Understanding Cognitive Variability in Alzheimer's Disease

Doherty, Thomas

https://doi.org/10.34737/w166v

Understanding Cognitive Variability in Alzheimer's Disease

Thomas Andrew Doherty

University of Westminster
Doctorate of Philosophy

2022

Abstract

Alzheimer’s Disease (AD) is highly heterogenous, both clinically and biologically. This variability is exacerbated by the ways within which, the clinical presentation is assessed with cognitive measures. This inhibits clinical trial success and earlier diagnosis of individuals. Marrying the clinical presentation to the pathology of the disease has so far proved troublesome. This thesis will look at how cognitive measures can best capture the clinical presentation of AD and how these measures can link to the underlying pathology using machine learning methods.

This thesis studied this problem across four analyses and two cohorts. Each study looked at a different aspect of cognitive testing within AD. This was done with the overarching aim to interrogate the cognitive variability across the spectrum of AD.

Study 1 showed a novel discrepancy score is different to memory measures at screening for AD. It also showed it tracks with AD severity, in the same way memory recall does. Studies 2 & 3 uncovered broad psychometric variance within amnestic measurement of impairment due to AD. This was done in two different populations across two different constructs of amnestic measurement, story recall and verbal list learning. These tests are frequently used interchangeably. These two studies show they should not be. Finally, Study 4 built models from cognitive measures to predict AD pathology. The performance of these models was moderate showing that even with novel cognitive measures, further work is needed to link the clinical and amyloid related biological presentations of AD.

Bridging the gap between clinical presentation and pathology of AD using clinical and cognitive markers alone is not possible. Even when using a novel measure of discrepancy score. The discrepancy measure shows promise but was limited due to the inability of the MMSE to measure verbal ability.

Conceptually a discrepancy score remains a promising avenue of research for screening, but broader language measures, as well as other AD biomarkers are needed to further test the construct validity of this measure.

Table of Contents

List of Tables & Figures

Figures

Tables

Acknowledgements

For Phyliss

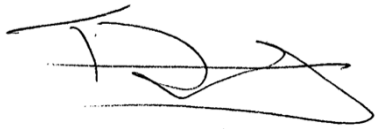
Memory, of all the powers of the mind, is the most delicate and frail.
Ben Jonson, 1640

With thanks to, Katie, Caroline, Alan, Anna & Kimbie. A big thank you to my supervisory team; Professor Catherine Loveday, Professor Tom Buchanan & Dr Samuel Evans. Many thanks to Dr Michelle Gee, Professor Bruce Albala & Dr Michael Irizarry for their assistance in sourcing the data from Mission AD. As well as to Dr Bob Smith, Dr Jeri Morris and Michael Smith for their assistance with data collection for the Alternate Story Recall dataset.

Word Count – 76,732

Authors Declaration

I declare that all material contained in this thesis is my own work



Definitions & Glossary of Terms

A	Amyloid
AAMI	Age Associated Memory Impairment
Aß	Amyloid Beta protein
ACE-III	Addenbrookes Cognitive Examination Version 3
AChEI	Acetylcholinesterase Inhibitor
AD	Alzheimer's Disease
ADAD	Autosomal Dominant Alzheimer's Disease
ADAM10/17	A Disintegrin and metalloproteinase domain-containing protein (10/17)
ADAS-Cog	Alzheimer's Disease Assessment Scale – Cognitive Subscale (# reflect # of items)
ADCOMS	Alzheimer's Disease Composite Scale
ADCS	Alzheimer's Disease Co-Operative Study
ADL	Activities of Daily Living
ADNI	Alzheimer's Disease Neuroimaging Initiative
AE	Adverse Event
AI	Artificial Intelligence
AIBL	Australian Imaging, Biomarker & Lifestyle Flagship Study of Aging
AICD	Amyloid Intracellular domain
ALC	Absolute Lymphocyte Count
ANOVA	Analysis of Variance
APP	Amyloid Precursor Protein
APOE	Apolipoprotein E (polymorphic alleles – e2, e3, e4; epsilon 2,3,4)
ARIA E/H	Amyloid Related Imaging Abnormality – Vasogenic Edema/Micro-Haemorrhage
AS	Alternate Story Recall
ATN	Amyloid, Tau, Neurodegeneration diagnostic framework
AT(X)N	ATN Framework + additional unknown (X) measures to be included later
BA	Bland & Altman Method Of Measuring Agreement
BACEi	BACE inhibitor
BACE1/2	Beta site APP cleaving enzyme (1 or 2)
BAN2401	Compound name of Lecanemab
BBB	Blood Brain Barrier
BNT	Boston Naming Test
BP	Blood Pressure
CAA	Cerebral Amyloid Angiopathy
CAD	Computer Aided Diagnosis
CAMD	Coalition Against Major Diseases
CBB	Cogstate Brief Battery
CERAD	Consortium to Establish a Registry for Alzheimer's Disease
CIBIC	The Clinician's Interview-Based Impression of Change
CD33	sialic acid binding Ig-like lectin 3
CDR/-SB/-GS	Clinical Dementia Rating Scale/ Sum Of Boxes / Global Score
CN	Cognitively Normal/Cognitively Unimpaired
CNS	Central Nervous System
CPAD	Critical Path for Alzheimer's Disease
CRBM	Conditional Restricted Boltzmann Machine

