fit across the whole dataset because such an approach is vulnerable to over-fitting, meaning that the model may not generalise well to other datasets and will be of limited usefulness for predicting new results. Cross-validation is less likely to over-estimate a model's accuracy and can give increased confidence in the generalisability of the results (Brownlee, 2018).

I conducted *k*-fold cross-validation, which involves splitting the dataset into *k* folds, where *k* is a number to be decided depending on the properties of the dataset (see below). The principle of the process is that one fold is left out and the model is fitted on the remaining *k*-1 folds and its accuracy evaluated, then this process is then repeated a total of *k* times, leaving out a different fold each time. Comparing the model's accuracy across each repetition allows relatively unbiased conclusions to be drawn about its accuracy.

When deciding how many folds to split the data into, the optimum number depends on the sample size and the distribution of the outcome to be predicted. A large number of folds is desirable for reducing bias, as the sub-sample used for each repetition (the *k*-1 folds) is increasingly likely to be representative of the whole sample. However, a larger number of folds incurs extra computational burden and means that the sub-samples used for each repetition increasingly overlap with each other and become basically equivalent to the full sample, so are similarly vulnerable to overfitting (Brownlee, 2016). I used 10 folds as this is widely accepted as the default recommended number (Brownlee, 2018).

In order to compare the 7 different models, I used an information criterion and a measure of accuracy, as explained below.

Information criteria measure how much information is lost by using the model to represent the "true" processes that generated the data and are commonly used for selecting a statistical model from several candidates. I used the Akaike information criterion (AIC), which balances goodness-of-fit against complexity to give an overall measure of the quality of a model. Balancing these two factors is desirable because goodness-of-fit nearly always improves with the addition of extra parameters but this extra complexity can lead to over-fitting – simpler models are generally preferred as they are more robust and generalisable (Bozdogan, 1987). AIC is defined as follows:

$$AIC = 2k - 2(\ln L) \qquad \textit{(Equation 1)}$$

where k is the number of parameters and L is the likelihood function (a measure of goodness-of-fit)

Thus, AIC penalises complexity by including a term that increases with the number of parameters. A lower value of AIC indicates a better model, one that minimises the information loss. The AIC values of two models can be compared to calculate the

probability that one minimises the information loss better than the other. For example, if model X has a lower AIC value than model Y, the probability that model Y minimises the information loss better than model X is:

$$e^{((AIC_X - AIC_Y)/2)} \qquad \textit{(Equation 2)}$$

While AIC allows models to be compared against each other, it does not provide an absolute measure of a model's quality. In order to quantify how accurately each model can predict an individual's dichotomous amyloid status, I used Receiver Operating Characteristic (ROC) curves, a common technique for assessing the accuracy of diagnostic tests where the classification is either positive or negative. A ROC curve plots the detection rate (rate of 'true positives') against "Type 1 errors" (rate of 'false positives') for every possible cut-point. In this case the possible cut-points are predicted probabilities of being Aβ+. The area under the curve (AUC) is a measure of the accuracy of the model, with an area of 0.5 representing a useless model (one that classifies individuals no better than chance) and an area of 1 representing a perfect model (one that correctly classifies all individuals).

Therefore, my strategy for selecting the best model was to minimise AIC and maximise AUC.

As explained earlier, after identifying the best model, I aimed to test how well this model can predict an individual's amyloid status. To avoid circularity, I reserved a proportion of the sample – hereafter referred to as the "validation set" – before conducting the cross-validation so that I could carry out this predictive testing using data that had played no part in the model selection process.

A summary of the methods is presented in Figure 9-3, with explanatory notes in the text below. Out of the 406 cognitively-normal participants with complete biomarker data, 21 were missing data for one or more of these cognitive measures so were excluded from this section of the analyses, leaving a sample of 385. As the missing data were mostly due to random technical problems (see section 3.3), excluding these individuals is unlikely to bias the sample, whereas including them would mean that models A to G were being tested on variable sample sizes.

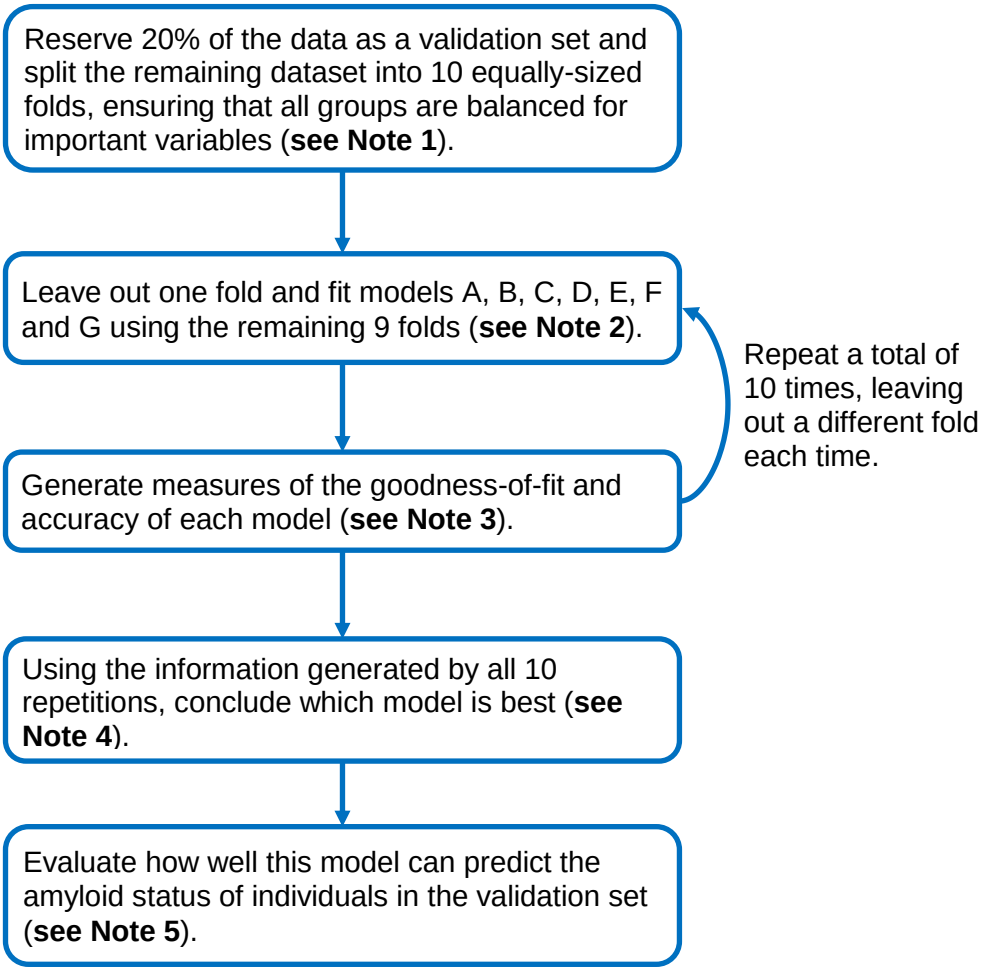


Figure 9-3. Flow-chart illustrating methods for cross-validation

Note 1. I chose to assign 20% of the sample to the validation set as this is a sufficient proportion to be representative of the dataset but allows the majority of the sample to be used in selecting the best model. It was important to ensure that each group contained a representative proportion of Aβ+ participants, since this is the outcome being modelled. I also balanced the groups for sex and childhood cognitive ability since these were the strongest and most consistent predictors of individual differences in cognition (see section 9.1). To achieve this, I ordered the dataset by amyloid status then sex then childhood cognitive ability, and applied the following repeating sequence to the dataset from top to bottom to assign each participant to a group: V, F1, F2, F3, F4, V, F5, F6, F7, F8, V, F9, F10, F1, F2, V, F3, F4, F5, F6, V, F7, F8, F9, F10. (V = validation set; F = fold).

Note 2. In this context, “fit the models” means run a logistic regression model with an outcome of amyloid status and predictors as specified earlier for models A to G. All

cognitive predictors were in the form of z-scores with a higher score indicating better performance (see Table 9-1).

Note 3. As explained earlier, the goodness-of-fit measure was Akaike's Information Criterion (AIC) and the accuracy measure was the area under the ROC curve (AUC).

Note 4. For each of models A to G, I calculated the mean AIC and mean AUC across the ten repetitions, and the standard errors of these means. I then chose the best model based on the aim to minimise AIC and maximise AUC (as explained above).

Note 5. The probability that an individual is amyloid positive can be predicted from the logistic regression model using an equation of the following form:

$$p(\mathrm{A\beta+}) = 1/(1 + e^{-(\mathrm{constant} + \mathrm{coef1.predictor1} + \mathrm{coef2.predictor2}\ldots)})$$

(Equation 3)

where p(Aβ+) is the probability of being amyloid positive, and 'coef' is the regression coefficient that corresponds to each predictor.

I fitted the chosen model to all ten folds to get the values for the constant and coefficients, then for each participant in the validation set I used the equation to predict their probability of being Aβ+. To assess the accuracy of these predictions, I compared them to the actual classifications of amyloid status derived from the PET scans. As the probability scores had a skewed distribution, I used a non-parametric test (Wilcoxon's rank sum) to determine whether the predicted probabilities were higher for the Aβ+ group than the Aβ- group.

9.2.3. Results

The characteristics of the groups are shown in Table 9-4.

Table 9-4. Characteristics of the validation set and the 10 folds

	n	% female	% Aβ+	Childhood cognitive ability (mean)	% *APOE*-ε4 carrier
Validation set	77	50.7	18.2	0.41	35.1
Fold 1	31	48.4	16.1	0.43	25.8
Fold 2	31	51.6	19.4	0.34	32.3
Fold 3	31	51.6	19.4	0.39	38.7
Fold 4	31	51.6	19.4	0.44	25.8
Fold 5	31	51.6	19.4	0.51	38.7
Fold 6	31	48.4	19.4	0.42	38.7
Fold 7	31	45.2	19.4	0.36	16.1
Fold 8	31	45.2	19.4	0.43	25.8
Fold 9	30	46.7	16.7	0.41	30.0
Fold 10	30	46.7	16.7	0.44	23.3

9.2.3.1. Selection of best model

Table 9-5 shows the results for each model in each of the ten repetitions. An example of the ROC curves for models A to G is provided in Figure 9-4, illustrating results from the first repetition. Figure 9-5 summarises the AUC and AIC statistics for each model averaged across the ten repetitions.

In each of the ten repetitions, the area under the curve (AUC) slightly increased from Models A to G (Table 9-5). This is consistent with the hypothesis that adding additional cognitive measures would improve the accuracy of the model. As illustrated in Figure 9-5, the rise in AUC value was greatest between models A and D, after which further gains appeared to be minimal. However, there was no evidence that these differences in accuracy were statistically significantly, as the confidence intervals of the AUCs of models A to G overlapped in each repetition.

Regarding Akaike's Information Criterion (AIC), each of the ten repetitions showed a consistent pattern whereby AIC increased slightly from model A to B, then decreased from model B to D (Table 9-5, Figure 9-5). Between models D to G the AIC values were more variable but remained broadly similar (Table 9-5, Figure 9-5). As a lower value of AIC is preferred, this suggests that the inclusion of the first three additional cognitive measures (PACC, Matrix Reasoning, Choice RT consistency) improved the quality of the

model (in spite of the fact that AIC is designed to penalize the number of parameters). The magnitude of these differences in AIC values is not negligible; for example, by comparing the mean AIC values for models A and D, using Equation 2 (page 262) above, we can conclude that the probability of model A being better than model D is only 5%. Based on these results, there was no evidence that the addition of further cognitive measures (Subtraction speed, Circle-tracing “indirect cost” to accuracy, and “What was where?” identification) improved the quality of the model.

In summary, the AUC and AIC statistics suggest that a model for amyloid status can be improved by the addition of cognitive measures (on top of demographic characteristics and *APOE*). These results tentatively suggest that model D appears the most promising, as adding additional cognitive measures beyond those included in model D may not yield improvements, and may not merit the extra time and effort involved in generating these measures (both for participants and researchers). However, it is important to note that these results are not conclusive and must be interpreted with caution given the lack of statistically significant differences in the AUCs.

Table 9-5. Results of the cross-validation process

Repetition	Model	AOC (95% CIs)	AIC
1	**A**	0.78 (0.71, 0.84)	242.6
	B	0.78 (0.72, 0.84)	243.6
	C	0.80 (0.47, 0.85)	240.9
	D	0.80 (0.75, 0.86)	239.9
	E	0.80 (0.75, 0.86)	241.9
	F	0.81 (0.76, 0.87)	240.3
	G	0.81 (0.76, 0.87)	242.3
2	**A**	0.77 (0.70, 0.84)	240.2
	B	0.79 (0.73, 0.85)	241.0
	C	0.80 (0.75, 0.86)	236.3
	D	0.82 (0.77, 0.88)	230.8
	E	0.82 (0.77, 0.88)	232.8
	F	0.83 (0.77, 0.88)	232.0
	G	0.83 (0.77, 0.88)	233.0
3	**A**	0.79 (0.72, 0.85)	237.3
	B	0.80 (0.74, 0.86)	237.3
	C	0.82 (0.76, 0.87)	232.5
	D	0.83 (0.77, 0.88)	229.4
	E	0.83 (0.77, 0.88)	231.1
	F	0.83 (0.78, 0.88)	231.2
	G	0.83 (0.78, 0.88)	233.1
4	**A**	0.77 (0.71, 0.84)	238.9
	B	0.79 (0.73, 0.85)	239.3
	C	0.81 (0.76, 0.87)	233.6
	D	0.82 (0.77, 0.87)	232.9
	E	0.82 (0.77, 0.88)	234.3
	F	0.83 (0.78, 0.88)	232.7
	G	0.83 (0.78, 0.88)	233.9
5	**A**	0.77 (0.71, 0.84)	240.4
	B	0.78 (0.72, 0.84)	241.7
	C	0.80 (0.74, 0.86)	237.1
	D	0.82 (0.77, 0.88)	232.9
	E	0.82 (0.76, 0.88)	234.8
	F	0.83 (0.77, 0.88)	233.6
	G	0.83 (0.78, 0.88)	235.0
6	**A**	0.78 (0.71, 0.84)	241.1
	B	0.79 (0.73, 0.85)	242.3
	C	0.80 (0.75, 0.86)	238.5
	D	0.82 (0.77, 0.88)	232.6
	E	0.82 (0.77, 0.88)	234.1
	F	0.83 (0.77, 0.88)	234.5
	G	0.83 (0.77, 0.88)	235.9

Repetition	Model	AOC (95% CIs)	AIC
7	**A**	0.78 (0.71, 0.84)	240.5
	B	0.79 (0.72, 0.85)	241.0
	C	0.79 (0.74, 0.85)	240.1
	D	0.81 (0.75, 0.87)	235.3
	E	0.81 (0.75, 0.87)	237.1
	F	0.81 (0.76, 0.87)	236.3
	G	0.82 (0.76, 0.87)	237.1
8	**A**	0.78 (0.72, 0.85)	238.5
	B	0.79 (0.73, 0.85)	239.8
	C	0.81 (0.75, 0.86)	234.4
	D	0.82 (0.76, 0.87)	233.6
	E	0.82 (0.76, 0.87)	235.6
	F	0.83 (0.78, 0.89)	229.6
	G	0.83 (0.78, 0.88)	231.3
9	**A**	0.77 (0.70, 0.83)	245.7
	B	0.77 (0.71, 0.84)	245.7
	C	0.79 (0.73, 0.85)	243.6
	D	0.80 (0.75, 0.86)	241.4
	E	0.80 (0.75, 0.86)	243.3
	F	0.81 (0.75, 0.87)	241.4
	G	0.81 (0.76, 0.87)	234.3
10	**A**	0.76 (0.69, 0.83)	244.6
	B	0.78 (0.72, 0.84)	244.4
	C	0.79 (0.73, 0.85)	243.3
	D	0.80 (0.75, 0.86)	240.5
	E	0.80 (0.75, 0.86)	242.5
	F	0.81 (0.75, 0.86)	242.0
	G	0.81 (0.76, 0.87)	243.4

Repetition numbers correspond to the fold that was left out (e.g. in repetition 1 the models were fitted using all participants in folds 2-10). See section 9.2.2 for details of the predictors included in models A to G.

AIC = Akaike's Information Criterion; AUC = Area under the curve.

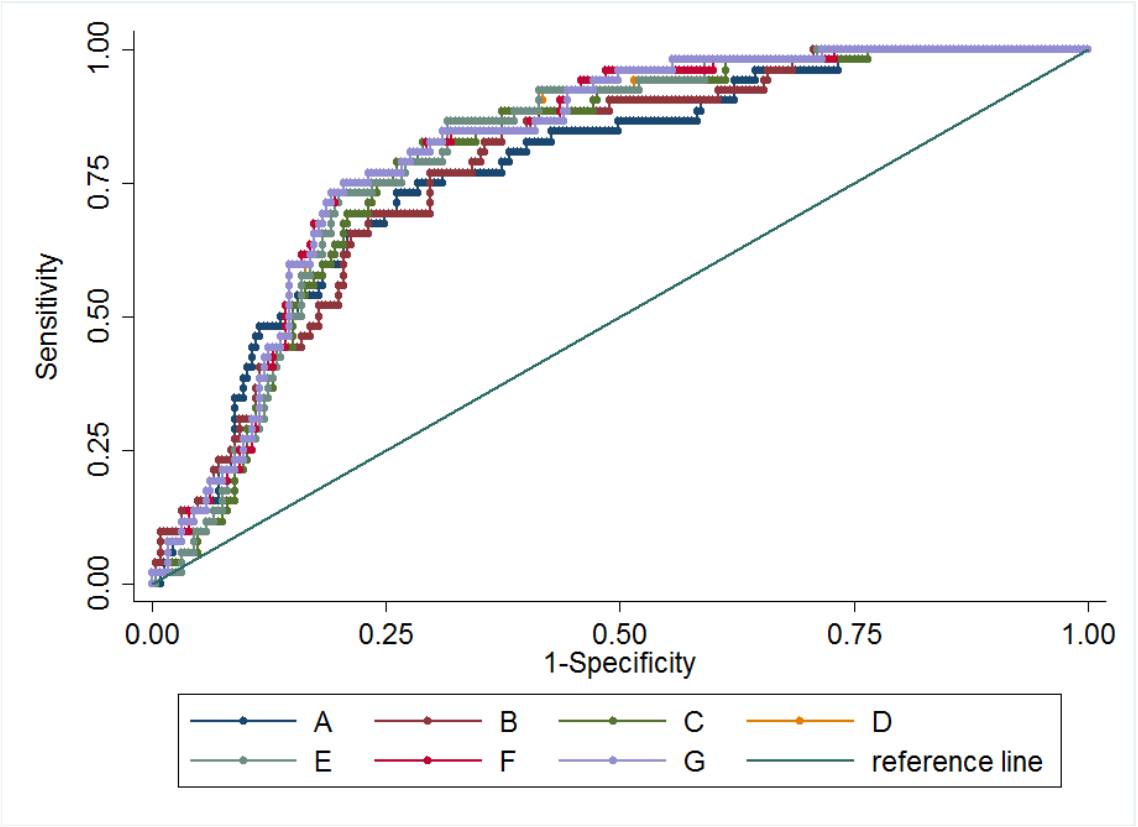


Figure 9-4. Receiver Operating Characteristic (ROC) curves for models A to G in repetition 1

Repetition 1 involved fitting the models to folds 2-10. Areas under the curves are quoted in Table 9-5. See section 9.2.2 for details of the predictors included in models A to G.

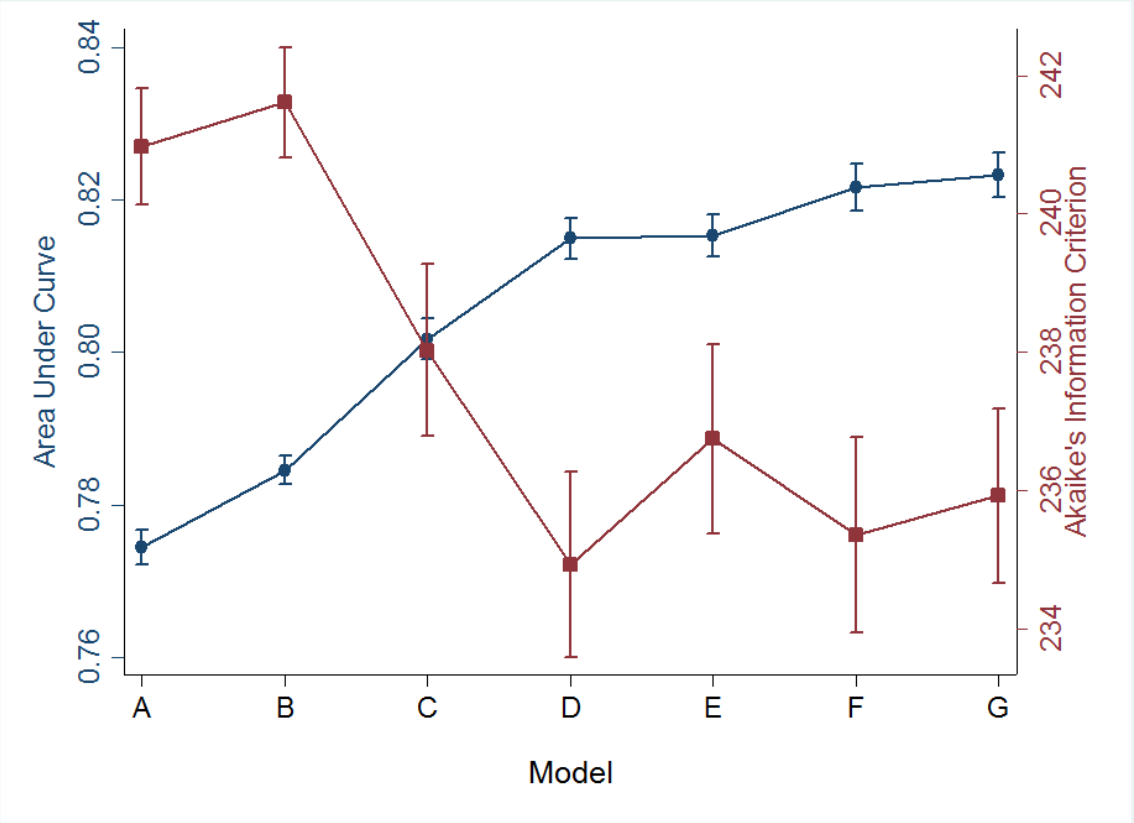


Figure 9-5. Mean AUC and AIC values for Models A to G

AUC (navy circles) is plotted on the left-hand y-axis. AIC (maroon squares) is plotted on the right-hand y-axis. Markers show the mean value across the ten repetitions of each model, and error bars show the standard error of the mean. A preferred model maximises AUC and minimises AIC. See section 9.2.2 for details of the predictors included in models A to G.
AUC = Area under the curve; AIC = Akaike's Information Criterion

9.2.3.2. Prediction of amyloid status in validation set

Based on the above results I selected model D for testing in the validation set (the group of 77 individuals whose data were not included in the model-selection process). I used model D to calculate a probability score for each participant – their predicted probability of being Aβ+ according to the model (see Note 5 of Figure 9-3).

The predicted probabilities were higher for Aβ+ (mean = 36%) than Aβ- participants (mean = 19%), although the difference was not statistically significant (z = -1.73, p = 0.08). To examine whether the model was doing any better than simply accounting for *APOE*-ε4, I compared the predicted probabilities for Aβ+ and Aβ- participants in the ε4-carrier and non-carrier groups separately. In the ε4-carrier group, the predicted probabilities were higher on average for Aβ+ participants than Aβ- participants (z = -3.10, p =0.002), but in the non-carrier group the predicted probabilities did not differ between Aβ+ and Aβ- participants (z = 0.21, p = 0.83) (Figure 9-6).

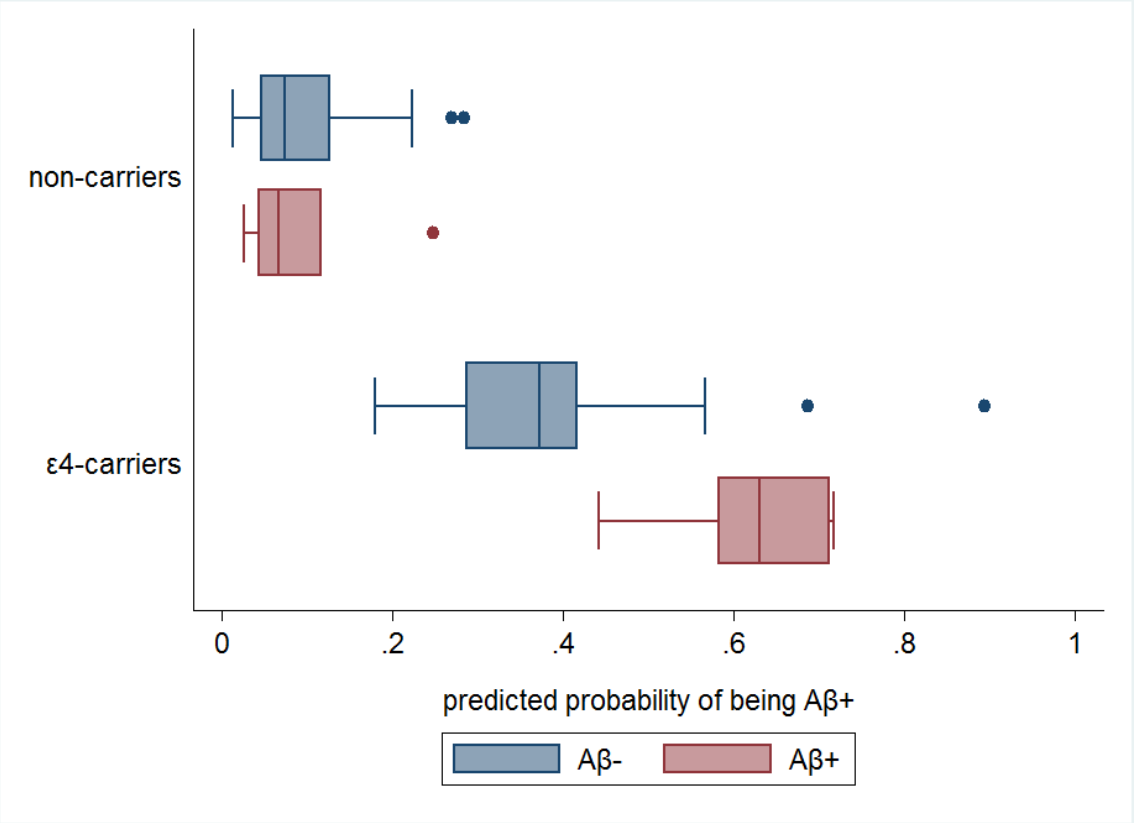


Figure 9-6. Probabilities of β-amyloid positivity for participants in the validation set, based on predictions from Model D

Box plots show the distribution of probability scores, which are predictions of the probability that an individual is β-amyloid positive. These probability scores were calculated based on the equations for Model D that were derived from the cross-validation process (see Note 5 of Figure 9-3). APOE-ε4 carriers and non-carriers are plotted separately, further separated into Aβ+ and Aβ- (the true amyloid status according to PET scan). Numbers of participants in each group are as follows: non-carrier Aβ- = 43; non-carrier Aβ+ = 7; ε4-carrier Aβ- = 20; ε4-carrier Aβ+ = 7.

9.2.4. Conclusions

In summary, it is not possible to conclude from these results that one combination of cognitive measures was more sensitive to dichotomous amyloid status than another. There was an indication that sensitivity (true positives) was improved by the inclusion of additional cognitive measures (Matrix Reasoning and Choice RT consistency) compared to the PACC alone, but the evidence was inconclusive due to the lack of statistically significant differences in accuracy between the various models. This is not unexpected given that the differences in cognition between Aβ+ and Aβ- participants were so subtle and the majority of variance in cognitive performance was unexplained (see 9.1.3), so cognitive scores are not a strong predictor of amyloid status on an individual level at age 70. Once follow-up data have been collected and changes in cognitive performance over

time can be used as predictors in the model, I would like to repeat this approach to investigate the most sensitive combination of measures for detecting Aβ-related decline in cognition.

The results from testing the chosen model in the validation set confirmed that the chosen model does not accurately predict dichotomous amyloid status on an individual level. On a group level (i.e. does the model predict amyloid status better than predictions purely based on *APOE* genotype?), the results were in the expected direction for ε4-carriers, but again do not constitute convincing evidence, and statistical comparisons were limited by the small sample size of the validation set. The question of whether these sorts of predictive models may have practical applications is explored in the following section.

9.3. Potential applications for clinical practice and clinical trials

The final section of this chapter focuses on potential applications of the results reported in this thesis. I decided to approach this in two ways. Firstly, it could be of value in a clinical setting to estimate the likelihood of Aβ positivity for an individual who does not meet criteria for dementia or MCI but falls into an "impaired" range on one or more cognitive tests. Secondly, eligibility for future therapeutic trials in preclinical AD is likely to be based on having an Aβ+ PET scan. This was the case in the first clinical trial in preclinical AD, the A4 trial (Sperling *et al.*, 2014), which is currently still ongoing, as well as in the Janssen EARLY trial https://clinicaltrials.gov/ct2/show/NCT02569398?term=JNJ-54861911, which was terminated in 2018 due to liver toxicity https://www.janssen.com/update-janssens-bace-inhibitor-program. It would be valuable to know whether the addition of a cognitive assessment into the screening process could improve efficiency by reducing the number of PET scans required to recruit the desired number of Aβ+ individuals. These two potential applications are explored in the following two sub-sections. This section is purely exploratory and partly a vehicle for me to reflect on the meaning of the analyses and results I have reported.

9.3.1. Application for clinical practice

9.3.1.1. Rationale

In clinical neuropsychology, the standard definition of cognitive impairment is below the fifth percentile of the normal healthy population. If an individual shows evidence of cognitive impairment on a clinical neuropsychological assessment, the next steps in the diagnostic process will be guided by evidence about the likelihood of possible causes. I decided to investigate the prevalence of Aβ positivity in the lowest-performing five percent of Insight 46 participants, as this could be of potential use to clinicians when dealing with patients who show similar evidence of possible cognitive impairment. I hypothesised that Aβ+ participants would be more likely to fall into the "impaired" range.

9.3.1.2. Methods

Before defining the fifth percentile of performance on each test, participants with dementia (n=3) were excluded. Participants with other neurological and psychiatric conditions (see section 3.2.3) were not excluded as the criteria for those conditions were primarily designed for analyses involving biomarkers (to ensure exclusion of potentially confounding comorbidities) and do not necessarily imply clinical symptoms or cognitive impairment. As this analysis aims to be relevant to the general population who may present to clinic without having had prior neuroimaging or screening for MCI, the sample was kept as representative as possible. From the sub-sample of dementia-free participants with complete biomarker data (n=443), participants who performed below the fifth percentile on each cognitive outcome were identified (see Table 9-1 for definitions of the 19 different outcomes). For each cognitive outcome, a chi-square test was conducted to assess whether the prevalence of Aβ positivity among individuals in this "impaired" group was higher than in the "non-impaired" group (those whose performance was greater than or equal to the fifth percentile).

For each participant, I totalled up the number of cognitive measures on which their performance was below the fifth percentile. As some participants were missing data for one or more cognitive tests (mostly due to technical problems – see section 3.3), this total was converted into a proportion by dividing it by the number of tests completed. Wilcoxon's rank sum test was used to assess whether Aβ+ participants tended to fall into the "impaired" range on a greater proportion of tests than Aβ- participants.

9.3.1.3. Results

The lowest-performing five percent of participants equates to approximately 22 individuals in this sample, but the exact number varies from one cognitive test to another depending on the distribution of scores.

In sixteen out of the nineteen cognitive outcomes, the prevalence of Aβ positivity in the lowest-performing five percent of participants was greater than 18.3%, which was the prevalence in the sample as a whole (Figure 9-7). However, these differences were not statistically significant, with the exception of the Matrix Reasoning test, on which the prevalence of Aβ positivity in the "impaired" group (38%) was significantly higher than in the "non-impaired" group ($\chi^2 = 6.0$, $p = 0.014$).

On average, Aβ+ participants performed below the fifth percentile on a greater proportion of tests than Aβ- participants (Aβ+ median = 0.05; Aβ- median = 0; $z = -2.24$, $p = 0.02$). For Aβ+ participants, the median value is equivalent to performing below the fifth percentile on one out of nineteen tests.

9.3.1.4. Conclusions

With the exception of the Matrix Reasoning test, these results do not support the hypothesis that Aβ+ participants were more likely to fall into the "impaired" range on cognitive tests, although statistical comparisons between the "impaired" and "non-impaired" groups were limited by the small sample size of the "impaired" group. It is interesting that Matrix Reasoning was the test on which Aβ+ participants were most likely to fall into the "impaired" range, as this complements the results of the main analyses where the Matrix Reasoning stood out as the test with the largest group difference between Aβ+ and Aβ- participants (see section 9.1). However it is important to note that Matrix Reasoning scores were generally high in Insight 46 as discussed previously (see Chapter 4) so it may be misleading to refer to the lowest 5% of scores as an "impaired" range relative to the wider population. Further studies would be needed to validate normative data for healthy adults of this age.

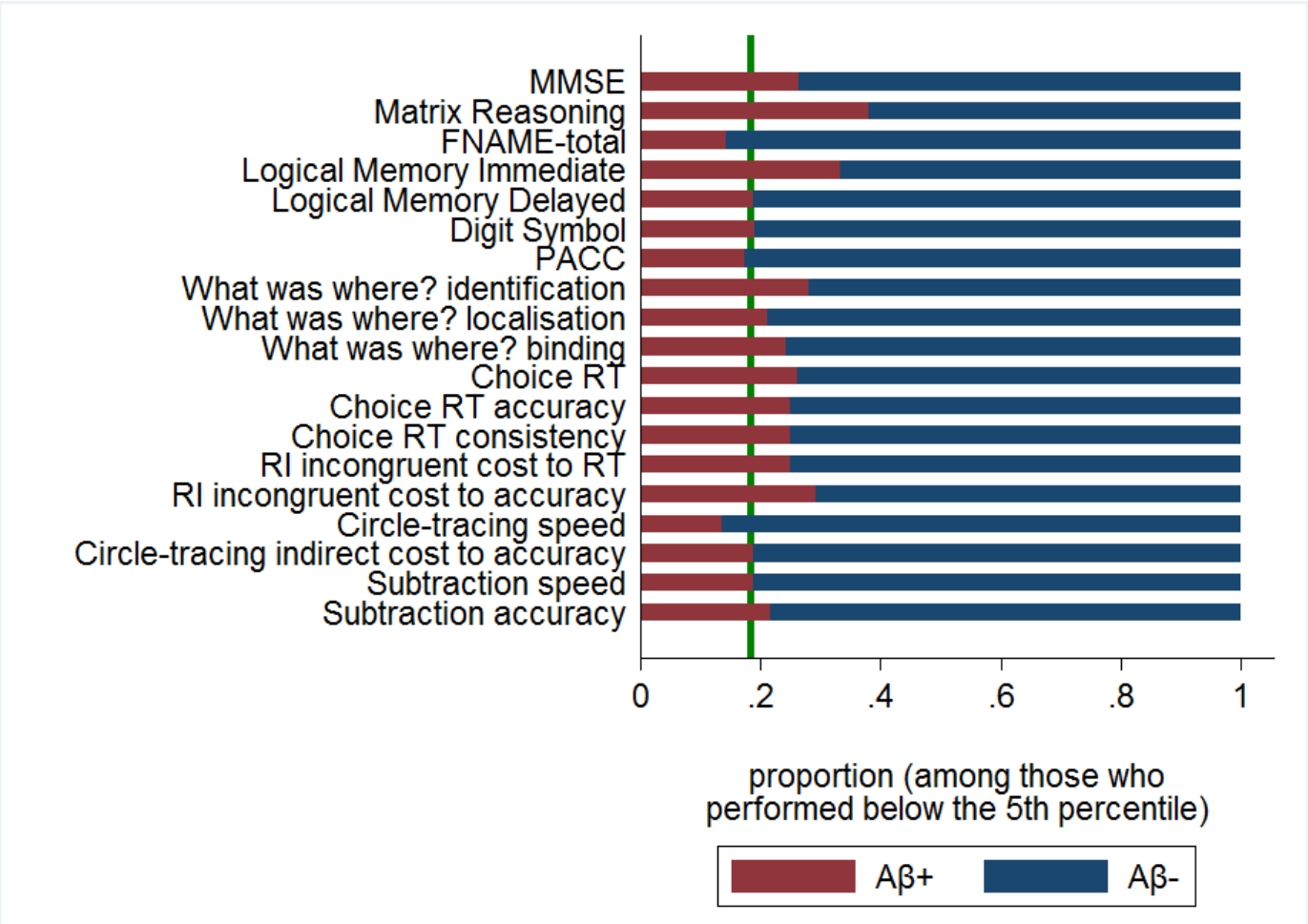


Figure 9-7. Amyloid status of the lowest-performing 5% of participants on each cognitive outcome

The vertical green line at 0.183 indicates the proportion of β-amyloid positive participants in the sample as a whole. MMSE = mini mental state examination; PACC = Preclinical Alzheimer Cognitive Composite; RI = Response Inhibition; RT = reaction time

9.3.2. Application for clinical trial recruitment

9.3.2.1. Rationale

The second application I chose to explore is whether a sensitive battery of cognitive measures may improve the efficiency of screening for a clinical trial in preclinical AD, where eligibility is conditional on having an Aβ+ PET scan. As the prevalence of Aβ positivity is still relatively low in healthy older adults who would be the target population for such a trial (Jansen *et al.*, 2015), screening is a time-consuming and expensive process (~£1500 per PET scan) with a low success rate. Cognitive assessment is non-invasive and cheap, so could potentially be used to pre-screen large numbers of people, in order to decide which individuals should proceed to have a PET scan. The purpose of the pre-screening process would be to exclude individuals who are highly likely to be Aβ-. A recent analysis of three large cohorts concluded that a pre-screening algorithm for Aβ positivity based on cognitive, genetic and socio-demographic predictors could reduce recruitment costs by about 20% by reducing the number of PET scans required (Ansart *et al.*, 2019).

In section 9.2 I concluded that the optimum combination of cognitive measures for detecting Aβ positivity was the following three, with the caveat that this model was not conclusively better than the others: PACC, Matrix Reasoning, Choice RT consistency. In this section, I aimed to quantify the extent to which this cognitive battery might improve the efficiency of the screening process and provide an estimate of cost-effectiveness. I decided to investigate this question in a scenario where *APOE* genotype is unknown, since that is the most realistic scenario for clinical trial recruitment. Although commercial companies offering *APOE* genotyping have appeared in the last few years, *APOE* testing is not available clinically due to its low prognostic value on an individual level, and is generally not recommended (National Institute on Aging, 2015; Alzheimer's Society, 2016a), so the vast majority of people do not know their *APOE* genotype. To date, clinical trials targeting Aβ+ cognitively-normal older adults – the A4 trial (Sperling *et al.*, 2014) and the Janssen EARLY trial https://clinicaltrials.gov/ct2/show/NCT02569398?term=JNJ-54861911 – have used Aβ-PET to determine eligibility and have not included *APOE* genotyping in the screening process.

9.3.2.2. Methods

I imagined a clinical trial which aims to recruit 100 Aβ+ individuals, and compared two screening scenarios: 1) no pre-screening; 2) pre-screening with a cognitive assessment.

In the "no pre-screening" scenario, the number of individuals who should have a PET scan to meet the recruitment target can be estimated from the prevalence of Aβ positivity among cognitively-normal participants in the Insight 46 sample (18.3%).

For the "pre-screening scenario" I conducted a Receiver Operating Characteristic (ROC) analysis to select an optimal cut-point that would be used as a threshold for determining whether or not participants proceeded to have a PET scan. The outcome was amyloid status, and the model included the following cognitive measures as predictors: PACC, Matrix Reasoning and Choice RT consistency (see Table 9-1 for definitions). The sample was the 406 cognitively-normal participants with complete biomarker data.

I chose an optimal cut-point by maximising Youden's index, which is defined as follows:

Youden's index = Sensitivity + Specificity – 1

Youden's index falls between 0 for a useless model (predicts the outcome no better than chance) and 1 for a perfect test (predicts the outcome perfectly).

For any screening test we have a 2x2 table as follows:

	TEST	
TRUE	Positive	Negative
Positive	true positive (TP)	false negative (FN)
Negative	false positive (FP)	true negative (TN)

These values relate to the ROC analysis according to the following three equations:

1. *Sensitivity = TP / (TP + FN)*
2. *Specificity = TN / (TN + FP)*
3. *Accuracy = (TP + TN) / (TP + TN + FN + FP)*

At any given cut-point, values are defined for the model's sensitivity, specificity and accuracy (proportion of subjects correctly classified). Therefore if a clinical trial needs to recruit 100 Aβ+ individuals, the value of "true positives" is set to 100 and the other three values (false negative, false positive, true negative) can be determined by simultaneously solving the three equations. The number of individuals required at each

stage of the screening process to achieve the recruitment target can then be calculated as follows:

Number required for pre-screening = TP + FN + TN + FP

Number who would proceed from pre-screening to PET scan = TP + FP

To evaluate cost-effectiveness I assumed that a PET scan costs £1500 and a cognitive assessment costs £50 (similar to the estimated cost assumed in Ansart *et al.* (2019)).

9.3.2.3. Results

A ROC curve for the pre-screening scenario is shown in Figure 9-8, with the chosen cut-point. Note that the lower bound of the 95% confidence intervals for the area under the curve is greater than 0.5, which is evidence that the cognitive assessment enables amyloid status to be predicted better than chance – this can also be concluded from the chi-square value of the model ($\chi^2 = 11.7$, $p = 0.009$).

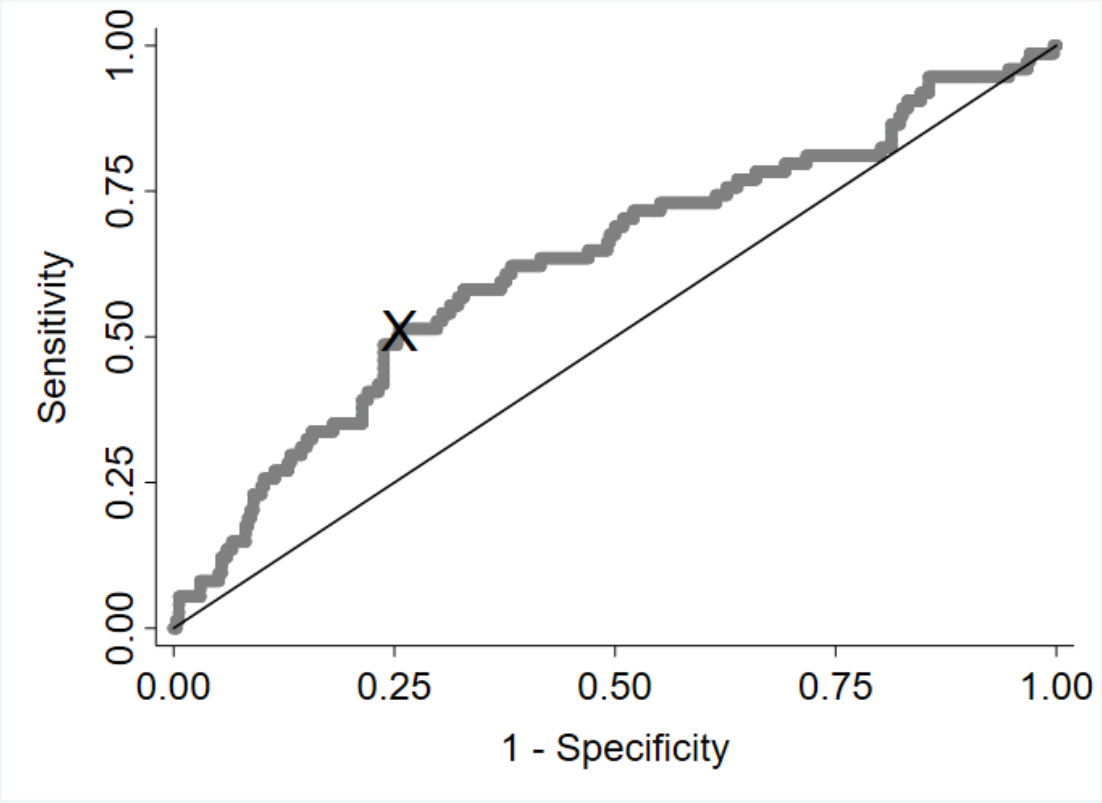


Figure 9-8. ROC curves for pre-screening scenario

Area under the curve = 0.63 (95% confidence intervals 0.55 to 0.70). Black cross indicates the chosen cut-point, which maximises Youden's index. Youden's index can be visualised as the vertical distance from the cross to the reference line.