CR1	Complement Receptor 1
CSF	Cerebral Spinal Fluid
CT	Computerised Tomography
CTF(89/99)	c-terminus fragment
CU	Cognitively Unimpaired/Cognitively Normal
CV	Cardiovascular
CVLT	California Verbal Learning Test
C99	Cell Membrane fragment 99
DET	Detection Test from the Cogstate Brief Battery
DIAD	Dominantly Inherited Alzheimer's Disease
DIAN/DIAN-TU	Dominantly Inherited Alzheimer's Disease Network (Trials Unit)
DKEFS	Delis-Kaplan Executive Function System
DLPFC	Dorso-lateral Pre-Frontal Cortex
DMN	Default Mode Network
DRS	Dementia Rating Scale
DS	Discrepancy Score
DSST	Digit Symbol Substitution Test
EC	Ethics Committee
ECG	Electro-cardiogram
ELISA	Enzyme-linked immunosorbent assay
EMA	European Medicines Agency
EMEA	Europe, The Middle East & Africa
EMIF-AD	European Medical Information Framework – Alzheimer's Disease
EPAD	European Prevention of Alzheimer's Disease
EQ-5D-5L	EuroQoL (Quality of Life) Scale – 5 dimension – 5 levels
FAQ	Functional Assessment Questionnaire
FCD	Functional Cognitive Disorder
FCSRT	Free & Cued Selective Reminding Task
FDA	Food & Drug Administration
FDG	Fluorodeoxyglucose (PET)
FLAIR	Fluid Attenuation Inversion Recovery - MRI sequence
fMRI	Functional Magnetic Resonance Imaging
g	General Intelligence Factor
GDS	Geriatric Depression Scale
GWAS	Genome Wide Association Study
HVLT	Hopkins Verbal Learning Test
IDE	Identification Test from the Cogstate Brief Battery
INR	International Normalised Ratio of Prothrombin protein
IQ	Intelligence Quotient
IRB	Institutional Review Board
ISLT	International Shopping List Task
ISLTDR	Delayed recall score from the ISLT
ISLTTR	Immediate recall score from the ISLT
IST	Isaacs Set Test
LDA	Linear Discriminant Analysis
LLN	Lower Limit of Normal
LLOA	Lower Limit of Agreement
Log10	Logarithmic Scale Base 10
LTP	Long term potentiation

N	Neurodegeneration
NART	National Adult Reading Test
NFL	Neurofilament Light chain protein
NFT	Neurofibrillary Tangles
ng	Neurogranin
NHST	Null Hypothesis Significance Testing
NIA-AA	National Institute of Aging – Alzheimer's Association
NMDA	N-methyl-D-aspartic acid
NPI-10	Neuropsychiatric Inventory – 10 item questionnaire
NPV	Negative Predictive Value
mAbs	Monoclonal Antibodies
MANOVA	Multivariate Analysis of Variance
MAP1/2	Microtubule Associated Proteins 1a,1b, 1, 2
MCI	Mild Cognitive Impairment
MD	Mean Difference
MoCA	Montreal Cognitive Assessment
MRI	Magnetic Resonance Imaging
MAO-A/B	Monoamine Oxidase A/B
ML	Machine Learning
MMSE	Mini Mental State Examination
MSD	Meso Scale Discovery Immunoassay
OCL	One Card Learning Test from the Cogstate Brief Battery
OLE	Open Label Extension
ONB	One Back Test from the Cogstate Brief Battery
PACC	Preclinical Alzheimer's Disease Cognitive Composite
PCA	Principal Components Analysis
PD	Pharmacodynamic
PGx	Pharmacogenomic
PK	Pharmacokinetic
PLS	Partial Least Squares
PET	Positron Emission Topography
PiB	Pittsburgh Compound B PET ligand
PHF	Paired Helical Filament
PPV	Positive Predictive Value
PSEN1/2	Presenilin protein 1 or 2
p-tau	Phosphorylated tau protein
QoL-AD	Quality of Life – Alzheimer's Disease Scale
RAVLT	Rey Auditory Verbal Learning Test
RBANS	Repeatable Battery for the Assessment of Neuropsychological Status
RNA	Ribonucleic Acid
ROC	Receiver Operating Characteristic
sAPP	Soluble amyloid precursor protein - extracellular fragment (alpha or beta)
SD	Standard Deviation
SEZ6	Seizure Protein 6
SIB	Severe Impairment Battery
SNAP	Suspected non-AD Pathology
SPECT	Single Photon Emission Computerised Tomography
sTREM2	Soluble Triggering Receptor Expressed On Myeloid Cells 2
SUVr	Standard Uptake Value Ratio – MRI measurement