The equation for the model takes the form of Equation 3 (page 264), and is given below:

$$p(\mathrm{A\beta+}) = 1/(1 + e^{(1.52 + 0.17*\mathrm{P} + 0.23*\mathrm{M} + 0.28*\mathrm{R})})$$

(Equation 4)

where p(Aβ+) is the probability of being amyloid positive; P = PACC score; M = Matrix Reasoning z-score; R = Choice Reaction Time consistency z-score

The chosen cut-point (see Figure 9-8) corresponds to a 20% probability of being Aβ+, meaning that participants with a probability of less than 20% would not proceed to a PET scan. Table 9-6 shows the solutions of the equations for this pre-screening scenario.

Table 9-6. Solution of the equations for the number of individuals required for pre-screening, and number of individuals who would proceed to a screening PET scan

Sensitivity	0.51
Specificity	0.74
Accuracy	0.70
TP	100
FN	95
TN	651
FP	224
Number of people required for pre-screening	1069
Number of people proceeding to PET scan	324

See 9.3.2.2 for explanation of the solution of the simultaneous equations for sensitivity, specificity and accuracy. TP = true positives; FN = false negatives; TN = true negatives; FP = false positives.

Based on the prevalence of Aβ positivity in the Insight 46 sample, if no pre-screening were conducted then 548 individuals would need to be scanned to meet the recruitment target of 100 Aβ+ individuals. By pre-screening 1069 individuals with the cognitive assessment, 324 PET scans would be required – a reduction of 41%. This would be cost-effective as the £53,450 cost of conducting the cognitive assessments would be outweighed by the £336,000 reduction in scan costs.

9.3.2.4. Conclusions

In this section I explored the potential application of a cognitive assessment to as part of the screening process for therapeutic trials in preclinical AD, where the target population is cognitively-normal older adults with elevated β-amyloid. The results suggested that my chosen battery of cognitive measures was able to predict an individual's amyloid status better than chance in the Insight 46 sample, and therefore including a cognitive assessment in the pre-screening process may be worth considering in the future as it could improve efficiency and cost-effectiveness by reducing the number of expensive PET scans required.

This was an exploratory analysis conducted primarily as a demonstration of method, and there are a number of factors that limit its relevance to real-word therapeutic trials. Firstly, as the sample only contains 69- to 71-year-olds, and increasing age is associated with both cognitive decline and increasing likelihood of Aβ positivity, the model would not generalise to a wider age range. Secondly, my approach was to choose a cut-point that optimised the accuracy of the predictive model and then to evaluate its cost-effectiveness, but if cost-effectiveness was the primary goal then it may be preferable to

choose a cut-point that minimised cost rather than maximising accuracy (as was done in (Ansart *et al.*, 2019)). Alternatively for a screening test it may be most important to optimise sensitivity (i.e. to minimise false negatives).

In summary, while it is clear that a cross-sectional cognitive assessment is not a good predictor of amyloid status on an individual basis (within cognitively-normal older adults), it may be of potential use in a clinical trial recruitment scenario for identifying individuals who are more likely to be Aβ+.

10. GENERAL DISCUSSION

10.1. Summary

In this large population-based sample of older adults of approximately the same age, I investigated predictors of performance on a range of cognitive measures, including some standard paper-and-pencil tests and some more novel computerised tests. My key findings are summarised as follows, with reference to the overarching hypotheses stated in section 2.7.2:

i) As expected, higher childhood cognitive ability was strongly associated with better performance on almost all cognitive tests, and there were independent effects of educational attainment and socioeconomic position on some cognitive outcomes.

ii) As hypothesised, subtle sex differences in cognition were detected on a range of cognitive measures. On average, women performed better on measures of memory and processing speed, whereas men tended to be faster at serial subtraction, more accurate at reporting the locations of objects, and slightly faster at responding to a complex reaction time task.

iii) Consistent with the hypothesis that cognitively-normal participants with elevated Aβ deposition would show evidence of subtle cognitive deficits, Aβ+ participants performed less well than Aβ- participants on a range of cognitive measures. Importantly, these subtle deficits covered a range of cognitive domains – not just memory – and included some measures where such differences have not been reported before. Differences between the group means were of the order 0.3 *SD*. The continuous measure of Aβ (SUVR) also showed associations with poorer performance on some cognitive measures.

iv) In general, white matter pathology and whole brain volume were not strongly associated with cognitive performance, but there was evidence of some specific associations that are consistent with previous studies: greater whole brain volume was associated with faster performance on three different timed measures and greater white matter hyperintensity volume (WMHV) was associated with slower processing speed.

v) The hypothesis that composite and computerised cognitive measures would be more sensitive to brain pathology than standard paper-and-pencil cognitive tests

was not supported overall. There was evidence that some computerised measures were among the most sensitive to Aβ, but some of the non-computerised measures performed similarly and the most sensitive measure in this study was in fact Matrix Reasoning (a standard test of non-verbal reasoning). The Preclinical Alzheimer's Cognitive Composite (PACC) was not better at discriminating Aβ+ and Aβ- individuals than some of the individual tests.

Regarding associations between *APOE*-ε4 and cognition (about which I did not make an overarching hypothesis), there was no evidence of a detrimental effect of *APOE*-ε4 on cognition after accounting for Aβ pathology, but rather there was evidence that *APOE*-ε4 was associated with better performance on measures of short-term memory.

These results are discussed in more detail below, followed by a discussion of the strengths and limitations of this study, and directions for future work.

10.2. Key results and interpretation

10.2.1. Effects of childhood cognitive ability, education and adult socioeconomic position

The association between higher childhood cognitive ability and better cognitive performance more than 60 years later was seen across most cognitive measures, covering a wide range of cognitive domains, and is consistent with previous NSHD analyses (Richards and Sacker, 2003; Richards *et al.*, 2004, 2019; Davis *et al.*, 2017; Philippou *et al.*, 2018). The finding that educational attainment and adult socioeconomic position were associated with many cognitive outcomes, *independent of childhood cognition,* is also consistent with previous NSHD analyses (Richards and Sacker, 2003; Richards *et al.*, 2019), which have concluded that these factors are only moderately correlated (see section 2.4.2). Clearly educational and occupational attainment are affected by many factors other than cognitive ability; among Insight 46 participants one of these factors was sex, as males had on average 2 years more education (likely a reflection of cultural norms at the time) and were more likely to work in jobs with a higher occupational complexity, despite there being no sex difference in childhood cognitive ability. Other predictors of education and adult socioeconomic position in NSHD are mother's education and father's socioeconomic position (Richards and Sacker, 2003; Richards et al., 2019). As most studies of ageing and neurodegenerative disease do not

have access to measures of prior cognitive ability, educational attainment is often used as a proxy, but the results reported in this thesis highlight the limitations of this.

The independent effects of education and adult socioeconomic position on later-life cognition are consistent with evidence that cognitive ability is not fixed from birth but is shaped by various influences throughout life (Richards and Deary, 2013). A recent meta-analysis of the effects of education on IQ reported an increase of approximately 1 to 5 IQ points per additional year of education, and found that these effects persisted into older age (Ritchie and Tucker-Drob, 2018). Another analysis from the NSHD found that education and training undertaken during adulthood was associated with higher cognitive ability at age 53, after accounting for childhood cognitive ability and educational attainment up to age 26 (Hatch *et al.*, 2007).

The evidence for independent effects of childhood cognitive ability, education and adult socioeconomic position on cognitive performance at age ~70 is also consistent with evidence that these factors may be protective against cognitive decline in later life. Multiple longitudinal studies have reported an association between higher childhood cognitive ability and reduced risk of dementia (Snowdon *et al.*, 1996; Whalley *et al.*, 2000; McGurn *et al.*, 2008; Russ *et al.*, 2017; Huang *et al.*, 2018). Higher educational and occupational attainment are similarly associated with a reduced risk of AD, and have been identified as potentially modifiable factors that could be considered as targets for intervention (Stern, 2012; Norton *et al.*, 2014; Xu *et al.*, 2015). Paradoxically, it has been reported that highly-educated cognitively-normal older adults tend to have a greater burden of Aβ pathology (Jansen *et al.*, 2015; Arenaza-Urquijo *et al.*, 2017) and neurodegeneration (Pettigrew *et al.*, 2016). The concept of 'cognitive reserve' has been proposed to account for this, positing that individual differences in the way that cognitive tasks are performed allow some individuals to tolerate a greater burden of pathology than others before they experience cognitive decline, and that this resilience or 'reserve' can be increased by factors such as education, stimulating activities, healthy lifestyle and social support (Stern, 2012; Chan *et al.*, 2018; Russ, 2018). There was no evidence of this divergence occurring in Insight 46, as the Aβ+ and Aβ- groups did not differ in their childhood cognitive ability, educational attainment or adult socioeconomic position (see section 3.6). However, as rates of cognitive impairment are extremely low at age 70, it may be that differences in resilience to pathology become apparent in future when some individuals begin progressing to dementia.

10.2.2. Sex differences

The sex differences in cognition detected in this study are consistent with typical patterns of sex differences reported in the literature. Women performed better than men on measures of verbal memory (Logical Memory immediate and delayed recall, recall for face-name and face-occupation associations on the FNAME-12 task). A female advantage on verbal tasks is well-documented, with one proposed explanation being differences in cognitive strategies: women are reported to show an increased tendency to cluster items to be remembered according to semantic and phonological categories (Andreano and Cahill, 2009), and to engage in elaborative processing of the meaning of information to be remembered (Wirth *et al.*, 2007). Another proposed explanation is a sex difference in distribution of language processing across the brain, with some studies reporting that women show more bilateral activity whereas men show more left-lateralised activity (Andreano and Cahill, 2009), although others do not find such differences (Sommer *et al.*, 2004).

Sex differences in verbal encoding could also explain my finding of a subtle advantage for women on recalling the identity of objects on the "What was where?" visual short-term memory binding task, as I noted from anecdotal reports that participants were adopting verbalisation strategies (Andreano and Cahill, 2009). Verbal encoding has also been proposed as an explanation for the consistent finding – replicated here in Insight 46 – of superior performance for women on the Digit-Symbol substitution task (Royer, 1978; Majeres, 1983). However, a review of sex differences in processing speed tasks (including Digit-Symbol Substitution) concluded that memory does not play a big role in performance and suggested that the effect may be underpinned by female superiority in reading and writing skills, although the concept of processing speed itself remains poorly defined (Roivainen, 2011).

The measures on which male participants showed superior performance – serial subtraction speed, location memory in the "What was where?" task, and response times in the Response Inhibition task – are also consistent with the literature. A male advantage has been widely reported on mathematical (e.g. (Hyde and Mertz, 2009)), visuospatial (e.g. (Andreano and Cahill, 2009; Voyer, Voyer and Saint-Aubin, 2017)) and reaction time tasks (e.g. (Der and Deary, 2006; Roivainen, 2011)). As discussed in Chapters 5 and 8 respectively, the superior performance of males in mathematics and visuospatial abilities is reduced in countries with greater gender equality (Hyde and Mertz, 2009; Coutrot *et al.*, 2018), but biological explanations have also been proposed.

In summary, a predictable pattern of sex differences has emerged across the Insight 46 cognitive battery, and differences between males and females were in the region of 0.5

SD on average. This highlights the importance of accounting for sex differences when seeking to detect and track subtle cognitive decline in the preclinical stages of AD, as average differences between men and women in this study were often larger than differences between Aβ+ and Aβ- groups (typically ~0.3 *SD*). Once Insight 46 follow-up assessments are completed, it will be important to see whether any sex differences emerge in terms of changes in cognition over the ~2 year interval, as a similar study that conducted longitudinal assessment of 755 cognitively-normal older adults on the Preclinical Alzheimer Cognitive Composite (PACC) found that, although sex was not a direct predictor of cognitive decline, Aβ+ women declined faster than Aβ+ men (Buckley *et al.*, 2018).

10.2.3. Age effects

Despite the very narrow age range of Insight 46 participants (2.6 years – due to the time taken to collect the data, as all participants were born during the same week), unexpected associations between age and cognition were observed on three cognitive tests. Older participants tended to score lower on the Matrix Reasoning task and to have slower and more variable reaction times on the Choice RT task. Both these findings follow the expected direction of age effects but the magnitudes are significantly greater than could be expected (see discussions in sections 4.5.2.3 and 6.5.3.1). On the Choice RT task, the slower RT of older participants was accompanied by higher accuracy, which can be explained by a speed-accuracy trade-off. Conversely, on the circle-tracing task, older participants tended to trace more quickly and to make correspondingly more errors – an effect which has no obvious explanation (see discussion in section 8.5.3.1).

I considered the possibility of a recruitment bias i.e. that participants tested towards the end of the data collection period may have differed in some ways to those seen earlier. While participants were invited in a random order, inevitably some participants delayed their visits due to health problems, life circumstances or being initially undecided about taking part. However we did not find any evidence of differences in childhood cognitive ability, education or adult socioeconomic position (which were controlled for in analyses anyway), or in general health based on measures of self-rated health status and overall disease burden (see (James *et al.*, 2018) for a description of these measures). In terms of biomarker measures, older age was associated with greater white matter disease burden and smaller whole brain volume, but again the age effects on cognition were present despite controlling for these variables.

10.2.4. Associations with β-amyloid pathology

Investigating associations between dichotomous amyloid status and cognition was the primary focus of my research. As hoped, this work has produced novel evidence that contributes to the understanding of subtle cognitive changes that may be associated with preclinical AD pathology. These results are discussed below, followed by a discussion of their potential implications for clinical trials.

10.2.4.1. Evidence for differences between Aβ+ and Aβ- groups

The dominant narrative is that episodic memory is the first cognitive function to decline in AD (e.g. (Grober *et al.*, 2008; Mortamais *et al.*, 2017; Farrell *et al.*, 2018; Landau *et al.*, 2018)), although a closer examination of the literature reveals a nuanced picture, with reports of early declines in several areas including episodic function and visuospatial ability (e.g. (Baker *et al.*, 2017; Duke Han *et al.*, 2017) – see section 2.3). One particularly notable feature of my results is that statistically significant differences were observed between cognitively-normal Aβ+ and Aβ- on eight different outcomes covering a wide range of cognitive domains: MMSE; Logical Memory immediate recall; Matrix Reasoning test of non-verbal reasoning; intra-individual variability in reaction time (RT) on the Choice RT task; memory for object identity on the "What was where?" task; serial subtraction speed; and visuomotor integration on the circle-tracing task (the difference in accuracy between the conditions of direct and indirect visual feedback). To my knowledge, some of these have not been reported before – particularly Matrix Reasoning, intra-individual variability in RT, visuomotor integration and serial subtraction speed – and such a wide range of differences has not been reported before in a single study.

It is also worth noting that the results were in the predicted direction (poorer performance for Aβ+ participants) on the vast majority of other measures: face-name associative memory exam (FNAME-12); Logical Memory delayed recall, Digit-Symbol Substitution; object location memory on the "What was where?" task; mean Choice RT, accuracy on the Choice RT task; difference in accuracy between the congruent and incongruent conditions of the Response Inhibition task; subtraction accuracy. While these results should not be over-interpreted, this is consistent with the overall picture of the literature, where differences between Aβ+ and Aβ- groups are not consistently detected cross-sectionally but emerge more clearly when comparing longitudinal change scores (Baker *et al.*, 2017; Duke Han *et al.*, 2017; Mormino *et al.*, 2017).

In fact, it would not have been overly surprising if no cross-sectional differences had been observed at all in Insight 46, especially given the relatively young age of the cohort: one meta-analysis of cross-sectional differences between Aβ+ and Aβ- groups concluded that differences in episodic memory and executive function were statistically significant only in samples aged 75 years and above (Baker *et al.*, 2017). It is likely that the availability of unusually rich data on our participants' prior cognitive ability, plus the ability to adjust for other brain pathologies and *APOE*-ε4, increased the sensitivity of analyses to detect subtle effects of Aβ pathology, as other sources of intra-individual variation in cognition were accounted for more robustly than usual. As mentioned earlier, accounting for prior cognitive ability may be particularly necessary when investigating differences in high-performing individuals (Rentz *et al.*, 2004, 2007). With the important caveats that Aβ positivity is not sufficient to meet criteria for preclinical AD (tau pathology is also needed – see section 2.2.1) and that not all Aβ+ individuals are expected to develop symptoms in their life-times (see section 2.2.2), these results make a novel contribution to growing evidence that preclinical AD is accompanied by subtle cognitive decline in a broad range of cognitive domains, and that this decline begins many years before the onset of symptoms.

10.2.4.2. Implications for clinical trials in preclinical AD

The results of one of the first clinical trials in preclinical AD (currently ongoing) rest upon the Preclinical Alzheimer Cognitive Composite, since this is the primary outcome measure on which Aβ+ individuals on active treatment and placebo will be compared (Sperling *et al.*, 2014). My analyses have provided novel information about life-course determinants of the PACC – childhood cognitive ability, educational attainment and adult socioeconomic position – which should be considered when interpreting any differences between groups. In addition, the results reported in this thesis can be interpreted both to support and challenge the argument for using the PACC as an outcome measure in such trials, i.e. the argument that composites may be sensitive to deficits which are too subtle to be detected on individual cognitive tests. On the one hand, I identified a large number of tests whose results appeared to suggest subtle non-statistically significant deficits in Aβ+ individuals, including three of the four tests which make up our version of the PACC (Logical Memory delayed recall, Digit-Symbol Substitution; FNAME). Therefore, the fact that a statistically significant difference between the Aβ+ and Aβ- groups was detected on the PACC is due to the additive contributions of these small effects on its sub-tests. However, the effect size of the difference on the PACC (0.17 *SD*) was less than half of that of the most sensitive individual measures (Matrix Reasoning and intra-individual variability in Choice RT – see section 9.1.3) suggesting that the PACC could be improved

and it may be important to widen the focus from episodic memory and executive function to consider a broader range of cognitive domains.

However, as the PACC was designed to measure change in cognitive performance over time, it will be important to see how these other measures compare to the PACC in terms of longitudinal performance once the Insight 46 follow-up data have been collected. Rates of cognitive decline may differ from one cognitive domain to another (Cloutier *et al.*, 2015), and also different tests are differentially influenced by practice effects and test-retest reliability, so measures that appear promising cross-sectionally may not perform best for tracking longitudinal change.

Looking ahead to future clinical trials in preclinical AD, my results suggest that it may be cost-effective to include cognitive assessment in the screening process, to exclude individuals who are highly likely to be Aβ- and thus reduce the number of PET scans required to recruit the required number of Aβ+ individuals. However, my data do not allow me to recommend a specific screening process for several reasons: 1) my analyses were inconclusive with respect to which combination of cognitive tests was most sensitive to amyloid status; 2) the results cannot be generalised to participants outside this very narrow age range; 3) a more robust cost-effectiveness analysis would be needed to weigh up the feasibility of different options, including, for example, whether certain cognitive tests could be administered remotely.

10.2.5. Associations with whole brain volume, white matter pathology and APOE-ε4

While my main aim was to study Aβ, I also investigated whether cognition was independently associated with whole brain volume, global WMHV and *APOE*-ε4. Some interesting results emerged, with each of these variables showing specific associations with particular cognitive domains. The consistency of effects across tests, and their consistency with previous literature, allays concerns about spurious findings due to multiple comparisons.

10.2.5.1. APOE-ε4

In general, *APOE*-ε4 was not associated with cognition after adjustment for amyloid status, consistent with evidence that *APOE*-ε4 confers increased risk of cognitive decline primarily by increasing the likelihood of accumulation of Aβ plaques (Kline, 2012). However, on the “What was where?” task my analyses replicated two previous studies

in identifying an *advantage* for ε4-carriers in reporting the locations of objects (Zokaei *et al.*, 2017; Zokaei, Čepukaitytė, *et al.*, 2019). In addition, I found evidence of advantages for ε4-carriers in recalling the identities of objects on this task, and in immediate recall of the Logical Memory story (after adjustment for Aβ positivity, which was associated with poorer performance on these same measures); these results have not been reported before to my knowledge. These tasks are both measures of short-term memory after delays of a few seconds and rely heavily on attention. Logical Memory delayed recall (delay of ~20 minutes) showed a trend in the same direction (p = 0.06) – unsurprising given its strong correlation with the immediate recall score – but there was no evidence of an effect of *APOE*-ε4 on the other memory test in the Insight 46 battery (face-name associative memory exam (FNAME)).

These results could be interpreted to support the hypothesis of antagonistic pleiotropy, whereby the *APOE*-ε4 allele is associated with both beneficial and detrimental effects (see section 2.6.1). Investigation of this phenomenon tends to focus on the transition from beneficial to detrimental effects with age (Duke Han and Bondi, 2008; Tuminello and Duke Han, 2011; Smith, Ashford and Perfetti, 2019), but my results highlight the possibility that beneficial effects may persist, albeit often counteracted by Aβ, such that ε4-carriers who avoid accumulation of significant Aβ pathology may be the most cognitively advantaged on certain measures. This is consistent with evidence that among clinically-normal individuals aged ~90 years and above (i.e. those who have survived dementia-free beyond the prime ages for developing dementias), the cognitive performance of ε4-carriers appears to be at least as good as non-carriers (see review in (Tuminello and Duke Han, 2011)). The model of antagonistic pleiotropy accounts for evidence that ε4-carriers show increased brain activation in task-relevant regions, particularly frontal and parietal regions in tasks of short-term memory and attention (Duke Han and Bondi, 2008; Tuminello and Duke Han, 2011; Rusted *et al.*, 2013), and further suggests that this increased activation may be pronounced in older age as individuals preferentially recruit frontal brain regions to compensate for preclinical AD pathology in other regions (Duke Han and Bondi, 2008; Tuminello and Duke Han, 2011; Scheller *et al.*, 2017).

This study was limited in its ability to study the effects of *APOE*-ε4 by the fact that we combined all participants into two categories (ε4-carriers and non-carriers), an approach that was taken due to the rarity of the ε4/ε4 genotype and the desire not to exclude participants with ε2 alleles, since the main focus of the analyses was on Aβ. Further studies are required to investigate the effects of *APOE*-ε4 and Aβ in large groups of older adults with each specific *APOE* genotype.

10.2.5.2. White matter hyperintensity volume

My finding of an association between greater WMHV and slower processing speed (as measured by the Digit-Symbol Substitution task) is consistent with the results of previous studies (Gunning-Dixon and Raz, 2000; Oosterman *et al.*, 2004; Prins *et al.*, 2005; van Dijk *et al.*, 2008). As discussed in Chapter 4, this has implications for the interpretation of the PACC as an outcome measure in research studies and clinical trials, and it is important to remember that processing speed is a sensitive but non-specific measure of brain function which is compromised in many disorders (Jaeger, 2018). There was also an indication of an association between greater WMHV and poorer visuomotor integration (as evidenced by disproportionately inaccurate circle-tracing under the condition of indirect visual feedback – see Chapter 8). Aside from this, these results indicate that WMHV appear to have little effect on the cognitive performance of cognitively-normal older adults with a generally low burden of white matter disease.

10.2.5.3. Whole brain volume

As expected for 70-year-olds, there did not appear to be evidence of significant brain atrophy among Insight 46 participants (excluding individuals with major neurological and psychiatric conditions), although inferences about atrophy based on cross-sectional measures of brain volume are limited by the variability that exists between individuals. The finding that larger whole brain volume was associated with faster performance on three different timed measures is consistent with previous studies (Jackson *et al.*, 2012; Magistro *et al.*, 2015; Takeuchi *et al.*, 2017) and is notable for the diversity of these measures: Digit-Symbol Substitution, serial subtraction rate, and circle-tracing speed. However, as serial subtraction and circle-tracing were administered concurrently, performance on each task was affected by the other so further studies would be needed to test whether these findings are replicated if the tasks are administered separately.

10.3. Strengths and limitations

This study has a number of major strengths, foremost of which is the very small age range. Neurodegenerative diseases are notoriously difficult to disentangle from so-called normal or healthy ageing, since ageing is accompanied by increasing brain pathologies and neurodegeneration (Jack *et al.*, 2014; Parnetti *et al.*, 2019), and decline in most cognitive abilities (Glisky, 2007). This age-homogenous sample allowed hypotheses about the effects of brain pathologies to be clearly tested with this confound all but eliminated. Another major strength is the prospective collection of cognitive and

demographic data from birth, which provided the rare opportunity to investigate the effects of childhood cognitive ability on cognitive performance more than 60 years later. Thirdly, the large sample size made it possible to detect subtle effects including sex differences, differences between Aβ+ and Aβ- participants and associations with whole brain volume and WMHV. A large sample size was particularly important to the comparison of Aβ+ and Aβ- groups, since there was a difference in numbers of participants in the these groups – in line with expectations of the proportion of Aβ+ individuals at this age (Jansen *et al.*, 2015) – which reduced statistical power to detect differences between them.

As an investigation of preclinical AD, the main limitation of this study was the absence of tau-PET imaging. Since standard criteria for preclinical AD are based on the presence of both Aβ and tau pathology (see section 2.2), it was not possible to identify participants who meet criteria for preclinical AD, nor to investigate how Aβ and tau pathology may interact to affect cognition. Based on previous studies, somewhere around 30% of Aβ+ 70-year-olds would be expected to have tau pathology (Jack *et al.*, 2017; Kern *et al.*, 2018), and tau pathology will be present in some Aβ- individuals as well (possibly around 15-20% (Jack *et al.*, 2017)). As subtle cognitive deficits are reported to be greater in individuals with both pathologies, as opposed to those with Aβ alone (Soldan *et al.*, 2016; Duke Han *et al.*, 2017; Bilgel *et al.*, 2018; Ho and Nation, 2018; Sperling *et al.*, 2018), the associations between Aβ and poorer cognition reported in this thesis may be partially explained by tau pathology.

Another limitation of the study is that the computerised cognitive tests were all shortened versions that contained relatively few trials compared to most neuropsychological research studies, which limited the potential for detailed examination of patterns of performance on any individual test. However, this was a necessary trade-off when designing a ~90-minute cognitive battery that could be incorporated into a busy assessment day, and on the whole these results indicate that the battery was highly successful at capturing subtle differences in cognitive performance across a wide range of domains, as well as being well-tolerated by participants.

The generalisability of findings from this study rests on the extent to which the sample is representative of the population. This is a source of both strengths and limitations, which are discussed in the following sub-section.

10.3.1. Representativeness of the NSHD and Insight 46

The NSHD sample was originally designed to be representative of the general population, with the following caveats: 1) only single babies were included (not twins or multiples); 2) the sample was stratified by social class, taking all babies whose fathers had an agricultural or non-manual occupation, and one in four babies whose fathers had a manual occupation; 3) only babies born to married mothers were included, since the stratification by social class was based on the father's occupation (and in the 1940s it was relatively uncommon for unmarried couples to co-habit) (Wadsworth *et al.*, 2006). A comparison with census data when study members were aged 43 concluded that the cohort remained broadly representative of the UK population of British-born adults of the same age (Wadsworth *et al.*, 1992), and a similar comparison at ages 60-64 concluded that the cohort was representative in terms of socioeconomic position and rates of unemployment, although they were more likely to own a home and less likely to have limiting illness (Stafford *et al.*, 2013). As a native-born cohort reflecting the general British post-war population, all NSHD participants are white, so do not represent the more contemporary ethnic and cultural diversity of the wider population.

Inevitably, there is a bias for healthier study members to still be alive and participating in the cohort. At age 69, active NSHD participants (52% of the original cohort) were more likely to have higher childhood cognitive ability, higher educational attainment, a non-manual occupation, and better health status than those no longer active (Kuh *et al.*, 2016; Richards *et al.*, 2019). Despite their better health status on average, it is notable that only 15% of NSHD participants had no clinical disorders at age 60-64 (based on a list of 15 disorders e.g. cancer, hypertension, diabetes), and the average was 2 disorders each – a telling illustration of population ageing (Pierce *et al.*, 2012).

By drawing participants from the NSHD, Insight 46 is likely to be more representative than most studies in dementia research which recruit convenience samples or recruit through memory clinics and which may be biased towards those with higher education, higher socioeconomic position, and better cognition (Hultsch *et al.*, 2002; Brodaty *et al.*, 2014). Convenience samples of so-called healthy controls have been reported to have a higher prevalence of a family history of AD (Brodaty *et al.*, 2014) and higher rates of hippocampal atrophy than population-based samples (Whitwell *et al.*, 2012). In addition, the inclusive recruitment criteria and wide geographical area (participants came from across England, Wales and Scotland) distinguish Insight 46 from most research and clinical trial cohorts which tend to recruit participants who live near major urban centres (Tanner *et al.*, 2015) and have strict exclusion criteria for physical health problems and neurological or psychiatric conditions. Insight 46 participants, however, were required to be willing and able to attend a research visit in London, and on average had slightly

higher education and socioeconomic position than those who decided not to participate or were ineligible (James *et al.*, 2018). Within Insight 46, participants with missing neuroimaging data were more likely to be obese and to have mental health problems (James *et al.*, 2018). As obesity and depression are associated with increased dementia risk (Norton *et al.*, 2014), this raises the possibility that individuals with brain pathology and associated subtle cognitive decline may be underrepresented in these analyses.

10.4. Future directions

Questions about the nature and timing of cognitive decline in the preclinical stages of AD can best be answered with longitudinal cognitive assessment. Insight 46 participants are currently undergoing follow-up assessments (interval = ~2 years) which will be completed in summer 2020, so it will then be possible to assess changes in performance on each cognitive task, and to investigate whether Aβ pathology at baseline is predictive of relatively poorer cognition at follow-up. Declines relative to baseline are expected, since numerous studies have reported an association between Aβ positivity and cognitive decline over a similar interval (see (Baker *et al.*, 2017; Duke Han *et al.*, 2017) for meta-analyses), but practice effects are also anticipated to play a role (Hassenstab *et al.*, 2015; Vemuri *et al.*, 2015; Machulda *et al.*, 2017). Predictive effects of baseline WMHV, whole brain volume and *APOE*-ε4 will also be explored.

Models of cognitive trajectories will also be able to incorporate measures of cognition across adulthood in the NSHD (see Table 3-2), which will allow examination of relationships between life-course cognition and later-life brain pathology.

It will also be important to examine changes in levels of brain pathologies over the ~2-year interval and how these relate to changes in cognition. As discussed in section 2.3.2, "amyloid accumulators" – individuals whose levels of Aβ are rising from an initially normal level – may be a group of particular importance for identifying the earliest changes in cognition and may be a suitable target group for future clinical trials (McMillan and Chételat, 2018).

Investigation of a broader range of biomarkers is also planned, including blood-based biomarkers, cortical thickness (a biomarker for neurodegeneration specified in the NIA-AA criteria for preclinical AD (Jack *et al.*, 2018)), microstructural neuroimaging measures, a polygenic risk score that incorporates other genetic risks for AD in addition to *APOE*-ε4, and measures of Aβ and tau pathology in CSF (as approximately 35% of the cohort have so far agreed to have a lumbar puncture at their follow-up assessment). Looking further ahead, approximately one third of Insight 46 participants have agreed to post-

mortem brain donation, which will ultimately allow direct investigation of pathologies and their relationship with cognition during life.

10.5. Closing summary

These results add to growing evidence of subtle cognitive decline associated with preclinical AD pathology and contribute novel data, with differences in cognition between Aβ+ and Aβ- groups reported in an unusually wide range of cognitive domains including some that have so far received little attention. These differences were detectable at an age when those who are destined to develop dementia are still likely to be many years from symptoms. The rich historical data available on this cohort enabled predictable variation between individuals to be accounted for in a manner unique among neuroimaging studies in the field of preclinical AD research, as well as making possible a detailed investigation of the life-course determinants of cognition at age 70. These results have implications for the interpretation of cognitive data measured in later life, and for the use of cognitive tests as outcome measures in future clinical trials that will hopefully provide long-awaited relief to those who live with the effects of this devastating disease.

REFERENCES

Abikoff, H. *et al.* (1987) 'Logical memory subtest of the wechsler memory scale: Age and education norms and alternate-form reliability of two scoring systems', *Journal of Clinical and Experimental Neuropsychology*. Taylor & Francis Group, 9(4), pp. 435–448. doi: 10.1080/01688638708405063.

Albanese, E. *et al.* (2012) 'No association between gain in body mass index across the life course and midlife cognitive function and cognitive reserve—The 1946 British birth cohort study', *Alzheimer's & Dementia*, 8(6), pp. 470–482. doi: 10.1016/j.jalz.2011.09.228.

Alegret, M. *et al.* (2015) 'Validation of the Spanish Version of the Face Name Associative Memory Exam (S-FNAME) in Cognitively Normal Older Individuals', *Archives of Clinical Neuropsychology*. Oxford University Press, 30(7), pp. 712–720. doi: 10.1093/arclin/acv050.

Alosco, M. L. *et al.* (2018) 'A Clinicopathological Investigation of White Matter Hyperintensities and Alzheimer's Disease Neuropathology', *Journal of Alzheimer's Disease*. IOS Press, 63(4), pp. 1347–1360. doi: 10.3233/JAD-180017.

Alzheimer's Association Calcium Hypothesis Workgroup (2017) 'Calcium Hypothesis of Alzheimer's disease and brain aging: A framework for integrating new evidence into a comprehensive theory of pathogenesis.', *Alzheimer's & dementia : the journal of the Alzheimer's Association*. Elsevier, 13(2), pp. 178-182.e17. doi: 10.1016/j.jalz.2016.12.006.

Alzheimer's Society (2016a) *Genetics of Dementia: Factsheet 405LP*. Available at: https://www.alzheimers.org.uk/sites/default/files/pdf/factsheet_genetics_of_dementia.pdf.

Alzheimer's Society (2016b) *Sight, perception and hallucinations in dementia: Factsheet 527LP*. Available at: https://www.alzheimers.org.uk/sites/default/files/pdf/sight_perception_and_hallucinations_in_dementia.pdf.

Alzheimer's Society (2018) *What is vascular dementia? Factsheet 402LP*. Available at: https://www.alzheimers.org.uk/sites/default/files/pdf/factsheet_what_is_vascular_dementia.pdf.

Andreano, J. M. and Cahill, L. (2009) 'Sex influences on the neurobiology of learning and memory', *Learning & Memory*. Cold Spring Harbor Laboratory Press, 16(4), pp. 248–266. doi: 10.1101/LM.918309.

Ansart, M. *et al.* (2019) 'Reduction of recruitment costs in preclinical AD trials: validation of automatic pre-screening algorithm for brain amyloidosis', *Statistical Methods in Medical Research*, 0(0), pp. 1–14. doi: 10.1177/0962280218823036.

Anstey, K. J. *et al.* (2005) 'Biomarkers, Health, Lifestyle, and Demographic Variables as Correlates of Reaction Time Performance in Early, Middle, and Late Adulthood', *The Quarterly Journal of Experimental Psychology Section A*, 58(1), pp. 5–21. doi: 10.1080/02724980443000232.

Arenaza-Urquijo, E. M. *et al.* (2017) 'Association between educational attainment and amyloid deposition across the spectrum from normal cognition to dementia: neuroimaging evidence for protection and compensation', *Neurobiology of Aging*. Elsevier, 59, pp. 72–79. doi: 10.1016/J.NEUROBIOLAGING.2017.06.016.

Aron, A. R. *et al.* (2004) 'A componential analysis of task-switching deficits associated

with lesions of left and right frontal cortex', *Brain*, 127, pp. 1561–1573. doi: 10.1093/brain/awh169.

ARUK Dementia Statistics Hub (no date). Available at: www.dementiastatistics.org (Accessed: 10 May 2019).

Aschenbrenner, A. J. *et al.* (2015) 'Alzheimer disease biomarkers, attentional control, and semantic memory retrieval: Synergistic and mediational effects of biomarkers on a sensitive cognitive measure in non-demented older adults.', *Neuropsychology*, 29(3), pp. 368–81. doi: 10.1037/neu0000133.

Austad, S. N. and Hoffman, J. M. (2018) 'Is antagonistic pleiotropy ubiquitous in aging biology?', *Evolution, Medicine, and Public Health*, pp. 287–294. doi: 10.1093/emph/eoy033.

Ayutyanont, N. *et al.* (2014) 'The Alzheimer's prevention initiative composite cognitive test score: sample size estimates for the evaluation of preclinical Alzheimer's disease treatments in presenilin 1 E280A mutation carriers.', *The Journal of clinical psychiatry*, 75(6), pp. 652–60. doi: 10.4088/JCP.13m08927.

Baddeley, A. (2000) 'The episodic buffer: a new component of working memory?', *Trends in Cognitive Sciences*. Elsevier Current Trends, 4(11), pp. 417–423. doi: 10.1016/S1364-6613(00)01538-2.

Baddeley, A. D. and Hitch, G. (1974) 'Working Memory', *Psychology of Learning and Motivation*. Academic Press, 8, pp. 47–89. doi: 10.1016/S0079-7421(08)60452-1.

Baker, J. E. *et al.* (2017) 'Cognitive impairment and decline in cognitively normal older adults with high amyloid-β: A meta-analysis', *Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring*, 6, pp. 108–121. doi: 10.1016/j.dadm.2016.09.002.

Baker, J. E. *et al.* (2018) 'Episodic Memory and Learning Dysfunction Over an 18-Month Period in Preclinical and Prodromal Alzheimer's Disease', *Journal of Alzheimer's Disease*. Edited by A. Gross. IOS Press, 65(3), pp. 977–988. doi: 10.3233/JAD-180344.

Ballesteros, S., Mayas, J. and Reales, J. M. (2013) 'Cognitive function in normal aging and in older adults with mild cognitive impairment.', *Psicothema*, 25(1), pp. 18–24. doi: 10.7334/psicothema2012.181.

Bateman, R. J. *et al.* (2010) 'Autosomal-dominant Alzheimer's disease: a review and proposal for the prevention of Alzheimer's disease', *Alzheimer's Research & Therapy*. BioMed Central, 3(1), p. 1. doi: 10.1186/alzrt59.

Bateman, R. J. *et al.* (2012) 'Clinical and biomarker changes in dominantly inherited Alzheimer's disease.', *The New England journal of medicine*, 367(9), pp. 795–804. doi: 10.1056/NEJMoa1202753.

Bateman, R. J. *et al.* (2017) 'The DIAN-TU Next Generation Alzheimer's prevention trial: Adaptive design and disease progression model and the DIAN-TU Pharma Consortium for the Dominantly Inherited Alzheimer Network', *Alzheimer's & Dementia*, 13, pp. 8–19. doi: 10.1016/j.jalz.2016.07.005.

Bélanger, S., Belleville, S. and Gauthier, S. (2010) 'Inhibition impairments in Alzheimer's disease, mild cognitive impairment and healthy aging: Effect of congruency proportion in a Stroop task', *Neuropsychologia*, 48(2), pp. 581–590. doi: 10.1016/J.NEUROPSYCHOLOGIA.2009.10.021.

Benton, A. (1986) 'Reaction time in brain disease: some reflections.', *Cortex; a journal devoted to the study of the nervous system and behavior*, 22(1), pp. 129–40. doi: https://doi.org/10.1016/S0010-9452(86)80037-5.

Bilgel, M. *et al.* (2018) 'Effects of amyloid pathology and neurodegeneration on cognitive change in cognitively normal adults', *Brain*. doi: 10.1093/brain/awy150.

Boenink, M. (2018) 'Gatekeeping and trailblazing: The role of biomarkers in novel guidelines for diagnosing Alzheimer's disease', *BioSocieties*, 13(1), pp. 213–231. doi: 10.1057/s41292-017-0065-0.

Bott, N. *et al.* (2018) 'Device-Embedded Cameras for Eye Tracking-Based Cognitive Assessment: Validation With Paper-Pencil and Computerized Cognitive Composites.', *Journal of medical Internet research*, 20(7), p. e11143. doi: 10.2196/11143.

Bozdogan, H. (1987) 'Model selection and Akaike's Information Criterion (AIC): The general theory and its analytical extensions', *Psychometrika*, 52(3), pp. 345–370. doi: 10.1007/BF02294361.

Braak, H. and Braak, E. (1991) 'Neuropathological stageing of Alzheimer-related changes', *Acta Neuropathologica*, 82(4), pp. 239–259. doi: 10.1007/BF00308809.

Brébion, G. (2001) 'Language Processing, Slowing, and Speed/Accuracy Trade-Off in the Elderly', *Experimental Aging Research*, 27(2), pp. 137–150. doi: 10.1080/036107301750073999.

Broadbent, A. (2014) 'Disease as a theoretical concept: The case of "HPV-itis"', *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences*, 48, pp. 250–257. doi: 10.1016/J.SHPSC.2014.07.010.

Brodaty, H. *et al.* (2014) 'Influence of population versus convenience sampling on sample characteristics in studies of cognitive aging.', *Annals of epidemiology*, 24(1), pp. 63–71. doi: 10.1016/j.annepidem.2013.10.005.

Brookmeyer, R. and Abdalla, N. (2018) 'Estimation of lifetime risks of Alzheimer's disease dementia using biomarkers for preclinical disease', *Alzheimer's & Dementia*. doi: 10.1016/j.jalz.2018.03.005.

Brownlee, J. (2016) *Gentle Introduction to the Bias-Variance Trade-Off in Machine Learning*. Available at: https://machinelearningmastery.com/gentle-introduction-to-the-bias-variance-trade-off-in-machine-learning/.

Brownlee, J. (2018) *A Gentle Introduction to k-fold Cross-Validation*. Available at: https://machinelearningmastery.com/k-fold-cross-validation/.

Buckley, R. F. *et al.* (2017) 'Functional network integrity presages cognitive decline in preclinical Alzheimer disease.', *Neurology*, 89(1), pp. 29–37. doi: 10.1212/WNL.0000000000004059.

Buckley, R. F. *et al.* (2018) 'Sex, amyloid, and APOE ε4 and risk of cognitive decline in preclinical Alzheimer's disease: Findings from three well-characterized cohorts', *Alzheimer's & Dementia*. doi: 10.1016/j.jalz.2018.04.010.

Bugg, J. M. *et al.* (2006) 'Age differences in fluid intelligence: contributions of general slowing and frontal decline.', *Brain and cognition*, 62(1), pp. 9–16. doi: 10.1016/j.bandc.2006.02.006.

Bulterijs, S. *et al.* (2015) 'It is time to classify biological aging as a disease.', *Frontiers in genetics*, 6, p. 205. doi: 10.3389/fgene.2015.00205.

Burnham, S. C. *et al.* (2016) 'Clinical and cognitive trajectories in cognitively healthy elderly individuals with suspected non-Alzheimer's disease pathophysiology (SNAP) or Alzheimer's disease pathology: a longitudinal study', *The Lancet Neurology*, 15(10), pp. 1044–1053. doi: 10.1016/S1474-4422(16)30125-9.

Burton, C. L. *et al.* (2006) 'Intraindividual Variability as a Marker of Neurological

Dysfunction: A Comparison of Alzheimer's Disease and Parkinson's Disease', *Journal of Clinical and Experimental Neuropsychology*. Taylor & Francis Group, 28(1), pp. 67–83. doi: 10.1080/13803390490918318.

Cardoso, J. *et al.* (2013) 'STEPS: Similarity and Truth Estimation for Propagated Segmentations and its application to hippocampal segmentation and brain parcelation.', *Medical image analysis*, 17(6), pp. 671–84. doi: 10.1016/j.media.2013.02.006.

Caselli, R. J. *et al.* (2014) 'The neuropsychology of normal aging and preclinical Alzheimer's disease.', *Alzheimer's & dementia : the journal of the Alzheimer's Association*, 10(1), pp. 84–92. doi: 10.1016/j.jalz.2013.01.004.

Caselli, R. J. *et al.* (2018) 'Personality Changes During the Transition from Cognitive Health to Mild Cognitive Impairment', *Journal of the American Geriatrics Society*. doi: 10.1111/jgs.15182.