SVM	Support Vector Machines
T	Tau
t-tau	Total tau
TB	Tuberculosis
TDP-43	Transactive response DNA-binding protein 43
TMT	Trial Making Test
TREM2	Triggering Receptor Expressed On Myeloid Cells 2
TSH	Thyroid Stimulating Hormone
ULN	Upper Limit of Normal
ULOA	Upper Limit of Agreement
US/USA	United States of America
VIP	Variable Importance Projection
VLL	Verbal List Learning
vMRI	Volumetric Magnetic Resonance Imaging
VRM	Verbal Recognition Memory
VRT	Visual Retention Test
W-TAR	Weschler Test of Adult Reading
WAIS-R	Weschler Adult Intelligence Scale - Revised
WISC-R	Weschler Intelligence Scale for Children - Revised
WMS/-LM	Weschler Memory Scale (Logical Memory Subscale)
xMAP	Proprietary Luminex multiplex assay
ZBI	Zarit Burden Interview
3R/4R tau	3/4 repeated chain of proteins within a tau protein
11C	Carbon-11 ligand base
18F	Fluorine-18 ligand base

Psychometric Properties & Validity Definitions

Content	How well do measures accurately index domains they are designed to measure
Criterion	How well do measures index the disease process itself within AD
Cross-Cultural	Do measures perform the same way across different demographics
Construct	How well scores from a measure, reflect the domain being measured
Concurrent	How well do measures with the same construct provide the same outcome
Convergent	How alike are measures that measure the same domain
Ecological	How a measure matches the real-world context it is being evaluated for
Interpretability	How do measures relate to everyday performance/a clinical presentation

Chapter I –Thesis Outline & Review of the Literature

Thesis Overview

Alzheimer's Disease (AD) is the most widely researched neurodegenerative disorder, yet the number of individuals living with the disease is expected to double nearly every twenty years (Prince et al., 2015). The diagnosis of the disease has greatly changed over the last decade due to a large number of small breakthroughs, new biomarkers, imaging methods and better agreement on disease staging criteria. Whilst great strides have been made, there are many yet to take. Cognitive impairment still remains the first and critical symptom of the disease, yet is not the first pathological domain. Marrying the clinical presentation to the early pathology of the disease still remains difficult despite many efforts. This thesis will look to how cognitive measures can best capture the clinical presentation of AD and how cognitive measures can be used to index cognitive performance and link this to the underlying pathology.

The lack of confluence of the clinical presentation and pathology within AD, has meant AD currently has no therapeutics that cure or modify the disease course. Since the last therapies were approved in the mid 90s, the diagnostic criteria have vastly changed in line with our understanding. Cognitive measures have always been used to find decline and measure treatment response. However, there are no set or agreed upon measures to find this impairment. Many are used and most were developed decades ago to measure mild to severe AD, with only 50% of currently used measures having published information about their validity (Soobiah et al., 2019). In order to link clinical disease to pathology, a better understanding of the variability of these measures is critical and a fundamental pillar of this thesis.

The focus of this thesis revolves around understanding this discrepancy within clinical trials for AD. Clinical trials provide a unique opportunity to reach more individuals than a conventional research study and provide highly controlled settings for cognitive measurement. The first cohort under study here is taken from the two largest clinical trials even conducted in AD and spans 29 countries. This should allow for wide contextualisation and scope of findings from this thesis.

General Aims & Thesis Outline

The general aim of this thesis is to better understand the measurement of impairment for diagnosis and classification of AD. This will be achieved through creating and assessing a discrepancy score of cognitive functions and exploring how these existing and new cognitive measures can help

bridge the gap between clinical presentation and the underlying biological pathology within AD. To accomplish this, our measurement of memory within AD needs to be better understood. **Chapter 1** will give an overview of the issues facing the field and the key topic areas for this thesis; Diagnosis & Classification, The Amyloid Hypothesis & Disease Biomarkers, Neuroimaging and finally Cognition Within AD. **Chapter 2** will detail the study outlines and general concepts regarding the studies to be undertaken. Plus, also going into great detail on the cohorts and scales under investigation. This will be done in order to answer a number of research questions.

This thesis aims to provide important new insights into the following questions;

- Why do cognitive measures vary so much within clinical trials and cohort studies?
- Why have all trials to date failed to meet their cognitive endpoints?
- How can you accurately diagnose Alzheimer's Disease with cognitive tests?
- Can you use existing (and commonly used) cognitive measures in a new way to better understand cognitive performance over the course of the disease

This will be achieved by addressing the following research questions and hypotheses;

- How does a novel discrepancy measure relates to memory measures for AD diagnosis?
- What can this novel discrepancy score tell us about cognitive processes in general?
- Does this novel discrepancy score capture different cognitive processes to that of amnestic impairment?
- How does this novel discrepancy score change throughout the progression of the disease?
- Can commonly used verbal list learning (memory) measures be used interchangeably?
- How do these measures perform within different populations;
    - Are they consistent enough for use in clinic screening?
    - How do they perform within individuals with confirmed AD and high levels of pathology?
- Does this measurement of amnestic impairment also apply across measures of memory with different construct paradigms?
- Can this new knowledge of cognitive measures in AD be used to predict amyloid pathology, the cornerstone of AD pathology?
- Can this help link the biological classification of AD to the clinical presentation?

Chapter 3 will aim to address the first four research questions on discrepancy scores. This will be done by first taking the full screening cohort

and computing a discrepancy measure from the existing cognitive measures, using a crystallised language score from the MMSE and a fluid cognitive score from the Cogstate Brief Battery. The analysis will look to understand how this discrepancy score relates to amnestic performance. The cohort will then be staged based upon level of impairment (as measured by the CDR) to understand how this relationship between amnestic performance and discrepancy score changes with disease severity.

Moving on to focus on the amnestic memory measures further, **Chapter 4** will analyse the psychometric validity of four commonly used memory measures within AD. This will be done across two studies using two separate cohorts and two separate pairs of measures but identical analyses. These pairs of measures are both purported to measure episodic memory, in that the first pair is a word list recall and the second pair is based upon story recall.

The fourth and final analysis in **Chapter 5** will then use these findings to explore whether it is possible to predict amyloid pathology within confirmed AD individuals, using existing and cognitive measures and the newly established discrepancy scores. This will be done using a machine learning framework for classification called support vector machines. The main findings from all of these studies will be discussed in **Chapter 6** followed by a discussion of the implications of these along with any issues that arise and future directions for research.

Literature Review Outline

The following literature review covers four fundamental areas of AD research: classification of AD, biomarkers, neuroimaging and cognition. As detailed in the thesis outline, diagnosis and classification is the ultimate goal of performing cognitive tests, so this cornerstone of AD research is discussed in detail along with how current approaches, tackle the clinical and biological phenotypes of the disease. The literature review then moves to look in more depth at the cognitive domains and measures used to identify and assess the course of the disease. Finally, the ways pathology is measured using neuroimaging and other biomarker modalities are broadly discussed, as this is key to helping bridge the gap between clinical presentation and AD pathology. To note, further reviews of the current clinical trial landscape as well as more details on pathology composition can be found in **Appendix 1 & 2** respectively.

A comprehensive literature search was performed in order to identify the publications that investigated the four areas of review. Longitudinal studies and cross-sectional analysis were discriminated where appropriate and extensive studies across a single area corresponding to the discussion were sought in order to give a complete overview of the constituent research area.

Firstly, a broad literature search of electronic databases Medline, clintrials.gov, University of Westminster Library and Science Direct was performed, using dates respective of the individual subject area relevance. Search terms were based upon each heading and sub heading within the review below. Eligible articles were those reporting the results, reviews or meta-analysis of studies investigating the diagnosis of Alzheimer's Disease, cognitive indices within the course of Alzheimer's Disease pathogenesis and progression and drug trials related to the key disease modifying treatments currently under development of Alzheimer's Disease. Cochrane databases were also searched for appropriate reviews of treatments and cerebrospinal fluid (CSF) biomarkers within the spectrum of Alzheimer's Disease. The search of clinical trials through the US database was also done to uncover current status, inclusion criteria, prior study results and cognitive measures employed within the studies.