Castel, A. D. *et al.* (2007) 'Spatial attention and response control in healthy younger and older adults and individuals with Alzheimer's disease: Evidence for disproportionate selection impairments in the simon task.', *Neuropsychology*, 21(2), pp. 170–182. doi: 10.1037/0894-4105.21.2.170.

Chan, D. *et al.* (2018) 'Lifestyle activities in mid-life contribute to cognitive reserve in late-life, independent of education, occupation, and late-life activities', *Neurobiology of Aging*. Elsevier, 70, pp. 180–183. doi: 10.1016/J.NEUROBIOLAGING.2018.06.012.

Christ, B. U., Combrinck, M. I. and Thomas, K. G. F. (2018) 'Both Reaction Time and Accuracy Measures of Intraindividual Variability Predict Cognitive Performance in Alzheimer's Disease.', *Frontiers in human neuroscience*, 12, p. 124. doi: 10.3389/fnhum.2018.00124.

Christensen, H. *et al.* (2005) 'Within-Occasion Intraindividual Variability and Preclinical Diagnostic Status: Is Intraindividual Variability an Indicator of Mild Cognitive Impairment?', *Neuropsychology*, 19(3), pp. 309–317. doi: 10.1037/0894-4105.19.3.309.

Clark, L. R. *et al.* (2012) 'Specific measures of executive function predict cognitive decline in older adults.', *J Int Neuropsychol Soc.*, 18(1), pp. 118–27. doi: 10.1017/S1355617711001524.

Cloutier, S. *et al.* (2015) 'Patterns of Cognitive Decline Prior to Dementia in Persons with Mild Cognitive Impairment', *Journal of Alzheimer's Disease*, 47(4), pp. 901–913. doi: 10.3233/JAD-142910.

Cooper, R., Richards, M. and Kuh, D. (2017) 'Childhood cognitive ability and age-related changes in physical capability from midlife', *Psychosomatic Medicine*, p. 1. doi: 10.1097/PSY.0000000000000482.

Coughlan, G. *et al.* (2019) 'Toward personalized cognitive diagnostics of at-genetic-risk Alzheimer's disease.', *Proceedings of the National Academy of Sciences of the United States of America*, 116(19), pp. 9285–9292. doi: 10.1073/pnas.1901600116.

Coutrot, A. *et al.* (2018) 'Global Determinants of Navigation Ability', *Current Biology*, 28(17), pp. 2861-2866.e4. doi: 10.1016/j.cub.2018.06.009.

Dang, C. *et al.* (2018) 'Relationship Between Amyloid-β Positivity and Progression to Mild Cognitive Impairment or Dementia over 8 Years in Cognitively Normal Older Adults', *Journal of Alzheimer's Disease*, 65(4), pp. 1313–1325. doi: 10.3233/JAD-180507.

Dang, J. S., Figueroa, I. J. and Helton, W. S. (2018) 'You are measuring the decision to be fast, not inattention: the Sustained Attention to Response Task does not measure sustained attention', *Experimental Brain Research*, pp. 1–8. doi: 10.1007/s00221-018-

5291-6.

Davis, D. *et al.* (2016) 'Verbal memory and search speed in early midlife are associated with mortality over 25 years' follow-up, independently of health status and early life factors: a British birth cohort study.', *International Journal of Epidemiology*. Oxford University Press, p. dyw100. doi: 10.1093/ije/dyw100.

Davis, D. *et al.* (2017) 'Decline in Search Speed and Verbal Memory Over 26 Years of Midlife in a British Birth Cohort.', *Neuroepidemiology*, 49, pp. 121–128. doi: 10.1159/000481136.

Davis, R. *et al.* (2014) 'THink: Inferring Cognitive Status from Subtle Behaviors.', *Proceedings of the AAAI Conference on Artificial Intelligence*, 2014, pp. 2898–2905. Available at: http://www.ncbi.nlm.nih.gov/pubmed/27066307 (Accessed: 28 July 2017).

de Boer, C. *et al.* (2016) 'The Effect of Neurodegeneration on Visuomotor Behavior in Alzheimer's Disease and Parkinson's Disease.', *Motor control*, 20(1), pp. 1–20. doi: 10.1123/mc.2014-0015.

De Groot, J. C. *et al.* (2002) 'Periventricular cerebral white matter lesions predict rate of cognitive decline.', *Annals of neurology*, 52(3), pp. 335–41. doi: 10.1002/ana.10294.

de la Torre, J. (2018) 'The Vascular Hypothesis of Alzheimer's Disease: A Key to Preclinical Prediction of Dementia Using Neuroimaging', *Journal of Alzheimer's Disease*, 63(1), pp. 35–52. doi: 10.3233/JAD-180004.

Deary, I. J. and Der, G. (2005) 'Reaction Time, Age, and Cognitive Ability: Longitudinal Findings from Age 16 to 63 Years in Representative Population Samples', *Aging, Neuropsychology, and Cognition*, 12(2), pp. 187–215. doi: 10.1080/13825580590969235.

Deary, I. J., Der, G. and Ford, G. (2001) 'Reaction times and intelligence differences: A population-based cohort study', *Intelligence*, 29(5), pp. 389–399. doi: 10.1016/S0160-2896(01)00062-9.

Deary, I. J., Johnson, W. and Starr, J. M. (2010) 'Are processing speed tasks biomarkers of cognitive aging?', *Psychology and Aging*, 25(1), pp. 219–228. doi: 10.1037/a0017750.

Debette, S. and Markus, H. S. (2010) 'The clinical importance of white matter hyperintensities on brain magnetic resonance imaging: systematic review and meta-analysis.', *BMJ*, 341, p. c3666. doi: 10.1136/bmj.c3666.

Della Sala, S. *et al.* (2010) 'Dual Task During Encoding, Maintenance, and Retrieval in Alzheimer's Disease', *Journal of Alzheimer's Disease*, 19(2), pp. 503–515. doi: 10.3233/JAD-2010-1244.

Department of Education and Science (1972) *Burnham further education committee grading courses.* London.

Der, G. and Deary, I. J. (2006) 'Age and sex differences in reaction time in adulthood: results from the United Kingdom Health and Lifestyle Survey.', *Psychology and aging*, 21(1), pp. 62–73. doi: 10.1037/0882-7974.21.1.62.

Der, G. and Deary, I. J. (2017) 'The relationship between intelligence and reaction time varies with age: Results from three representative narrow-age age cohorts at 30, 50 and 69 years', *Intelligence*, 64, pp. 89–97. doi: 10.1016/J.INTELL.2017.08.001.

Diamond, A. (2013) 'Executive functions.', *Annual review of psychology*, 64, pp. 135–68. doi: 10.1146/annurev-psych-113011-143750.

Doebler, P. and Scheffler, B. (2016) 'The relationship of choice reaction time variability and intelligence: A meta-analysis', *Learning and Individual Differences*, 52, pp. 157–

166. doi: 10.1016/J.LINDIF.2015.02.009.

Donohue, M. C. *et al.* (2014) 'The preclinical Alzheimer cognitive composite: measuring amyloid-related decline.', *JAMA neurology*, 71(8), pp. 961–70. doi: 10.1001/jamaneurol.2014.803.

Donohue, M. C., Sperling, R. A., *et al.* (2017) 'Association Between Elevated Brain Amyloid and Subsequent Cognitive Decline Among Cognitively Normal Persons', *JAMA*, 317(22), pp. 2305–2316. doi: 10.1001/jama.2017.6669.

Donohue, M. C., Sun, C.-K., *et al.* (2017) 'Cross-validation of optimized composites for preclinical Alzheimer's disease', *Alzheimer's & Dementia: Translational Research & Clinical Interventions*, 3(1), pp. 123–129. doi: 10.1016/j.trci.2016.12.001.

Dubois, B. *et al.* (2007) 'Research criteria for the diagnosis of Alzheimer's disease: revising the NINCDS–ADRDA criteria', *The Lancet Neurology*, 6(8), pp. 734–746. doi: 10.1016/S1474-4422(07)70178-3.

Dubois, B. *et al.* (2010) 'Revising the definition of Alzheimer's disease: a new lexicon', *The Lancet Neurology*, 9(11), pp. 1118–1127. doi: 10.1016/S1474-4422(10)70223-4.

Dubois, B. *et al.* (2014) 'Advancing research diagnostic criteria for Alzheimer's disease: the IWG-2 criteria', *The Lancet Neurology*, 13(6), pp. 614–629. doi: 10.1016/S1474-4422(14)70090-0.

Dubois, B. *et al.* (2016) 'Preclinical Alzheimer's disease: Definition, natural history, and diagnostic criteria.', *Alzheimer's & dementia : the journal of the Alzheimer's Association*, 12(3), pp. 292–323. doi: 10.1016/j.jalz.2016.02.002.

Duke Han, S. *et al.* (2017) 'Detectable Neuropsychological Differences in Early Preclinical Alzheimer's Disease: A Meta-Analysis', *Neuropsychology Review*, pp. 1–21. doi: 10.1007/s11065-017-9345-5.

Duke Han, S. and Bondi, M. W. (2008) 'Revision of the apolipoprotein E compensatory mechanism recruitment hypothesis', *Alzheimer's & Dementia*, 4(4), pp. 251–254. doi: 10.1016/J.JALZ.2008.02.006.

Dykiert, D. *et al.* (2012a) 'Age Differences in Intra-Individual Variability in Simple and Choice Reaction Time: Systematic Review and Meta-Analysis', *PLoS ONE*, 7(10), p. e45759. doi: 10.1371/journal.pone.0045759.

Dykiert, D. *et al.* (2012b) 'Sex differences in reaction time mean and intraindividual variability across the life span.', *Developmental Psychology*, 48(5), pp. 1262–1276. doi: 10.1037/a0027550.

Ecuyer-Dab, I. and Robert, M. (2004) 'Have sex differences in spatial ability evolved from male competition for mating and female concern for survival?', *Cognition*, 91(3), pp. 221–257. doi: 10.1016/j.cognition.2003.09.007.

Edmonds, E. C. *et al.* (2015) 'Subtle Cognitive Decline and Biomarker Staging in Preclinical Alzheimer's Disease', *Journal of Alzheimer's Disease*, 47(1), pp. 231–242. doi: 10.3233/JAD-150128.

Emery, L., Hale, S. and Myerson, J. (2008) 'Age differences in proactive interference, working memory, and abstract reasoning.', *Psychology and aging*, 23(3), pp. 634–45. doi: 10.1037/a0012577.

Farrell, M. E. *et al.* (2017) 'Association of Longitudinal Cognitive Decline With Amyloid Burden in Middle-aged and Older Adults', *JAMA Neurology*, 74(7), p. 830. doi: 10.1001/jamaneurol.2017.0892.

Farrell, M. E. *et al.* (2018) 'Regional amyloid accumulation and cognitive decline in initially amyloid-negative adults.', *Neurology*, 91(19), pp. e1809–e1821. doi:

10.1212/WNL.0000000000006469.

Feeney, J. *et al.* (2013) 'Low macular pigment optical density is associated with lower cognitive performance in a large, population-based sample of older adults', *Neurobiology of Aging*, 34(11), pp. 2449–2456. doi: 10.1016/J.NEUROBIOLAGING.2013.05.007.

Fernández, G. *et al.* (2018) 'Visual Processing during Short-Term Memory Binding in Mild Alzheimer's Disease', *Journal of Alzheimer's Disease*, 63(1), pp. 185–194. doi: 10.3233/JAD-170728.

Fischer, A. G. *et al.* (2016) 'Gender Influences on Brain Responses to Errors and Post-Error Adjustments', *Scientific Reports*, 6(1), p. 24435. doi: 10.1038/srep24435.

Fjell, A. M. *et al.* (2011) 'Reduced white matter integrity is related to cognitive instability.', *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 31(49), pp. 18060–72. doi: 10.1523/JNEUROSCI.4735-11.2011.

Folstein, M. F., Folstein, S. E. and McHugh, P. R. (1975) '"Mini-mental state." A practical method for grading the cognitive state of patients for the clinician.', *Journal of Psychiatric Research*, 12(3), pp. 189–198. doi: 10.1016/0022-3956(75)90026-6.

Forster, S. and Lavie, N. (2008) 'Attentional capture by entirely irrelevant distractors', *Visual Cognition*, 16(2–3), pp. 200–214. doi: 10.1080/13506280701465049.

Fox, N. C. *et al.* (1998) 'Presymptomatic cognitive deficits in individuals at risk of familial Alzheimer's disease. A longitudinal prospective study.', *Brain*, 121, pp. 1631–1639. Available at: http://www.ncbi.nlm.nih.gov/pubmed/9762953 (Accessed: 15 October 2018).

Fox, N. C. and Schott, J. M. (2004) 'Imaging cerebral atrophy: normal ageing to Alzheimer's disease', *The Lancet*. Elsevier, 363(9406), pp. 392–394. doi: 10.1016/S0140-6736(04)15441-X.

Freeston, J. and Rooney, K. (2014) 'Throwing Speed and Accuracy in Baseball and Cricket Players', *Perceptual and Motor Skills*. SAGE PublicationsSage CA: Los Angeles, CA, 118(3), pp. 637–650. doi: 10.2466/30.PMS.118k25w4.

Gaertner, B. *et al.* (2018) 'Normative data for the Digit Symbol Substitution Test in a population-based sample aged 65–79 years: Results from the German Health Interview and Examination Survey for Adults (DEGS1)', *The Clinical Neuropsychologist*, 32(sup1), pp. 114–132. doi: 10.1080/13854046.2018.1484168.

Gale, C. R. *et al.* (2012) 'Cognitive function in childhood and lifetime cognitive change in relation to mental wellbeing in four cohorts of older people', *PLoS One*, 7(9), p. e44860. doi: 10.1371/journal.pone.0044860.

Galvin, J. E. *et al.* (2005) 'The AD8: A brief informant interview to detect dementia', *Neurology*, 65(4), pp. 559–564. doi: 10.1212/01.wnl.0000172958.95282.2a.

Garrett, D. D., Macdonald, S. W. S. and Craik, F. I. M. (2012) 'Intraindividual reaction time variability is malleable: feedback- and education-related reductions in variability with age.', *Frontiers in human neuroscience*, 6, p. 101. doi: 10.3389/fnhum.2012.00101.

Gaysina, D. *et al.* (2014) 'Cortisol and cognitive function in midlife: the role of childhood cognition and educational attainment.', *Psychoneuroendocrinology*, 47, pp. 189–98.

Girelli, L. and Delazer, M. (2001) 'Numerical abilities in dementia', *Aphasiology*, 15(7), pp. 681–694. doi: 10.1080/02687040143000122.

Gladyshev, T. V and Gladyshev, V. N. (2016) 'A Disease or Not a Disease? Aging As a Pathology.', *Trends in molecular medicine*, 22(12), pp. 995–996. doi:

10.1016/j.molmed.2016.09.009.

Glisky, E. L. (2007) ‘Changes in Cognitive Function in Human Aging’, in Riddle, D. (ed.) *Brain Aging: Models, Methods, and Mechanisms*. Boca Raton, FL: CRC Press/Taylor & Francis. Available at: http://www.ncbi.nlm.nih.gov/pubmed/21204355.

Godefroy, O. *et al.* (2010) ‘Age-related slowing: perceptuomotor, decision, or attention decline?’, *Experimental aging research*, 36(2), pp. 169–89. doi: 10.1080/03610731003613615.

Goldstein, S. and Naglieri, J. A. (eds) (2014) *Handbook of Executive Functioning*. New York: Springer. doi: 10.1007/978-1-4614-8106-5.

Grober, E. *et al.* (2008) ‘Memory impairment, executive dysfunction, and intellectual decline in preclinical Alzheimer’s disease.’, *Journal of the International Neuropsychological Society*, 14(2), pp. 266–78. doi: 10.1017/S1355617708080302.

Grothe, M. J. *et al.* (2017) ‘In vivo staging of regional amyloid deposition’, *Neurology*, 89(20). doi: 10.1212/WNL.0000000000004643.

Guiso, L. *et al.* (2008) ‘Culture, Gender and Math’, *Science*, 320, pp. 1164–1165.

Gunning-Dixon, F. M. and Raz, N. (2000) ‘The cognitive correlates of white matter abnormalities in normal aging: a quantitative review.’, *Neuropsychology*, 14(2), pp. 224–32.

Halpern, D. F. *et al.* (2007) ‘The Science of Sex Differences in Science and Mathematics.’, *Psychological science in the public interest : a journal of the American Psychological Society*, 8(1), pp. 1–51. doi: 10.1111/j.1529-1006.2007.00032.x.

Harrington, K. D., Lim, Y. Y., Ames, D., Hassenstab, J., Laws, S. M., *et al.* (2017) ‘Amyloid β–associated cognitive decline in the absence of clinical disease progression and systemic illness’, *Alzheimer’s & Dementia: Diagnosis, Assessment & Disease Monitoring*, 8, pp. 156–164. doi: 10.1016/J.DADM.2017.05.006.

Harrington, K. D., Lim, Y. Y., Ames, D., Hassenstab, J., Rainey-Smith, S., *et al.* (2017) ‘Using Robust Normative Data to Investigate the Neuropsychology of Cognitive Aging’, *Archives of Clinical Neuropsychology*, 32(2), pp. 142–154. doi: 10.1093/arclin/acw106.

Harrington, K. D. *et al.* (2018) ‘Estimates of age-related memory decline are inflated by unrecognized Alzheimer’s disease’, *Neurobiology of Aging*, 70, pp. 170–179. doi: 10.1016/J.NEUROBIOLAGING.2018.06.005.

Harrington, M. G. *et al.* (2013) ‘Executive function changes before memory in preclinical Alzheimer’s pathology: a prospective, cross-sectional, case control study.’, *PloS one*. Public Library of Science, 8(11), p. e79378. doi: 10.1371/journal.pone.0079378.

Hassenstab, Jason *et al.* (2015) ‘Absence of practice effects in preclinical Alzheimer’s disease.’, *Neuropsychology*, 29(6), pp. 940–8. doi: 10.1037/neu0000208.

Hassenstab, J *et al.* (2015) ‘Neuropsychological Markers of Cognitive Decline in Persons With Alzheimer Disease Neuropathology’, *J Neuropathol Exp Neurol*. 2015/10/16, 74(11), pp. 1086–1092. doi: 10.1097/nen.0000000000000254.

Hassenstab, J. *et al.* (2016) ‘Certified normal: Alzheimer’s disease biomarkers and normative estimates of cognitive functioning’, *Neurobiology of Aging*, 43, pp. 23–33. doi: 10.1016/j.neurobiolaging.2016.03.014.

Hassenstab, J. *et al.* (2017) ‘Reliability and reproducibility of cognitive composites for Alzheimer’s disease secondary prevention trials: The power PACC’. Vienna. Available at: https://cmoffice.kenes.com/ADPD17/CM.NET.WebUI/CM.NET.WEBUI.scpr/SCPRfunct

iondetail.aspx?confID=05000000-0000-0000-0000-000000000195&sesID=05000000-0000-0000-0000-000000037872&absID=07000000-0000-0000-0000-000000556608.

Hassenstab, J. *et al.* (2018) 'Comparing smartphone-administered cognitive assessments with conventional tests and biomarkers in sporadic and dominantly inherited Alzheimer disease', *Alzheimer's & Dementia*, 14(7), pp. P224–P225. doi: 10.1016/j.jalz.2018.06.2350.

Hatch, S. L. *et al.* (2007) 'The continuing benefits of education: adult education and midlife cognitive ability in the British 1946 birth cohort.', *The journals of gerontology. Series B, Psychological sciences and social sciences*, 62B(6), pp. S404-14. Available at: http://www.ncbi.nlm.nih.gov/pubmed/18079429.

Hawkins, K. M. and Sergio, L. E. (2014) 'Visuomotor impairments in older adults at increased Alzheimer's disease risk.', *Journal of Alzheimer's disease : JAD*, 42(2), pp. 607–21. Available at: http://www.ncbi.nlm.nih.gov/pubmed/24919768 (Accessed: 6 November 2015).

Haworth, J. *et al.* (2016) 'Measuring Information Processing Speed in Mild Cognitive Impairment: Clinical Versus Research Dichotomy.', *Journal of Alzheimer's disease : JAD*, 51(1), pp. 263–75. doi: 10.3233/JAD-150791.

Hedden, T. *et al.* (2013) 'Meta-analysis of amyloid-cognition relations in cognitively normal older adults.', *Neurology*, 80(14), pp. 1341–8. doi: 10.1212/WNL.0b013e31828ab35d.

Hedden, T. *et al.* (2016) 'Multiple Brain Markers are Linked to Age-Related Variation in Cognition', *Cerebral Cortex*, 26(4), pp. 1388–1400. doi: 10.1093/cercor/bhu238.

Heneka, M. T. *et al.* (2015) 'Neuroinflammation in Alzheimer's disease.', *The Lancet. Neurology*, 14(4), pp. 388–405. doi: 10.1016/S1474-4422(15)70016-5.

Hirono, N. *et al.* (1998) 'Regional metabolism: associations with dyscalculia in Alzheimer's disease.', *Journal of neurology, neurosurgery, and psychiatry*, 65(6), pp. 913–6. doi: 10.1136/JNNP.65.6.913.

Ho, J. K. and Nation, D. A. (2018) 'Neuropsychological Profiles and Trajectories in Preclinical Alzheimer's Disease', *Journal of the International Neuropsychological Society*, pp. 1–10. doi: 10.1017/S135561771800022X.

Hohman, T. J. *et al.* (2017) 'Evaluating Alzheimer's disease biomarkers as mediators of age-related cognitive decline.', *Neurobiology of aging*, 58, pp. 120–128. doi: 10.1016/j.neurobiolaging.2017.06.022.

Holden, M. P., Duff-Canning, S. J. and Hampson, E. (2015) 'Sex differences in the weighting of metric and categorical information in spatial location memory', *Psychological Research*, 79(1), pp. 1–18. doi: 10.1007/s00426-013-0539-z.

Hsieh, S. *et al.* (2013) 'Validation of the Addenbrooke's Cognitive Examination III in frontotemporal dementia and Alzheimer's disease.', *Dementia and geriatric cognitive disorders*, 36(3–4), pp. 242–50. doi: 10.1159/000351671.

Huang, A. R. *et al.* (2018) 'Adolescent Cognitive Aptitudes and Later-in-Life Alzheimer Disease and Related Disorders', *JAMA Network Open*, 1(5), p. e181726. doi: 10.1001/jamanetworkopen.2018.1726.

Hultsch, D. F. *et al.* (2000) 'Intraindividual variability in cognitive performance in older adults: comparison of adults with mild dementia, adults with arthritis, and healthy adults.', *Neuropsychology*, 14(4), pp. 588–598. doi: http://dx.doi.org/10.1037/0894-4105.14.4.588.

Hultsch, D. F. *et al.* (2002) 'Sampling and generalisability in developmental research:

Comparison of random and convenience samples of older adults', *International Journal of Behavioral Development*, 26(4), pp. 345–359. doi: 10.1080/01650250143000247.

Hurst, L. *et al.* (2013) 'Lifetime Socioeconomic Inequalities in Physical and Cognitive Aging', *American Journal of Public Health*, 103(9), pp. 1641–1648. doi: 10.2105/AJPH.2013.301240.

Hyde, J. S. and Mertz, J. E. (2009) 'Gender, culture, and mathematics performance.', *Proceedings of the National Academy of Sciences of the United States of America*, 106(22), pp. 8801–7. doi: 10.1073/pnas.0901265106.

Jack, C. R. *et al.* (2012) 'An operational approach to National Institute on Aging-Alzheimer's Association criteria for preclinical Alzheimer disease', *Annals of Neurology*, 71(6), pp. 765–775. doi: 10.1002/ana.22628.

Jack, C. R. *et al.* (2013) 'Tracking pathophysiological processes in Alzheimer's disease: an updated hypothetical model of dynamic biomarkers.', *The Lancet. Neurology*, 12(2), pp. 207–16. doi: 10.1016/S1474-4422(12)70291-0.

Jack, C. R. *et al.* (2014) 'Age-specific population frequencies of cerebral β-amyloidosis and neurodegeneration among people with normal cognitive function aged 50-89 years: a cross-sectional study.', *The Lancet. Neurology*, 13(10), pp. 997–1005. doi: 10.1016/S1474-4422(14)70194-2.

Jack, C. R. *et al.* (2016) 'A/T/N: An unbiased descriptive classification scheme for Alzheimer disease biomarkers.', *Neurology*, 87(5), pp. 539–47. doi: 10.1212/WNL.0000000000002923.

Jack, C. R. *et al.* (2017) 'Age-specific and sex-specific prevalence of cerebral β-amyloidosis, tauopathy, and neurodegeneration in cognitively unimpaired individuals aged 50–95 years: a cross-sectional study', *The Lancet Neurology*. doi: 10.1016/S1474-4422(17)30077-7.

Jack, C. R. *et al.* (2018) 'NIA-AA Research Framework: Toward a biological definition of Alzheimer's disease', *Alzheimer's & Dementia*, 14(4), pp. 535–562. doi: 10.1016/J.JALZ.2018.02.018.

Jackson, J. D. *et al.* (2012) 'White matter integrity and reaction time intraindividual variability in healthy aging and early-stage Alzheimer disease', *Neuropsychologia*, 50(3), pp. 357–366. doi: 10.1016/J.NEUROPSYCHOLOGIA.2011.11.024.

Jaeger, J. (2018) 'Digit Symbol Substitution Test: The case for sensitivity over specificity in neuropsychological testing', *Journal of Clinical Psychopharmacology*, 38(5), pp. 513–519. doi: 10.1097/JCP.0000000000000941.

James, S.-N. *et al.* (2018) 'Using a birth cohort to study brain health and preclinical dementia: Recruitment and participation rates in Insight 46', *BMC Research Notes*, 11(885). doi: https://doi.org/10.1186/s13104-018-3995-0.

Janac, S., Clarke, B. and Gems, D. (2017) 'Aging: natural or disease? The view from medical textbooks.', in Vaiserman, A. (ed.) *Anti-aging Drugs From Basic Research to Clinical Practice*. Cambridge: Royal Society of Chemistry, pp. 11–34.

Jansen, W. J. *et al.* (2015) 'Prevalence of cerebral amyloid pathology in persons without dementia: a meta-analysis.', *JAMA*, 313(19), pp. 1924–38. doi: 10.1001/jama.2015.4668.

Jansen, W. J. *et al.* (2018) 'Association of Cerebral Amyloid-β Aggregation With Cognitive Functioning in Persons Without Dementia', *JAMA Psychiatry*, 75(1), p. 84. doi: 10.1001/jamapsychiatry.2017.3391.

Jessen, F. *et al.* (2014) 'A conceptual framework for research on subjective cognitive

decline in preclinical Alzheimer's disease', *Alzheimer's & Dementia*, 10(6), pp. 844–852. doi: 10.1016/j.jalz.2014.01.001.

Jonaitis, E. M. *et al.* (2019) 'Measuring longitudinal cognition: Individual tests versus composites', *Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring*. Elsevier, 11, pp. 74–84. doi: 10.1016/J.DADM.2018.11.006.

Jonasson, Z. (2005) 'Meta-analysis of sex differences in rodent models of learning and memory: a review of behavioral and biological data.', *Neuroscience and biobehavioral reviews*, 28(8), pp. 811–25. doi: 10.1016/j.neubiorev.2004.10.006.

Jones, M. and Thompson, J. (2017) 'Atypical presentations of Alzheimer's disease', in Waldemar, G. and Burns, A. (eds) *Allzheimer's Disease*. 2nd edn. Oxford University Press. doi: 10.1093/med/9780198779803.003.0005.

Kaskikallio, A. *et al.* (2019) 'Cognitive Effects of White Matter Pathology in Normal and Pathological Aging', *Journal of Alzheimer's Disease*, 67(2), pp. 489–493. doi: 10.3233/JAD-180554.

Kern, S. *et al.* (2018) 'Prevalence of preclinical Alzheimer disease: Comparison of current classification systems.', *Neurology*, 90(19), pp. e1682-1691. doi: 10.1212/WNL.0000000000005476.

Kirova, A.-M. *et al.* (2015) 'Working Memory and Executive Function Decline across Normal Aging, Mild Cognitive Impairment, and Alzheimer's Disease', *BioMed Research International*, 2015, pp. 1–9. doi: 10.1155/2015/748212.

Kline, A. (2012) 'Apolipoprotein E, amyloid-ß clearance and therapeutic opportunities in Alzheimer's disease.', *Alzheimer's research & therapy*, 4(4), p. 32. doi: 10.1186/alzrt135.

Knopman, D. S. *et al.* (2018) 'Joint associations of β-amyloidosis and cortical thickness with cognition.', *Neurobiology of aging*, 65, pp. 121–131. doi: 10.1016/j.neurobiolaging.2018.01.017.

Koen, J. D. *et al.* (2017) 'Visual short-term memory for high resolution associations is impaired in patients with medial temporal lobe damage', *Hippocampus*, 27(2), pp. 184–193. doi: 10.1002/hipo.22682.

Koike, S. *et al.* (2017) 'Cognitive profiles in childhood and adolescence differ between adult psychotic and affective symptoms: a prospective birth cohort study', *Psychological Medicine*, pp. 1–12. doi: 10.1017/S0033291717000393.

Kormas, C. *et al.* (2018) 'Development of the Greek version of the Face Name Associative Memory Exam (GR-FNAME12) in cognitively normal elderly individuals', *The Clinical Neuropsychologist*, 32, pp. 152–163. doi: 10.1080/13854046.2018.1495270.

Kozauer, N. and Katz, R. (2013) 'Regulatory Innovation and Drug Development for Early-Stage Alzheimer's Disease', *New England Journal of Medicine*, 368(13), pp. 1169–1171. doi: 10.1056/NEJMp1302513.

Kuh, D. *et al.* (2009) 'Lifetime cognitive performance is associated with midlife physical performance in a prospective national birth cohort study.', *Psychosomatic medicine*, 71(1), pp. 38–48. doi: 10.1097/PSY.0b013e31818a1620.

Kuh, D. *et al.* (2016) 'The MRC National Survey of Health and Development reaches age 70: maintaining participation at older ages in a birth cohort study', *European Journal of Epidemiology*. Springer Netherlands, 31(11), pp. 1135–1147. doi: 10.1007/s10654-016-0217-8.

Kunz, L. *et al.* (2015) 'Reduced grid-cell-like representations in adults at genetic risk for

Alzheimer's disease.', *Science*, 350(6259), pp. 430–3. doi: 10.1126/science.aac8128.

Landau, S. M. *et al.* (2018) 'Memory decline accompanies subthreshold amyloid accumulation.', *Neurology*, 90(17), pp. e1452–e1460. doi: 10.1212/WNL.0000000000005354.

Lane, C. A. *et al.* (2017) 'Study protocol: Insight 46 – a neuroscience sub-study of the MRC National Survey of Health and Development', *BMC Neurology*, 17(75). doi: 10.1186/s12883-017-0846-x.

Langbaum, J. B. *et al.* (2014) 'An empirically derived composite cognitive test score with improved power to track and evaluate treatments for preclinical Alzheimer's disease', *Alzheimer's & Dementia*, 10(6), pp. 666–674. doi: 10.1016/j.jalz.2014.02.002.

Leal, S. L. *et al.* (2018) 'Subthreshold Amyloid Predicts Tau Deposition in Aging.', *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 38(19), pp. 4482–4489. doi: 10.1523/JNEUROSCI.0485-18.2018.

Lemay, M. *et al.* (2005) 'Sensorimotor mapping affects movement correction deficits in early Huntington's disease', *Experimental Brain Research*, 165(4), pp. 454–460. doi: 10.1007/s00221-005-2315-9.

Leung, K. K. *et al.* (2011) 'Brain MAPS: an automated, accurate and robust brain extraction technique using a template library.', *NeuroImage*, 55(3), pp. 1091–108. doi: 10.1016/j.neuroimage.2010.12.067.

Liang, Y. *et al.* (2016) 'Visual short-term memory binding deficit in familial Alzheimer's disease', *Cortex*, 78, pp. 150–164. doi: 10.1016/j.cortex.2016.01.015.

Liang, Y. *et al.* (2017) 'Visual short-term memory binding deficits in Alzheimer's disease: a reply to Parra's commentary.', *Cortex*, 88, pp. 201–204. doi: 10.1016/J.CORTEX.2016.11.003.

Lim, Y. Y. *et al.* (2015) 'APOE ε4 moderates amyloid-related memory decline in preclinical Alzheimer's disease', *Neurobiology of Aging*, 36(3), pp. 1239–1244. doi: 10.1016/j.neurobiolaging.2014.12.008.

Lim, Y. Y., Laws, S. M., *et al.* (2016) 'Aβ-related memory decline in APOE ε4 noncarriers: Implications for Alzheimer disease.', *Neurology*, 86(17), pp. 1635–42. doi: 10.1212/WNL.0000000000002604.

Lim, Y. Y., Snyder, P. J., *et al.* (2016) 'Sensitivity of composite scores to amyloid burden in preclinical Alzheimer's disease: Introducing the Z-scores of Attention, Verbal fluency, and Episodic memory for Nondemented older adults composite score.', *Alzheimer's & dementia : Diagnosis, Assessment & Disease Monitoring*, 2, pp. 19–26. doi: 10.1016/j.dadm.2015.11.003.

Lim, Y. Y. *et al.* (2018) 'Association of β-Amyloid and Apolipoprotein E ε4 With Memory Decline in Preclinical Alzheimer Disease', *JAMA Neurology*, 75(4), p. 488. doi: 10.1001/jamaneurol.2017.4325.

Lock, M. (2013) *The Alzheimer Conundrum: Entanglements of Dementia and Aging*. Princeton: Princeton University Press. Available at: https://www.jstor.org/stable/j.ctt3fgz45 (Accessed: 26 June 2018).

Logie, R. H. *et al.* (2007) 'Dual Task Demands On Encoding And Retrieval Processes: Evidence From Healthy Adult Ageing', *Cortex*, 43(1), pp. 159–169. doi: 10.1016/S0010-9452(08)70453-2.

Louie, R. (2019) 'The 2018 NIA-AA research framework: Recommendation and comments', *Alzheimer's & Dementia*, 15(1), pp. 182–183. doi: 10.1016/J.JALZ.2018.06.3062.

Lu, K. *et al.* (2019) 'Cognition at age 70: life course predictors and associations with brain pathologies', *Neurology*, Accepted.

Ma, W. J., Husain, M. and Bays, P. M. (2014) 'Changing concepts of working memory', *Nature Neuroscience*, 17(3), pp. 347–356. doi: 10.1038/nn.3655.

MacDonald, S. W. S., Nyberg, L. and Bäckman, L. (2006) 'Intra-individual variability in behavior: links to brain structure, neurotransmission and neuronal activity', *Trends in Neurosciences*, 29(8), pp. 474–480. doi: 10.1016/J.TINS.2006.06.011.

Machulda, M. M. *et al.* (2017) 'Practice effects and longitudinal cognitive change in clinically normal older adults differ by Alzheimer imaging biomarker status', *The Clinical Neuropsychologist*, 31(1), pp. 99–117. doi: 10.1080/13854046.2016.1241303.

MacLeod, C. M. (1991) 'Half a century of research on the Stroop effect: an integrative review.', *Psychological bulletin*, 109(2), pp. 163–203. doi: http://dx.doi.org/10.1037/0033-2909.109.2.163.

Macpherson, K. *et al.* (2017) 'Visuomotor Integration in presymptomatic familial Alzheimer's disease', *Alzheimer's & Dementia*, 13(7), p. P815. doi: 10.1016/j.jalz.2017.06.1115.

MacPherson, S. E. *et al.* (2012) 'Dual Task Abilities as a Possible Preclinical Marker of Alzheimer's Disease in Carriers of the E280A Presenilin-1 Mutation', *Journal of the International Neuropsychological Society*, 18, pp. 234–241.

Magistro, D. *et al.* (2015) 'The Relationship between Processing Speed and Regional White Matter Volume in Healthy Young People', *PLoS ONE*, 10(9), p. e0136386. doi: 10.1371/JOURNAL.PONE.0136386.

Majeres, R. L. (1983) 'Sex differences in symbol-digit substitution and speeded matching', *Intelligence*. JAI, 7(4), pp. 313–327. doi: 10.1016/0160-2896(83)90007-7.

Malone, I. B. *et al.* (2015) 'Accurate automatic estimation of total intracranial volume: a nuisance variable with less nuisance.', *NeuroImage*, 104, pp. 366–372. doi: 10.1016/j.neuroimage.2014.09.034.

Martin, R. C. *et al.* (2003) 'Loss of Calculation Abilities in Patients With Mild and Moderate Alzheimer Disease', *Archives of Neurology*, 60(11), pp. 1585–1589. doi: 10.1001/archneur.60.11.1585.

Masi, S. *et al.* (2018) 'Patterns of adiposity, vascular phenotypes and cognitive function in the 1946 British Birth Cohort.', *BMC medicine*, 16(1), p. 75. doi: 10.1186/s12916-018-1059-x.

Mayes, A., Montaldi, D. and Migo, E. (2007) 'Associative memory and the medial temporal lobes.', *Trends in cognitive sciences*, 11(3), pp. 126–35. doi: 10.1016/j.tics.2006.12.003.

McCleery, J. *et al.* (2018) 'When is Alzheimer's not dementia—Cochrane commentary on The National Institute on Ageing and Alzheimer's Association Research Framework for Alzheimer's Disease', *Age and Ageing*. doi: 10.1093/ageing/afy167.

McCleery, J. *et al.* (2019) 'The National Institute on Aging and Alzheimer's Association research framework: A commentary from the Cochrane Dementia and Cognitive Improvement Group', *Alzheimer's & Dementia*, 15(1), pp. 179–181. doi: 10.1016/J.JALZ.2018.10.007.

McColgan, P. *et al.* (2017) 'Topological length of white matter connections predicts their rate of atrophy in premanifest Huntington's disease.', *JCI insight*, 2(8). doi: 10.1172/jci.insight.92641.

McDaniel, M. A. (2005) 'Big-brained people are smarter: A meta-analysis of the

relationship between in vivo brain volume and intelligence', *Intelligence*, 33, pp. 337–346. doi: 10.1016/j.intell.2004.11.005.

McGurn, B. *et al.* (2008) 'Childhood cognitive ability and risk of late-onset Alzheimer and vascular dementia.', *Neurology*, 71(14), pp. 1051–6. doi: 10.1212/01.wnl.0000319692.20283.10.

McMillan, C. T. and Chételat, G. (2018) 'Amyloid "accumulators": The next generation of candidates for amyloid-targeted clinical trials?', *Neurology*, 90(17), pp. 759–760. doi: 10.1212/WNL.0000000000005362.

Medeiros, A. de M. and Silva, R. H. (2019) 'Sex Differences in Alzheimer's Disease: Where Do We Stand?', *Journal of Alzheimer's Disease*, 67(1), pp. 35–60. doi: 10.3233/JAD-180213.

Medina, M. *et al.* (2017) 'Toward common mechanisms for risk factors in Alzheimer's syndrome.', *Alzheimer's & Dementia: Translational Research & Clinical Interventions*, pp. 571–578. doi: 10.1016/j.trci.2017.08.009.

Merluzzi, A. P. *et al.* (2019) 'Differential effects of neurodegeneration biomarkers on subclinical cognitive decline.', *Alzheimer's & Dementia: Translational Research & Clinical Interventions*. Elsevier, pp. 129–138. doi: 10.1016/j.trci.2019.02.004.

Mollica, M. A. *et al.* (2017) 'Subtle visuomotor difficulties in preclinical Alzheimer's disease.', *Journal of neuropsychology*, 11(1), pp. 56–73. doi: 10.1111/jnp.12079.

Montgomery, V. *et al.* (2017) 'Effects of Delay Duration on the WMS Logical Memory Performance of Older Adults with Probable Alzheimer's Disease, Probable Vascular Dementia, and Normal Cognition', *Archives of Clinical Neuropsychology*, 32(3), pp. 375–380. doi: 10.1093/arclin/acx005.

Mormino, E. C. *et al.* (2014) 'Amyloid and APOE ε4 interact to influence short-term decline in preclinical Alzheimer disease.', *Neurology*, 82(20), pp. 1760–7. doi: 10.1212/WNL.0000000000000431.

Mormino, E. C. *et al.* (2017) 'Early and late change on the preclinical Alzheimer's cognitive composite in clinically normal older individuals with elevated β-amyloid', *Alzheimer's & Dementia*, 13, pp. 1004–1012. doi: 10.1016/j.jalz.2017.01.018.

Morris, G. P., Clark, I. A. and Vissel, B. (2018) 'Questions concerning the role of amyloid-β in the definition, aetiology and diagnosis of Alzheimer's disease', *Acta Neuropathologica*, 136, pp. 663–689. doi: 10.1007/s00401-018-1918-8.

Mortamais, M. *et al.* (2017) 'Detecting cognitive changes in preclinical Alzheimer's disease: A review of its feasibility', *Alzheimer's & Dementia*, 13(4), pp. 468–492. doi: 10.1016/j.jalz.2016.06.2365.

Mura, T. *et al.* (2017) 'Predicting dementia using socio-demographic characteristics and the Free and Cued Selective Reminding Test in the general population.', *Alzheimer's research & therapy*. BioMed Central, 9(1), p. 21. doi: 10.1186/s13195-016-0230-x.

Murphy, D. (2015) *Concepts of Health and Disease*, *The Stanford Encyclopedia of Philosophy*. Edited by E. N. Zalta. Available at: https://plato.stanford.edu/archives/spr2015/entries/health-disease.

Nagengast, A. J., Braun, D. A. and Wolpert, D. M. (2011) 'Risk sensitivity in a motor task with speed-accuracy trade-off.', *Journal of neurophysiology*, 105(6), pp. 2668–74. doi: 10.1152/jn.00804.2010.

National Institute on Aging (2015) *Alzheimer's Disease Genetics Fact Sheet*. Available at: https://www.nia.nih.gov/health/alzheimers-disease-genetics-fact-sheet#testing.

Neu, S. C. *et al.* (2017) 'Apolipoprotein E Genotype and Sex Risk Factors for Alzheimer Disease', *JAMA Neurology*, 74(10), p. 1178. doi: 10.1001/jamaneurol.2017.2188.

Nissan, J., Liewald, D. and Deary, I. J. (2013) 'Reaction time and intelligence: Comparing associations based on two response modes', *Intelligence*, 41(5), pp. 622–630. doi: 10.1016/J.INTELL.2013.08.002.

Nordenfelt, L. (2007) 'The concepts of health and illness revisited', *Medicine, Health Care and Philosophy*, 10(1), pp. 5–10. doi: 10.1007/s11019-006-9017-3.

Norton, S. *et al.* (2014) 'Potential for primary prevention of Alzheimer's disease: an analysis of population-based data', *The Lancet Neurology*, 13(8), pp. 788–794. doi: 10.1016/S1474-4422(14)70136-X.

O'Callaghan, C., Hodges, J. R. and Hornberger, M. (2013) 'Inhibitory dysfunction in frontotemporal dementia: a review.', *Alzheimer disease and associated disorders*, 27(2), pp. 102–8. doi: 10.1097/WAD.0b013e318265bbc9.

O'Donoghue, M. C. *et al.* (2018) 'APOE genotype and cognition in healthy individuals at risk of Alzheimer's disease: A review', *Cortex*, 104, pp. 103–123. doi: 10.1016/J.CORTEX.2018.03.025.

Oosterman, J. M. *et al.* (2004) 'Timed executive functions and white matter in aging with and without cardiovascular risk factors.', *Reviews in the neurosciences*, 15(6), pp. 439–62.

Otero, T. M. and Barker, L. A. (2014) 'The Frontal Lobes and Executive Functioning', in *Handbook of Executive Functioning*. New York, NY, pp. 29–44. doi: 10.1007/978-1-4614-8106-5_3.

Palmqvist, S. *et al.* (2018) 'CSF, plasma and MRI biomarker trajectories during the development of Alzheimer's disease', *Alzheimer's & Dementia*. Elsevier, 14(7), p. P641. doi: 10.1016/J.JALZ.2018.06.2688 (figure available at https://www.alzforum.org/news/conference-coverage/memory-slips-soon-amyloid-appears-two-decades-dementia).

Papp, K. V. *et al.* (2017) 'Optimizing the preclinical Alzheimer's cognitive composite with semantic processing: The PACC5', *Alzheimer's & Dementia: Translational Research & Clinical Interventions*, 3(4), pp. 668–677. doi: 10.1016/j.trci.2017.10.004.