Defining Alzheimer's Disease

The History of AD Diagnosis & Classification (1906-2016)

Alzheimer's Disease as it became known was first discovered by Alois Alzheimer way back in 1906, when he discovered "an unusual disease of the cerebral cortex". He made this discovery when studying the brain of a patient of his, over many years, Auguste Deter. Using staining techniques Alzheimer identified pathological anomalies in her brain that were amyloid plaques and neurofibrillary tangles (Hippius & Neundorfer, 2003). These two pathologies form the fundamental basis of the post-mortem diagnosis of AD. Clinically this patient presented with paranoia, progressive sleep and memory disturbance, aggression and confusion, which were present consistently up until her death, 5 years later. This pathological and clinical presentation was first documented as Alzheimer's Disease in 1910 in the 3rd edition of Psychiatrie (Kraepelin, 1910). Alzheimer also published 3 further cases whereby they only had a presence of plaques (which were subsequently determined to consist of amyloid), without any tangle pathology. Upon re-examination these cases were confirmed to be different stages of the same process (Maurer, Volk & Gerbaldo, 1997). Thus, confirming the progressive nature of AD in initial studied cases.

Up until the mid 80's this was generally described as senile dementia and commonly viewed as a natural process of aging. The reality is clearly fundamentally different. In 1984, the first international consensus clinical diagnostic criteria were developed (McKhann et al., 1984). This defined the clinical presentation of dementia with memory changes and another cognitive impairment. This was primarily done by ruling all other diagnoses and diseases out before AD. What is important to highlight here is the ambiguous and varying use of different domains or measures to used to find these changes/impairments. This is something that has been carried through to the 2021 criteria (**Figure 1.1.**). After the 1984 criteria, there began a 7-year period between 2007 and 2014 whereby the diagnostic criteria began to be updated for research and clinical settings (Dubios et al., 2007; 2010; 2014; Mckhann et al., 2011; Albert et al., 2011; Sperling et al., 2011). This was done four times as the field rapidly developed in vivo techniques to measure amyloid and tau pathology. This was down to the development in our understanding of the pathological processes that drive the neurodegeneration of the cortex. With the rapid advent of pathologic biomarkers in-vivo for AD, Mild Cognitive Impairment (MCI) due to AD and progressed AD has evolved (see **Figure 1.1.** for measurements of these). On top of this preclinical AD was also defined more clearly and considered to occur when these aforementioned markers are present in cognitively normal individuals (McKhann et al., 2011). The concept of preclinical AD primarily arose in the late 20th century, with it initially defined as individuals who were cognitively unimpaired but who also displayed brain lesions of AD

nature upon post-mortem examination (Hubbard et al., 1990). However, the frameworks in 2011 and 2014 began the movement towards extending the disease from one merely of symptoms to one that pre-dates symptoms by up to two decades. However, even with the changes in framework the ambiguity of clinical measurement of cognitive impairment remained. This is highly relevant for symptomatic and disease-modifying trials within AD. As within these trials, the treatments that are tested aim to improve cognition and function (Vellas et al., 2008; Andrews et al., 2019).
At this point it is also important to mention the definition of atypical variants of AD. These rare variants include the behaviour-frontal, posterior-cortical and logopenic-language variants of AD. These atypical variants are currently estimated to represent around 6% of AD cases in the elderly (Graff-Radford et al., 2021; Koedam et al., 2010). Given their distinctive clinical presentation they are easily diagnosed when seen by clinicians, with the exception of the posterior-cortical variant which needs further imaging to be diagnosed (Dubios et al., 2014).