Papp, K. V *et al.* (2014) 'Development of a psychometrically equivalent short form of the Face-Name Associative Memory Exam for use along the early Alzheimer's disease trajectory', *Clin Neuropsychol*. 2014/05/13, 28(5), pp. 771–785. doi: 10.1080/13854046.2014.911351.

Parnetti, L. *et al.* (2019) 'Prevalence and risk of progression of preclinical Alzheimer's disease stages: a systematic review and meta-analysis', *Alzheimer's Research & Therapy*, 11(1), p. 7. doi: 10.1186/s13195-018-0459-7.

Parra, M. A. *et al.* (2009) 'Short-term memory binding deficits in Alzheimer's disease.', *Brain : a journal of neurology*, 132(Pt 4), pp. 1057–66. doi: 10.1093/brain/awp036.

Parra, M. A. *et al.* (2010) 'Visual short-term memory binding deficits in familial Alzheimer's disease.', *Brain*, 133(9), pp. 2702–13. doi: 10.1093/brain/awq148.

Parra, M. A. *et al.* (2011) 'Specific deficit of colour–colour short-term memory binding in sporadic and familial Alzheimer's disease', *Neuropsychologia*, 49(7), pp. 1943–1952. doi: 10.1016/J.NEUROPSYCHOLOGIA.2011.03.022.

Parra, M. A. *et al.* (2015) 'Memory binding and white matter integrity in familial Alzheimer's disease', *Brain*, 138(5), pp. 1355–1369. doi: 10.1093/brain/awv048.

Parra, M. A. (2017) 'A commentary on Liang et al.'s paper with regard to emerging

views of memory assessment in Alzheimer's disease', *Cortex*, 88, pp. 198–200. doi: 10.1016/J.CORTEX.2016.06.006.

Parra, M. A. *et al.* (2017) 'Brain Information Sharing During Visual Short-Term Memory Binding Yields a Memory Biomarker for Familial Alzheimer's Disease', *Current Alzheimer Research*, 14(12). doi: 10.2174/1567205014666170614163316.

Payton, N. M. *et al.* (2018) 'Combining Cognitive, Genetic, and Structural Neuroimaging Markers to Identify Individuals with Increased Dementia Risk', *Journal of Alzheimer's Disease*, 64, pp. 533–542. doi: 10.3233/JAD-180199.

Pearson, H. (2016) *The Life Project: The Extraordinary Story of Our Ordinary Lives*. London, UK: Allen Lane.

Pereira, M. L. G. de F. *et al.* (2014) 'Eye movement analysis and cognitive processing: detecting indicators of conversion to Alzheimer's disease', *Neuropsychiatric Disease and Treatment*, 10, pp. 1273–1285. doi: 10.2147/NDT.S55371.

Pertzov, Y. *et al.* (2012) 'Forgetting what was where: the fragility of object-location binding.', *PloS one*, 7(10), p. e48214. doi: 10.1371/journal.pone.0048214.

Pertzov, Y. *et al.* (2013) 'Binding deficits in memory following medial temporal lobe damage in patients with voltage-gated potassium channel complex antibody-associated limbic encephalitis.', *Brain*, 136, pp. 2474–2485. doi: 10.1093/brain/awt129.

Pertzov, Y. *et al.* (2015) 'Effects of healthy ageing on precision and binding of object location in visual short term memory.', *Psychology and aging*, 30(1), pp. 26–35. doi: 10.1037/a0038396.

Petersen, R. C. *et al.* (2013) 'Mild cognitive impairment due to Alzheimer disease in the community.', *Annals of neurology*. NIH Public Access, 74(2), pp. 199–208. doi: 10.1002/ana.23931.

Pettigrew, C. *et al.* (2016) 'Cognitive reserve and cortical thickness in preclinical Alzheimer's disease', *Brain Imaging and Behavior*. Springer US, pp. 1–11. doi: 10.1007/s11682-016-9581-y.

Philippou, E. *et al.* (2018) 'Dietary glycaemic index and cognitive function: prospective associations in adults of the 1946 British birth cohort', *Public Health Nutrition*, pp. 1–10. doi: 10.1017/S136898001800352X.

Phillips, M. *et al.* (2013) 'Intra-Individual Reaction Time Variability in Mild Cognitive Impairment and Alzheimer's Disease: Gender, Processing Load and Speed Factors', *PLoS ONE*, p. e65712. doi: 10.1371/journal.pone.0065712.

Pierce, M. B. *et al.* (2012) 'Clinical disorders in a post war British cohort reaching retirement: evidence from the First National Birth Cohort study.', *PloS one*, 7(9), p. e44857. doi: 10.1371/journal.pone.0044857.

Pietto, M. *et al.* (2016) 'Behavioral and Electrophysiological Correlates of Memory Binding Deficits in Patients at Different Risk Levels for Alzheimer's Disease.', *Journal of Alzheimer's disease*, 53(4), pp. 1325–1340. doi: 10.3233/JAD-160056.

Pigeon, D. A. (1964) 'Tests used in the 1954 and 1957 surveys.', in Douglas, J. B. W. (ed.) *The home and the school.* London: Macgibbon & Key.

Pletnikova, O. *et al.* (2018) 'The spectrum of preclinical Alzheimer's disease pathology and its modulation by ApoE genotype', *Neurobiology of Aging*, 71, pp. 72–80. doi: 10.1016/j.neurobiolaging.2018.07.007.

Podcasy, J. L. and Epperson, C. N. (2016) 'Considering sex and gender in Alzheimer disease and other dementias.', *Dialogues in clinical neuroscience*, 18(4), pp. 437–446. Available at: http://www.ncbi.nlm.nih.gov/pubmed/28179815.

Primativo, S. *et al.* (2017) 'Eyetracking metrics reveal impaired spatial anticipation in behavioural variant frontotemporal dementia', *Neuropsychologia*, 106, pp. 328–340. doi: 10.1016/J.NEUROPSYCHOLOGIA.2017.10.014.

Prince, M. *et al.* (2014) *Dementia UK: Update. Second edition*. Alzheimer's Society.

Prins, N. D. *et al.* (2005) 'Cerebral small-vessel disease and decline in information processing speed, executive function and memory.', *Brain : a journal of neurology*, 128(Pt 9), pp. 2034–41. doi: 10.1093/brain/awh553.

Prins, N. D. and Scheltens, P. (2015) 'White matter hyperintensities, cognitive impairment and dementia: an update', *Nature Reviews Neurology*, 11(3), pp. 157–165. doi: 10.1038/nrneurol.2015.10.

Puccioni, O. and Vallesi, A. (2012) 'Conflict resolution and adaptation in normal aging: The role of verbal intelligence and cognitive reserve.', *Psychology and Aging*, 27(4), pp. 1018–1026. doi: 10.1037/a0029106.

Rabin, J. S. *et al.* (2018) 'Interactive Associations of Vascular Risk and β-Amyloid Burden With Cognitive Decline in Clinically Normal Elderly Individuals', *JAMA Neurology*, 75(9), pp. 1124–1131. doi: 10.1001/jamaneurol.2018.1123.

Rawle, M. J. *et al.* (2018) 'Apolipoprotein-E (Apoe) ε4 and cognitive decline over the adult life course', *Translational Psychiatry*, 8(18). doi: 10.1038/s41398-017-0064-8.

Reiman, E. M. *et al.* (2011) 'Alzheimer's Prevention Initiative: a plan to accelerate the evaluation of presymptomatic treatments.', *Journal of Alzheimer's disease*, 26 Suppl 3, pp. 321–9. doi: 10.3233/JAD-2011-0059.

Reimers, S. and Maylor, E. (2006) 'Gender Effects on Reaction Time Variability and Trial-to-Trial Performance: Reply to Deary and Der (2005)', *Aging, Neuropsychology, and Cognition*, 13(3–4), pp. 479–489. doi: 10.1080/138255890969375.

Rentz, D. M. *et al.* (2004) 'Use of IQ-adjusted norms to predict progressive cognitive decline in highly intelligent older individuals.', *Neuropsychology*, 18(1), pp. 38–49. doi: 10.1037/0894-4105.18.1.38.

Rentz, D. M. *et al.* (2007) 'Intelligence quotient-adjusted memory impairment is associated with abnormal single photon emission computed tomography perfusion.', *Journal of the International Neuropsychological Society*, 13(5), pp. 821–31. doi: 10.1017/S1355617707071056.

Rentz, D. M. *et al.* (2011) 'Face-name associative memory performance is related to amyloid burden in normal elderly', *Neuropsychologia*. 2011/06/22, 49(9), pp. 2776–2783. doi: 10.1016/j.neuropsychologia.2011.06.006.

Rentz, D. M. *et al.* (2013) 'Promising developments in neuropsychological approaches for the detection of preclinical Alzheimer's disease: a selective review', *Alzheimer's Research & Therapy*, 5(6), p. 58. doi: 10.1186/alzrt222.

Rentz, D. M. *et al.* (2017) 'Sex differences in episodic memory in early midlife: impact of reproductive aging', *Menopause*, 24(4), pp. 1–9. doi: 10.1097/GME.0000000000000771.

Rey-Mermet, A. and Gade, M. (2018) 'Inhibition in aging: What is preserved? What declines? A meta-analysis', *Psychonomic Bulletin & Review*, 25(5), pp. 1695–1716. doi: 10.3758/s13423-017-1384-7.

Richards, M. *et al.* (2004) 'Cognitive ability in childhood and cognitive decline in mid-life: longitudinal birth cohort study.', *BMJ*, 328(7439), p. 552. doi: 10.1136/bmj.37972.513819.EE.

Richards, M. *et al.* (2005) 'Lung Function and Cognitive Ability in a Longitudinal Birth

Cohort Study', *Psychosomatic Medicine*, 67(4), pp. 602–608. doi: 10.1097/01.psy.0000170337.51848.68.

Richards, M. *et al.* (2014) 'Lifetime affect and midlife cognitive function: prospective birth cohort study', *The British Journal of Psychiatry*, 204(3), pp. 194–199. doi: 10.1192/bjp.bp.113.128942.

Richards, M. *et al.* (2019) 'Identifying the lifetime cognitive and socioeconomic antecedents of cognitive state: seven decades of follow-up in a British birth cohort study', *BMJ Open*, 9(4), p. e024404. doi: 10.1136/bmjopen-2018-024404.

Richards, M. and Deary, I. J. (2013) 'A life course approach to cognitive capability', in Kuh, D. et al. (eds) *A Life Course Approach to Healthy Ageing*. Oxford, UK: Oxford University Press. doi: 10.1093/acprof:oso/9780199656516.001.0001.

Richards, M., Hardy, R. and Wadsworth, M. (2005) 'Alcohol consumption and midlife cognitive change in the British 1946 birth cohort study', *Alcohol and Alcoholism*, 40(2), pp. 112–117. doi: 10.1093/alcalc/agh126.

Richards, M. and Sacker, A. (2003) 'Lifetime Antecedents of Cognitive Reserve', *Journal of Clinical and Experimental Neuropsychology*, 25(5), pp. 614–624. doi: 10.1076/jcen.25.5.614.14581.

Richards, M. and Sacker, A. (2011) 'Is education causal? Yes', *International Journal of Epidemiology*, 40(2), pp. 516–518. doi: 10.1093/ije/dyq166.

Richards, M., Stephen, A. and Mishra, G. (2010) 'Health returns to cognitive capital in the British 1946 birth cohort', *Longitudinal and Life Course Studies*, 1(3), pp. 281–296. doi: http://dx.doi.org/10.14301/llcs.v1i3.94.

Ritchie, K. *et al.* (2017) 'Recommended cognitive outcomes in preclinical Alzheimer's disease: Consensus statement from the European Prevention of Alzheimer's Dementia project', *Alzheimer's & Dementia*, 13(2), pp. 186–195. doi: 10.1016/j.jalz.2016.07.154.

Ritchie, S. J. *et al.* (2016) 'Predictors of ageing-related decline across multiple cognitive functions.', *Intelligence*, 59, pp. 115–126. doi: 10.1016/j.intell.2016.08.007.

Ritchie, S. J. and Tucker-Drob, E. M. (2018) 'How Much Does Education Improve Intelligence? A Meta-Analysis', *Psychological Science*, 29(8), pp. 1358–1369. doi: 10.1177/0956797618774253.

Roe, C. *et al.* (2018) 'Using the A/T/N Framework to Examine Driving in Preclinical Alzheimer's Disease', *Geriatrics*, 3(2), p. 23. doi: 10.3390/geriatrics3020023.

Roe, C. M. *et al.* (2017) 'Preclinical Alzheimer's disease and longitudinal driving decline.', *Alzheimer's & Dementia: Translational Research & Clinical Interventions*, 3, pp. 74–82. doi: 10.1016/j.trci.2016.11.006.

Roivainen, E. (2011) 'Gender differences in processing speed: A review of recent research', *Learning and Individual Differences*. JAI, 21(2), pp. 145–149. doi: 10.1016/J.LINDIF.2010.11.021.

Ross, J. M. and Simpson, H. R. (1971a) 'The National Survey of Health and Development: 1. Educational Attainment', *British Journal of Educational Psychology*, 41(1), pp. 49–61. doi: 10.1111/j.2044-8279.1971.tb00657.x.

Ross, J. M. and Simpson, H. R. (1971b) 'The National Survey of Health and Development: 2. Rate of school progress between 8 and 15 years and between 15 and 18 years', *British Journal of Educational Psychology*, 41(2), pp. 125–135. doi: 10.1111/j.2044-8279.1971.tb02244.x.

Rothman, K. J. (1990) 'No adjustments are needed for multiple comparisons.', *Epidemiology*, 1(1), pp. 43–46. Available at:

http://www.ncbi.nlm.nih.gov/pubmed/2081237.

Rowe, C. C. *et al.* (2007) 'Imaging beta-amyloid burden in aging and dementia', *Neurology*, 68(20), pp. 1718–25. doi: 10.1212/01.wnl.0000261919.22630.ea.

Royer, F. L. (1978) 'Sex differences in symbol-digit substitution task performance', *Intelligence*, 2(2), pp. 145–151. doi: 10.1016/0160-2896(78)90005-3.

Rozand, V. *et al.* (2015) 'Effect of mental fatigue on speed–accuracy trade-off', *Neuroscience*, 297, pp. 219–230. doi: 10.1016/J.NEUROSCIENCE.2015.03.066.

Russ, T. C. *et al.* (2017) 'Childhood Cognitive Ability and Incident Dementia: The 1932 Scottish Mental Survey Cohort into their 10th Decade.', *Epidemiology*, 28(3), pp. 361–364. doi: 10.1097/EDE.0000000000000626.

Russ, T. C. (2018) 'Intelligence, Cognitive Reserve, and Dementia', *JAMA Network Open*, 1(5), p. e181724. doi: 10.1001/jamanetworkopen.2018.1724.

Rusted, J. M. *et al.* (2013) 'APOE e4 polymorphism in young adults is associated with improved attention and indexed by distinct neural signatures', *NeuroImage*, 65, pp. 364–373. doi: 10.1016/j.neuroimage.2012.10.010.

Ryan, N. S., Rossor, M. N. and Fox, N. C. (2015) 'Alzheimer's disease in the 100 years since Alzheimer's death', *Brain*, 138(12), pp. 3816–3821. doi: 10.1093/brain/awv316.

Salthouse, T. A. (1996) 'The processing-speed theory of adult age differences in cognition.', *Psychological review*, 103(3), pp. 403–428. doi: http://dx.doi.org/10.1037/0033-295X.103.3.403.

Salthouse, T. A. (2000) 'Aging and measures of processing speed', *Biological Psychology*, 54, pp. 35–54. doi: 10.1016/S0301-0511(00)00052-1.

Sanabria, A. *et al.* (2018) 'The Spanish version of Face-Name Associative Memory Exam (S-FNAME) performance is related to amyloid burden in Subjective Cognitive Decline', *Scientific Reports*, 8(3828). doi: 10.1038/s41598-018-21644-y.

Saucier, D. M. *et al.* (2002) 'Are sex differences in navigation caused by sexually dimorphic strategies or by differences in the ability to use the strategies?', *Behavioral neuroscience*, 116(3), pp. 403–410. doi: 10.1037//0735-7044.116.3.403.

Say, M. J. *et al.* (2011) 'Visuomotor integration deficits precede clinical onset in Huntington's disease.', *Neuropsychologia*, 49(2), pp. 264–70. doi: 10.1016/j.neuropsychologia.2010.11.016.

Scheller, E. *et al.* (2017) 'APOE moderates compensatory recruitment of neuronal resources during working memory processing in healthy older adults', *Neurobiology of Aging*, 56, pp. 127–137. doi: 10.1016/J.NEUROBIOLAGING.2017.04.015.

Schneegans, S. and Bays, P. M. (2019) 'New perspectives on binding in visual working memory.', *British journal of psychology*, 110, pp. 207–244. doi: 10.1111/bjop.12345.

Selkoe, D. J. and Hardy, J. (2016) 'The amyloid hypothesis of Alzheimer's disease at 25 years.', *EMBO molecular medicine*, 8(6), pp. 595–608. doi: 10.15252/emmm.201606210.

Sheppard, L. D. and Vernon, P. A. (2008) 'Intelligence and speed of information-processing: A review of 50 years of research', *Personality and Individual Differences*, 44(3), pp. 535–551. doi: 10.1016/J.PAID.2007.09.015.

Shirk, S. D. *et al.* (2011) 'A web-based normative calculator for the uniform data set (UDS) neuropsychological test battery', *Alzheimer's Research & Therapy*, 3(6), p. 32. doi: 10.1186/alzrt94.

Silverberg, N. B. *et al.* (2011) 'Assessment of cognition in early dementia.', *Alzheimer's & dementia : the journal of the Alzheimer's Association*, 7(3), pp. e60–e76. doi: 10.1016/j.jalz.2011.05.001.

Silverwood, R. J. *et al.* (2014) 'Cognitive and kidney function: results from a British birth cohort reaching retirement age', *PLoS One*, 9(1), p. e86743. doi: 10.1371/journal.pone.0086743.

Simon, J. R. (1969) 'Reactions toward the source of stimulation.', *Journal of experimental psychology*, 81(1), pp. 174–176. Available at: http://www.ncbi.nlm.nih.gov/pubmed/5812172 (Accessed: 20 October 2017).

Sliwinski, M. J. *et al.* (2018) 'Reliability and Validity of Ambulatory Cognitive Assessments', *Assessment*, 25(1), pp. 14–30. doi: 10.1177/1073191116643164.

Slot, R. E. R. *et al.* (2018) 'Subjective Cognitive Impairment Cohort (SCIENCe): study design and first results', *Alzheimer's Research & Therapy*, 10(1), p. 76. doi: 10.1186/s13195-018-0390-y.

Smith, C. J., Ashford, J. W. and Perfetti, T. A. (2019) 'Putative Survival Advantages in Young Apolipoprotein ε4 Carriers are Associated with Increased Neural Stress', *Journal of Alzheimer's Disease*, 68(3), pp. 885–923. doi: 10.3233/JAD-181089.

Snowdon, D. A. *et al.* (1996) 'Linguistic Ability in Early Life and Cognitive Function and Alzheimer's Disease in Late Life', *JAMA*, 275(7), p. 528. doi: 10.1001/jama.1996.03530310034029.

Soldan, A. *et al.* (2016) 'Hypothetical Preclinical Alzheimer Disease Groups and Longitudinal Cognitive Change', *JAMA Neurology*, 73(6), pp. 698–705. doi: 10.1001/jamaneurol.2016.0194.

Sommer, I. E. C. *et al.* (2004) 'Do women really have more bilateral language representation than men? A meta-analysis of functional imaging studies', *Brain*, 127(8), pp. 1845–1852. doi: 10.1093/brain/awh207.

Sperling, R. A. *et al.* (2011) 'Toward defining the preclinical stages of Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease.', *Alzheimer's & dementia*, 7(3), pp. 280–92. doi: 10.1016/j.jalz.2011.03.003.

Sperling, R. A. *et al.* (2014) 'The A4 study: stopping AD before symptoms begin?', *Science translational medicine*, 6(228). doi: 10.1126/scitranslmed.3007941.

Sperling, R. A. *et al.* (2018) 'The impact of Aβ and tau on prospective cognitive decline in older individuals', *Annals of Neurology*, 85, pp. 181–193. doi: 10.1002/ana.25395.

Sperling, R., Mormino, E. and Johnson, K. (2014) 'The Evolution of Preclinical Alzheimer's Disease: Implications for Prevention Trials', *Neuron*, 84(3), pp. 608–622. doi: 10.1016/j.neuron.2014.10.038.

Stafford, M. *et al.* (2013) 'Using a birth cohort to study ageing: representativeness and response rates in the National Survey of Health and Development.', *European journal of ageing*, 10(2), pp. 145–157. doi: 10.1007/s10433-013-0258-8.

StataCorp (2017) 'Stata statistical software: release 15'. College Station, Texas.

Stern, Y. (2012) 'Cognitive reserve in ageing and Alzheimer's disease.', *The Lancet. Neurology*, 11(11), pp. 1006–12. doi: 10.1016/S1474-4422(12)70191-6.

Stewart, R., Hardy, R. and Richards, M. (2015) 'Associations between skeletal growth in childhood and cognitive function in mid-life in a 53-year prospective birth cohort study.', *PloS one*, 10(4), p. e0124163. doi: 10.1371/journal.pone.0124163.

Stirling, L. A. *et al.* (2013) 'Use of a tracing task to assess visuomotor performance: effects of age, sex, and handedness.', *The journals of gerontology. Series A, Biological sciences and medical sciences*, 68(8), pp. 938–945. doi: 10.1093/gerona/glt003.

Stites, S. D., Milne, R. and Karlawish, J. (2018) 'Advances in Alzheimer's imaging are changing the experience of Alzheimer's disease', *Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring*, 10, pp. 285–300. doi: 10.1016/j.dadm.2018.02.006.

Stoet, G. (2017) 'Sex differences in the Simon task help to interpret sex differences in selective attention.', *Psychological research*, 81(3), pp. 571–581. doi: 10.1007/s00426-016-0763-4.

De Strooper, B. and Karran, E. (2016) 'The Cellular Phase of Alzheimer's Disease.', *Cell*, 164(4), pp. 603–615. doi: http://dx.doi.org/10.1016/j.cell.2015.12.056.

Stuss, D. T. *et al.* (2005) 'Multiple frontal systems controlling response speed', *Neuropsychologia*. Pergamon, 43(3), pp. 396–417. doi: 10.1016/J.NEUROPSYCHOLOGIA.2004.06.010.

Sudre, C. H. *et al.* (2015) 'Bayesian Model Selection for Pathological Neuroimaging Data Applied to White Matter Lesion Segmentation', *IEEE Transactions on Medical Imaging*, 34(10), pp. 2079–2102. doi: 10.1109/TMI.2015.2419072.

Sweeney, M. D. *et al.* (2019) 'Vascular dysfunction—The disregarded partner of Alzheimer's disease', *Alzheimer's & Dementia*, 15(1), pp. 158–167. doi: 10.1016/J.JALZ.2018.07.222.

Takeuchi, H. *et al.* (2017) 'Global associations between regional gray matter volume and diverse complex cognitive functions: evidence from a large sample study', *Scientific Reports*, 7(1), p. 10014. doi: 10.1038/S41598-017-10104-8.

Tales, A. *et al.* (2012) 'Intra-Individual Reaction Time Variability in Amnestic Mild Cognitive Impairment: A Precursor to Dementia?', *Journal of Alzheimer's Disease*, 32(2), pp. 457–466. doi: 10.3233/JAD-2012-120505.

Tang, Y., Lutz, M. W. and Xing, Y. (2018) 'A systems-based model of Alzheimer's disease', *Alzheimer's & Dementia*. doi: 10.1016/J.JALZ.2018.06.3058.

Tanner, A. *et al.* (2015) 'Barriers to Medical Research Participation as Perceived by Clinical Trial Investigators: Communicating with Rural and African American Communities', *Journal of Health Communication*, 20(1), pp. 88–96. doi: 10.1080/10810730.2014.908985.

Tariot, P. N. *et al.* (2018) 'The Alzheimer's Prevention Initiative Autosomal-Dominant Alzheimer's Disease Trial: A study of crenezumab versus placebo in preclinical PSEN1 E280A mutation carriers to evaluate efficacy and safety in the treatment of autosomal-dominant Alzheimer's diseas', *Alzheimer's & Dementia: Translational Research & Clinical Interventions*, 4, pp. 150–160. doi: 10.1016/J.TRCI.2018.02.002.

Thakkar, K. N. *et al.* (2014) 'Women are more sensitive than men to prior trial events on the Stop-signal task', *British Journal of Psychology*, 105(2), pp. 254–272. doi: 10.1111/bjop.12034.

Tippett, W. J., Krajewski, A. and Sergio, L. E. (2007) 'Visuomotor Integration Is Compromised in Alzheimer's Disease Patients Reaching for Remembered Targets', *European Neurology*, 58(1), pp. 1–11. doi: 10.1159/000102160.

Tippett, W. J. and Sergio, L. E. (2006) 'Visuomotor integration is impaired in early stage Alzheimer's disease.', *Brain research*, 1102(1), pp. 92–102. doi: 10.1016/j.brainres.2006.04.049.

Tippett, W. J., Sergio, L. E. and Black, S. E. (2012) 'Compromised visually guided

motor control in individuals with Alzheimer's disease: Can reliable distinctions be observed?', *Journal of Clinical Neuroscience*, 19(5), pp. 655–660. doi: 10.1016/j.jocn.2011.09.013.

Tuminello, E. R. and Duke Han, S. (2011) 'The apolipoprotein e antagonistic pleiotropy hypothesis: review and recommendations.', *International journal of Alzheimer's disease*, 2011, p. 726197. doi: 10.4061/2011/726197.

van Bergen, J. M. G. *et al.* (2018) 'Simultaneous quantitative susceptibility mapping and Flutemetamol-PET suggests local correlation of iron and β-amyloid as an indicator of cognitive performance at high age', *NeuroImage*, 174, pp. 308–316. doi: 10.1016/J.NEUROIMAGE.2018.03.021.

van Dijk, E. J. *et al.* (2008) 'Progression of Cerebral Small Vessel Disease in Relation to Risk Factors and Cognitive Consequences', *Stroke*, 39(10), pp. 2712–2719. doi: 10.1161/STROKEAHA.107.513176.

Van Hoorn, J. F. *et al.* (2010) 'Handwriting, visuomotor integration, and neurological condition at school age', *Developmental Medicine & Child Neurology*, 52(10), pp. 941–947. doi: 10.1111/j.1469-8749.2010.03715.x.

Vaportzis, E. *et al.* (2015a) 'Dual task performance in Huntington's disease: A comparison of choice reaction time tasks.', *Neuropsychology*, 29(5), pp. 703–712. doi: 10.1037/neu0000172.

Vaportzis, E. *et al.* (2015b) 'Effects of task difficulty during dual-task circle tracing in Huntington's disease.', *Journal of neurology*, 262(2), pp. 268–76. doi: https://doi.org/10.1007/s00415-014-7563-9.

Vaportzis, E., Georgiou-Karistianis, N. and Stout, J. C. (2013) 'Dual task performance in normal aging: a comparison of choice reaction time tasks.', *PloS one*. Public Library of Science, 8(3), p. e60265. doi: 10.1371/journal.pone.0060265.

Vaportzis, E., Georgiou-Karistianis, N. and Stout, J. C. (2014) 'Age and task difficulty differences in dual tasking using circle tracing and serial subtraction tasks.', *Aging clinical and experimental research*, 26(2), pp. 201–11. doi: https://doi.org/10.1007/s40520-013-0151-5.

Vemuri, P. *et al.* (2015) 'Vascular and amyloid pathologies are independent predictors of cognitive decline in normal elderly', *Brain*, 138(3), pp. 761–771. doi: 10.1093/brain/awu393.

Verheij, S. *et al.* (2012) 'Visuomotor impairment in early-stage Alzheimer's disease: changes in relative timing of eye and hand movements.', *Journal of Alzheimer's disease*, 30(1), pp. 131–43. doi: 10.3233/JAD-2012-111883.

Vernon, P. A. (1983) 'Speed of information processing and general intelligence', *Intelligence*, 7(1), pp. 53–70. doi: 10.1016/0160-2896(83)90006-5.

Vincent, A. S. *et al.* (2012) 'Automated Neuropsychological Assessment Metrics (v4) Traumatic Brain Injury Battery: military normative data.', *Military medicine*, 177(3), pp. 256–69. doi: 10.7205/milmed-d-11-00289.

Visser, P. J. *et al.* (2012) 'Comparison of International Working Group criteria and National Institute on Aging–Alzheimer's Association criteria for Alzheimer's disease', *Alzheimer's & Dementia*, 8(6), pp. 560–563. doi: 10.1016/J.JALZ.2011.10.008.

Vos, S. J. B. *et al.* (2013) 'Preclinical Alzheimer's disease and its outcome: A longitudinal cohort study', *The Lancet Neurology*, 12(10). doi: 10.1016/S1474-4422(13)70194-7.

Voyer, D. *et al.* (2007) 'Gender differences in object location memory: a meta-

analysis.', *Psychonomic bulletin & review*, 14(1), pp. 23–38. doi: https://doi.org/10.3758/BF03194024.

Voyer, D., Voyer, S. and Bryden, M. P. (1995) 'Magnitude of sex differences in spatial abilities: a meta-analysis and consideration of critical variables.', *Psychological bulletin*, 117(2), pp. 250–70. doi: http://dx.doi.org/10.1037/0033-2909.117.2.250.

Voyer, D., Voyer, S. D. and Saint-Aubin, J. (2017) 'Sex differences in visual-spatial working memory: A meta-analysis', *Psychonomic Bulletin & Review*, 24(2), pp. 307–334. doi: 10.3758/s13423-016-1085-7.

Wadsworth, M. *et al.* (2006) 'Cohort Profile: The 1946 National Birth Cohort (MRC National Survey of Health and Development)', *Int J Epidemiol*. 2005/10/06, 35(1), pp. 49–54. doi: 10.1093/ije/dyi201.

Wadsworth, M. E. *et al.* (1992) 'Loss and representativeness in a 43 year follow up of a national birth cohort.', *Journal of epidemiology and community health*, 46(3), pp. 300–4. doi: 10.1136/jech.46.3.300.

Wang, F. *et al.* (2015) 'Cerebral amyloidosis associated with cognitive decline in autosomal dominant Alzheimer disease.', *Neurology*, 85(9), pp. 790–8. doi: 10.1212/WNL.0000000000001903.

Wang, G. *et al.* (2018) 'Simultaneously evaluating the effect of baseline levels and longitudinal changes in disease biomarkers on cognition in dominantly inherited Alzheimer's disease.', *Alzheimer's & Dementia: Translational Research & Clinical Interventions*, 4, pp. 669–676. doi: 10.1016/j.trci.2018.10.009.

Wechsler, D. (1981) *Wechsler Adult Intelligence Scale–Revised*. San Antonio, TX: The Psychological Corporation.

Wechsler, D. (1987) *Wechsler Memory Scale - Revised: Manual*. San Antonio, TX: The Psychological Corporation.

Wechsler, D. (1999) *The Wechsler abbreviated scale of intelligence*. San Antonio, TX: The Psychological Corporation.

Weintraub, S. *et al.* (2009) 'The Alzheimer's Disease Centers' Uniform Data Set (UDS)', *Alzheimer Disease & Associated Disorders*, 23(2), pp. 91–101. doi: 10.1097/WAD.0b013e318191c7dd.

Weintraub, S. *et al.* (2018) 'Measuring cognition and function in the preclinical stage of Alzheimer's disease', *Alzheimer's & Dementia: Translational Research & Clinical Interventions*, 4, pp. 64–75. doi: 10.1016/J.TRCI.2018.01.003.

Whalley, L. J. *et al.* (2000) 'Childhood mental ability and dementia.', *Neurology*, 55(10), pp. 1455–9. doi: 10.1212/WNL.55.10.1455.

Whitwell, J. L. *et al.* (2012) 'Comparison of Imaging Biomarkers in the Alzheimer Disease Neuroimaging Initiative and the Mayo Clinic Study of Aging', *Archives of Neurology*, 69(5), p. 614. doi: 10.1001/archneurol.2011.3029.

Williams, B. R. *et al.* (2005) 'Inconsistency in Reaction Time Across the Life Span.', *Neuropsychology*, 19(1), pp. 88–96. doi: 10.1037/0894-4105.19.1.88.

Wirth, M. *et al.* (2007) 'Sex Differences in Semantic Processing: Event-Related Brain Potentials Distinguish between Lower and Higher Order Semantic Analysis during Word Reading', *Cerebral Cortex*, 17(9), pp. 1987–1997. doi: 10.1093/cercor/bhl121.

Woods, D. L. *et al.* (2015) 'Age-related slowing of response selection and production in a visual choice reaction time task', *Frontiers in Human Neuroscience*. Frontiers, 9. doi: 10.3389/fnhum.2015.00193.

World Health Organization (2019) *Dementia fact sheet*. Available at: https://www.who.int/news-room/fact-sheets/detail/dementia.

Xu, W. *et al.* (2015) 'Meta-analysis of modifiable risk factors for Alzheimer's disease', *Journal of Neurology, Neurosurgery & Psychiatry*. BMJ Publishing Group Ltd, 86(12), pp. 1299–1306. doi: 10.1136/jnnp-2015-310548.

Zetterberg, H. *et al.* (2009) 'Additional evidence for antagonistic pleiotropic effects of APOE', *Alzheimer's & Dementia*, 5(1), p. 75. doi: 10.1016/J.JALZ.2008.10.005.

Zetterberg, H. (2019) 'Blood-based biomarkers for Alzheimer's disease—An update', *Journal of Neuroscience Methods*, 319, pp. 2–6. doi: 10.1016/J.JNEUMETH.2018.10.025.

Zhao, Q.-F. *et al.* (2016) 'The prevalence of neuropsychiatric symptoms in Alzheimer's disease: Systematic review and meta-analysis', *Journal of Affective Disorders*, 190, pp. 264–271. doi: 10.1016/j.jad.2015.09.069.

Zokaei, N. *et al.* (2017) 'Sex and APOE: A memory advantage in male APOE ε4 carriers in midlife.', *Cortex*, 88, pp. 98–105. doi: 10.1016/j.cortex.2016.12.016.

Zokaei, N., Nour, M. M., *et al.* (2019) 'Binding deficits in visual short-term memory in patients with temporal lobe lobectomy.', *Hippocampus*, 29(2), pp. 63–67. doi: 10.1002/hipo.22998.

Zokaei, N., Čepukaitytė, G., *et al.* (2019) 'Dissociable effects of the apolipoprotein-E (APOE) gene on short- and long-term memories', *Neurobiology of Aging*, 73, pp. 115–122. doi: 10.1016/J.NEUROBIOLAGING.2018.09.017.

STATEMENT OF ATTRIBUTION

Chapter 3

Insight 46 was conceived and planned by Professors Jonathan Schott, Nick Fox, Marcus Richards and Diana Kuh. Dr Christopher Lane, Dr Thomas Parker, Dr David Cash, Elizabeth Donnachie, Heidi Murray-Smith, Suzie Barker and Dr Michelle Byford were instrumental in designing the study protocol and preparing the ethics application. The cognitive battery was designed by Professors Sebastian Crutch and Marcus Richards. The imaging protocol and processing pipelines were developed by Dr David Cash, Dr Ian Malone, Dr Marc Modat, Dr Carole Sudre, Dr David Thomas, Dr Gary Zhang, Dr Anna Barnes, Dr John Dickson, and Professor Sebastien Ourselin. Recruitment and clinical assessments were performed by Drs Christopher Lane, Thomas Parker, Ashvini Keshavan, Sarah Buchanan and Sarah Keuss. Co-ordination and booking of participants' travel and accommodation was performed by Heidi Murray-Smith, Claudia Cramer, Molly Cooper, Elizabeth Burgnon, Jessica Collins and me. Neuropsychology assessments were performed by me (225 assessments), Jessica Collins, Sarah James and Elizabeth Donnachie. Jana Klimova and Will Coath performed QC of volumetric T1, T2 and FLAIR images. Dr Ian Malone and Elizabeth Gordon managed the volumetric pipeline that generated whole brain volume and hippocampal volumes and were responsible for manual editing. The BaMoS pipeline was run by Dr Carole Sudre and BaMoS QC and manual editing was performed by Dr Christopher Lane as required. B-amyloid PET processing, imputation work and determination of the cut-point for positivity was performed by Dr David Cash. Andrew Wong and Heidi Murray-Smith co-ordinated the processing of *APOE* genotyping performed by LGC Hoddesdon on blood samples collected by Drs Christopher Lane, Thomas Parker, Ashvini Keshavan, Sarah Buchanan and Sarah Keuss. Processing of computerised cognitive tests and extraction of outcome variables was performed by me. Statistical advice was provided by Dr Jennifer Nicholas.

Chapter 4

I conceived and designed this study with advice from Dr Susie Henley and Professors Schott, Richards and Crutch. I designed the statistical models with advice from Dr Jennifer Nicholas. I performed statistical analysis and interpretation of the results.

Chapter 5

The “What was where?” task was designed by Dr Yoni Pertzov and Professor Masud Husain. I conceived and designed this study with advice from Dr Susie Henley and Professors Schott, Richards and Crutch. I designed the statistical models with advice from Dr Jennifer Nicholas. I performed statistical analysis and interpretation of the results. Dr Yoni Pertzov provided valuable feedback on this chapter.

Chapter 6

I conceived and designed this study with advice from Dr Susie Henley and Professors Schott, Richards and Crutch. I designed the statistical models with advice from Dr Jennifer Nicholas. I performed statistical analysis and interpretation of the results.

Chapter 7

I conceived and designed this study with advice from Dr Susie Henley and Professors Schott, Richards and Crutch. I designed the statistical models with advice from Dr Jennifer Nicholas. I performed statistical analysis and interpretation of the results.

Chapter 8

The circle-tracing task was designed by Professor Julie Stout and her colleagues at the University of Monash, who provided advice on extraction of the outcome variables. I conceived and designed this study with advice from Dr Susie Henley and Professors Schott, Richards and Crutch. I designed the statistical models with advice from Dr Jennifer Nicholas. I performed statistical analysis and interpretation of the results.

Chapter 9

I conceived and designed the cross-validation work (section 9.2) in collaboration with Drs Nick Firth, Neil Oxtoby, Susie Henley, and Professor Sebastian Crutch. Drs Nick Firth and Neil Oxtoby provided advice on the methods. Neil Oxtoby provided valuable feedback on this section of the chapter. Drs Ashvini Keshavan and Jennifer Nicholas provided advice on the statistical methods used in section 9.3.2. I performed statistical analysis and interpretation of the results.

Chapter 10

Comparison of Insight 46 individuals with those still in the main cohort at age 69 was performed by Dr Sarah James. I was responsible for the interpretation of the results.

PUBLICATIONS

Publications that have arisen to date as a direct result of the work in this thesis:

Chapter 3

Lane C A, Parker T D, Cash D M, **Macpherson K**, Donnachie E, Murray-Smith H., … Schott J M. (2017). Study protocol: Insight 46 – a neuroscience sub-study of the MRC National Survey of Health and Development. BMC Neurology, 17(1), 75.

Chapter 4

Lu K, Nicholas J M, Collins J D, James S-J, Parker T D, Lane C A,… Richards M, Schott J M. (2019) Cognition at age 70: life course predictors and associations with brain pathologies, *Neurology*, Accepted.

ACKNOWLEDGEMENTS

Firstly, I would like to thank the Insight 46 participants for their contribution of time and enthusiasm, without which this study would not have been possible. I very much enjoyed meeting such an interesting group of people and was inspired by their extraordinary commitment to research.

I am indebted to my supervisors Susie Henley, Sebastian Crutch, Jonathan Schott and Marcus Richards, whose support, guidance, knowledge and enthusiasm enabled me to have such a positive PhD experience. It has been a privilege to learn from you all. In particular, I would like to thank my primary supervisor, Susie, who continued to supervise me after leaving UCL. I am extremely grateful for your generosity of time and spirit, your interest in every detail of my work, and your endless encouragement.

The success of Insight 46 is due to a large team of talented and hard-working individuals: our study coordinator Heidi Murray-Smith; the study doctors Chris Lane, Tom Parker, Ashvini Keshavan, Sarah Buchanan and Sarah Keuss; the psychologists Jessica Collins, Elizabeth Donnachie and Sarah James, the neuroimaging team Dave Cash, Ian Malone, Dave Thomas, Anna Barnes, John Dickson, Marc Modat, Carole Sudre, Jana Klimova, Elizabeth Gordon and Will Coath; Suzie Barker our research governance manager; administrator Molly Cooper; radiographers at the Macmillan Cancer Centre who performed the MRI/PET scans; and colleagues from the Lifelong Health and Ageing Unit including Andy Wong, Sarah James, Marcus Richards and Diana Kuh. I am very grateful to Jenny Nicholas for her statistical expertise and her patience in sharing it. I would also like to acknowledge other colleagues, many of whom have become friends, who make the Dementia Research Centre such a good place to work, particularly Ivanna Pavisic, Emma Harding, Chris Hardy, Keir Yong, Aida Suarez Gonzalez, Hannah Carr, Coraline Daeninck, Jane Douglas and Jake Cinco. I am also grateful to Jo Barnes and Elizabeth Halton for their work in supporting PhD students and providing practical advice, and to Nick Fox for his leadership and encouragement.

Last but by no means least, I would like to thank my wonderful husband Yang and my parents and siblings, for all their love and support. I am particularly grateful to Yang for cheering me on during the thesis-writing process. And to my God, in whom I live.

APPENDIX

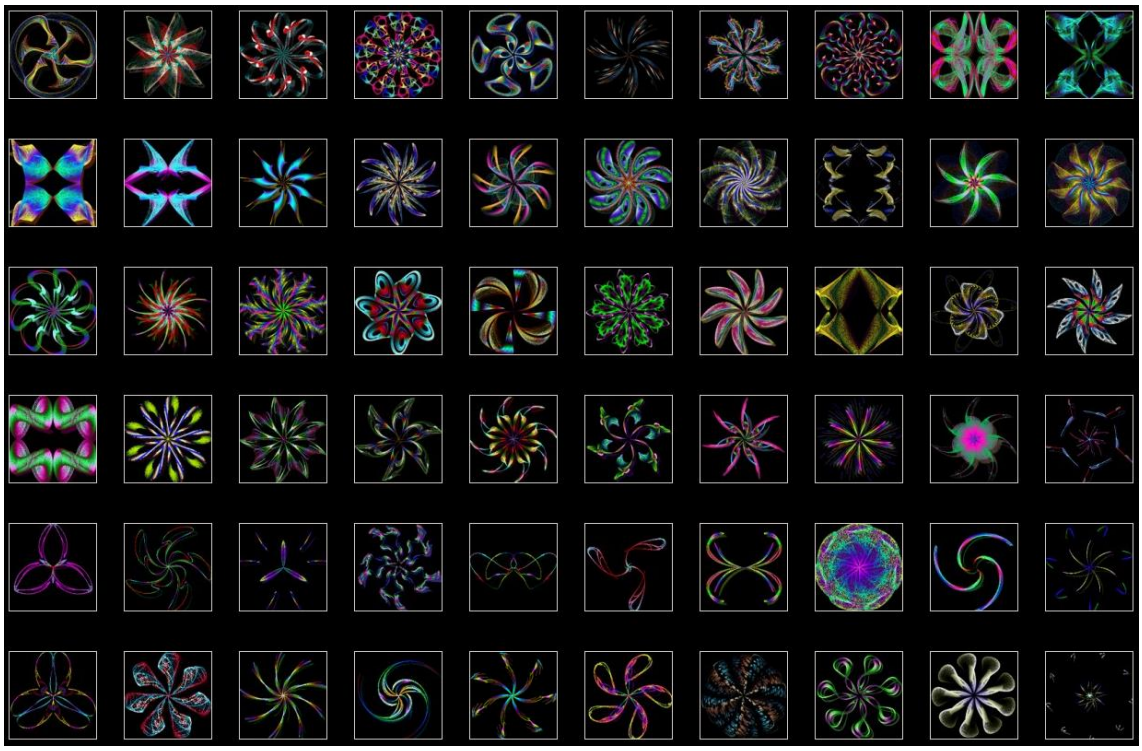


Figure A1. Fractals used in the "What was where?" task

Figure reprinted from Liang et al. (2016) under the terms of the Creative Commons Attribution License (CC BY).