On top of these variants, pathological comorbidities are broadly common within the general elderly population (Ferreira, Nordberg and Westman, 2020). Recent observational and memory clinic cohort studies have also shown, that broad comorbidities are present in normal elderly, MCI and AD individuals (de Jager et al., 2018). Whilst within a singular disease there will be prominent related pathology it is likely that other pathological processes are at play. Disentangling the heterogeneity of a singular neurodegenerative disease from a pathological standpoint is particularly critical when assessing drugs aimed at targeting a singular pathology. However, given the absence of in-vivo imaging methods for other pathological abnormalities such as alpha-synuclein and TDP-43 amongst others, bridging pathology and clinical presentation requires methodological improvements from both aspects, concurrently. Recently there has also been further definition of LATE pathology, Limbic-predominant Age-related TDP-43 Encephalopathy (Nelson et al., 2019). This is currently thought to be only really present in those 80 and above, with the pathology seen post mortem defined as separate from AD due to the predominance of TDP-43 pathology. However, measurement of TDP-43 in vivo remains challenging and given that this disease presents clinically in a very similar way to AD understanding the disease process in this older age group becomes, shows the heterogeneity of pathology only gets broader the older the individual becomes. The crux of the current landscape is the distinction between the clinical phenotype of AD and the biological definition of AD. Both carry great importance and are equally as important to see further understanding within.

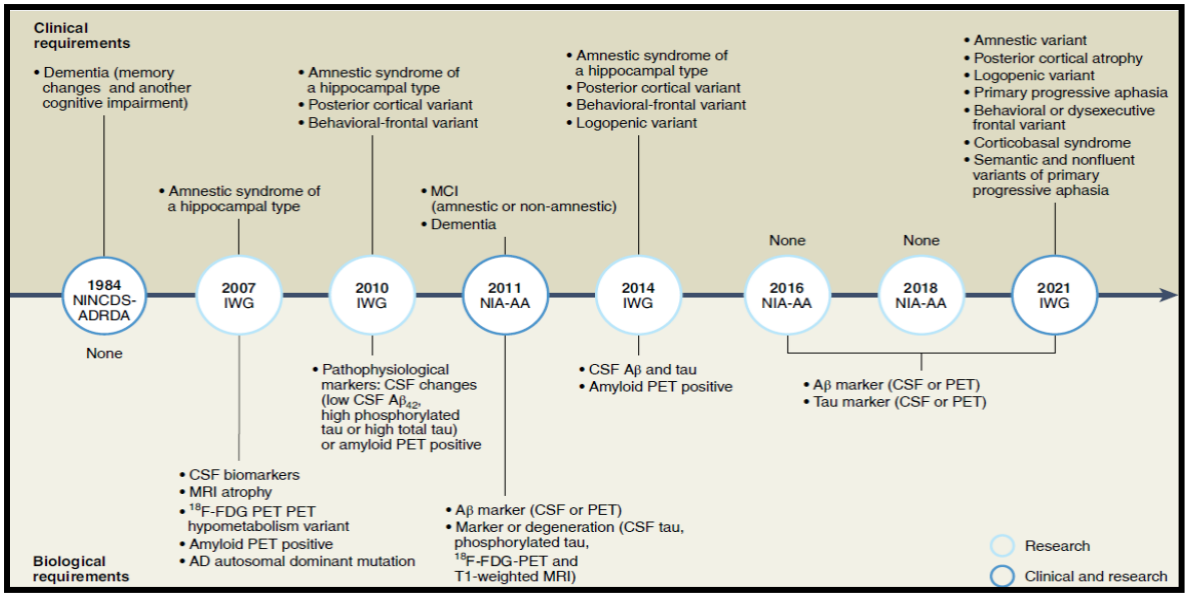


Figure 1.1. Evolution of the diagnostic criteria for AD from biological and clinical perspectives. Adapted from Hampel et al., 2022.

Amyloid, Tau & Neurodegeneration (ATN) Diagnostic Criteria (2016-Present)

The conceptual shift and a push for earlier detection has resulted in the development of clear diagnostic frameworks founded in the biological understanding of the pathological process within AD, termed the ATN framework (Dubois et al., 2016; Jack et al., 2013; 2014; 2016; 2018; 2021). This framework explores in much greater detail the earlier stage of the disease and splits the continuum of AD into six discrete yet continuous stages. This merging of modalities and classification update is however, not intended for clinical care yet and sits within the interventional and observational research spaces. The differences between the two entities have led to two distinct ways of describing AD: prototypical clinical syndromes, without biomarker/imaging verification and verified AD neuropathologic changes. Whilst the importance and relevance of clinical symptomology must not be underestimated, a syndrome is by definition a clinical manifestation of one or more diseases and not an aetiology. Shifting towards a biological empirically quantifiable definition of AD allows for a greater understanding, primarily of the mechanisms driving the clinical manifestations across the disease course. And importantly with new interventions having specific pathological targets, having a biological definition allows for more rigorous population categorisation and targeting earlier in the disease course. Which is in line with the current avenues drug development is accelerating towards (Cummings et al., 2021).

Development of AD biomarkers and cogent disease progression indicators have been copious and consequently well validated (Hansson, 2021). With new measures coming through consistently, any criteria must be flexible

enough to include these new measures for future classifications. This is particularly prevalent within AD classification as various imaging and cerebrospinal fluid (CSF) biomarkers are implemented within AD research and it is upon such biomarkers this updated disease staging is founded. This ATN classification refers to Amyloid [aggregated amyloid beta (Aβ) or associated pathological state], Tau [aggregated tau – neurofibrillary tangles or associated pathological state] and Neurodegeneration [marker/s of neurodegeneration or neuronal injury measured in vitro] (Jack et al., 2016; 2017). These three general groups mirror the nature of the pathological process within AD, biomarkers of Aβ [A] are currently cortical PET amyloid ligand binding and low levels of CSF Aβ (Fagan et al., 2007; Mattsson et al., 2009; Visser et al., 2009). Tau [T] includes markers of phosphorylated tau (p-tau) elevated in CSF and cortical PET tau ligand binding (Mattsson et al., 2009; Buerger et al., 2006; Brier et al., 2016; Chhatwal et al., 2016). The final part of this diagnostic criteria is Neurodegeneration [N], which is indexed through CSF total tau (t-tau), Fluorodeoxyglucose (FDG) PET hypometabolism and atrophy present on an MRI scan (Blennow, 2010; Seab et al., 1988; Fox et al., 2001; Minoshima et al., 1997; Besson et al., 1990; Dickerson et al., 2009; Knopman et al., 2013; Landau et al., 2011). The N component also portrays the lack of convergence with neuropathic findings for diagnosis of AD and is not specific in the way A and T are within the pathogenesis of AD. Nevertheless, as N is more ambiguous and lacking a direct correspondent to a diagnosis of AD, it is open to more flexibility. With the speed biomarkers such as neurofilament light chain (NFL) and neurogranin are becoming highly specific and utilised more widely, this criterion allows for potential future markers of neurodegeneration to be integrated into any model and is a key to the adoption of it into wider clinical and research practices.

Consequently, moving into new territory brings a new set of challenges. For many of these aforementioned modalities, moving away from a clinical classification to a biomarker one, involves having a single cut point in order to define a positive or negative grouping. Having such stringent criteria does help clearly define AD classification but also hinders a more nuanced approach that has a degree of flexibility in clinic. Clifford Jack and colleagues (2018) suggested looking to the oncology field where they have a grading system of 0-2 for each of their biomarkers allowing for two, rather than a singular, cut points giving a more nuanced view of the disease spectrum in relation to these biomarkers. This has however yet to be widely adopted into routine practice within AD, which currently still leaves the field with inherent sensitivity issues and dichotomised patients in both clinical settings and global trials.

Another key element to this is the consistency and validity of the ligands, assays and other measures being employed to identify these biomarkers. For PET ligand binding there are currently three tracers commercially available

all with differences in how they bind to Aβ and tau (see the imaging and biomarker sections below for full discussion). This inhibits consistency in trials and comparisons across cohorts if different compounds are being used in each one. The difference is not large but as already mentioned, when there is a single cut point to define AD diagnosis, then this can lead to heterogeneity of individuals within a single 'homogenous' group. Further to this, cerebral spinal fluid (CSF) and blood-based assays contain analogous inherent flaws. CSF samples for AD, are predominantly analysed by single-analyte enzyme-linked immunosorbent assay (ELISA), a multiplexing xMAP assay, or an immunoassay with electrochemoluminescence detection. These and other assays that measure biomarker levels, have been found to differ among studies, which could be the result of a number of preanalytical, analytical, or assay-related considerations. On top of this, previous collection methods, particularly for Aβ have been imprecise, as this protein is very sticky and collection and storage procedures need to be maintained and standardised for cross comparisons of samples (Bjerke et al., 2010; Andreasson et al.,2012; Sancesario et al., 2010). The discrepancies seen within laboratories and within assays has led to several efforts to standardise these both across assays and across regions. The largest of these has been undertaken by the Alzheimer's Association (Mattsson, 2011). Nine rounds of testing were undertaken across 84 laboratories. Three samples were sent to each participating laboratory in each round, these samples came from a human CSF pool prepared by a central laboratory who conducted a number of standard procedures during the processes of preparation to ensure the homogeneity of the samples being sent out (Mattsson et al., 2013). The key finding was that across all assays of the key AD proteins ($A\beta_{42}$, P-Tau & T-tau) the coefficient variance for inter-laboratory measurements was between 20-30% and when looking at the inter-run variance of the actual assays the ELISA showed consistently sub 4% across all rounds of testing, whereas xMAP and MSD 6E10 assays had between 2.5-7% and 2-6% variances respectively.

This does not suggest that these measures are not suitable for diagnostic use in AD, it does however show that employing single point cut scores on assays is not a current workable concept when conducting reproducible and global studies, unless a single central laboratory is utilised (as in clinical trials). The intra-run variations shown within the assays is still something that needs to be improved if single point cut scores are to be employed outside clinical trials, with the ELISA being shown to be the currently most consistent and reliable assay for all three AD biomarkers. This new ATN diagnostic criteria has subsequently led to a model of comprising of a number of different combinations of pathology, some related to AD and some which are not (**Table 1.1**).

Table 1.1. Biomarker profiles & categories

AT(N) Profiles	Biomarker Category
$A^-T^-(N)^-$	Normal AD Biomarkers
$A^+T^-(N)^-$	AD pathologic change
$A^+T^+(N)^-$	AD
$A^+T^+(N)^+$	AD
$A^+T^-(N)^+$	AD & concomitant non AD pathology
$A^-T^+(N)^-$	Non AD pathologic change
$A^-T^-(N)^+$	Non AD pathologic change
$A^-T^+(N)^+$	Non AD pathologic change

For diagnosis of AD it is a widely held view that amyloid biomarkers represent the earliest evidence of AD neuropathology in vivo and as such in combination with paired helical filament tau (p-tau) are seen as categorical determinants for the definition of AD pathology (Montine, et al., 2012; Hyman et al., 2012). Furthermore, abnormal amyloid being the earliest pathologic change can be argued to be the defining signature of AD (Jack et al., 2018). As seen in **Table 1.1.**, a positive amyloid biomarker is critical for the categorisation of AD. Without it, any positive biomarker is indicative of suspected non-AD pathologic change and without a positive amyloid AND tau biomarker a positive N biomarker is suggestive of concomitant non-AD pathology. As suggested earlier this methodology is envisioned to define AD more so as a biological construct. One that is permissive to a more etiologically based, biologic characterisation, to explain the pathologic cortical events that lead to cognitive impairment in AD, as well as delivering a model of wider multifactorial aetiology of other possible dementias.

Whilst this new ATN criteria is fundamental to our current biological definition of AD, the marriage to the clinical phenotype requires further work. Frisoni and colleagues (2022) recently published a position paper on how to best combine the two. They suggest that AD be treated as three separate diseases in a probabilistic model of AD. Sporadic AD is split by APOE genotype (carrier of the ε4 allele and those who do not) and the third subtype is autosomal dominant AD. Non APOE- ε4 sporadic AD is driven by varied genetic and environmental factors. Whilst this model shows promise it is yet to be fully accepted within research contexts. The marriage of the clinical phenotype with that of the biological categorisation model still needs work. But it is critical to consider when looking at applicability of results within these frameworks.

ATN & Cognition

Marrying the biological definition and clinical presentation also means having a fine-tuned fundamental understanding of how to measure cognition within the typical clinical presentation of AD. Currently, the relationship with cognitive impairments and ATN biology is equally important. Cognition is

fairly well characterised at the latter stages of AD and has been fundamentally defined across mild, moderate and severe AD, however it is less so at the very earliest stages of the disease and as such is where the main focus of this thesis will be. Cognition is one thing that is highly variable both across the population and has a number of confounding variables that interact to influence an individuals' ability and performance. Thus, in order to accurately incorporate cognition into any model, a fundamental understanding of the reasons behind the clinical and cognitive variability is crucial.

Cognitively impaired, in the context of AD, is at the very first stage of the clinical presentation of the disease (subsequent to preclinical AD) an individual has an impairment on a cognitive measure that is equivalent to or greater than 1-1.5 standard deviations below the appropriate normed population scores. The definition throughout all criteria is not specific to the cognitive domains or measures used. However, for the clinical presentation of typical AD individuals normally present with memory issues, if there is no memory impairment the definition within this spectrum is questionable. As seen with the biomarker categorisation, individuals who fall into some of these categories of cognitive impairment, may not have AD or even concomitant pathologies. Those who present with cognitive impairment to the aforementioned degree but without an amnestic aspect are very much questionable in terms of a diagnosis of AD.

Traditionally cognition in AD has been categorised into three distinct categories, cognitively unimpaired or preclinical AD, mild cognitive impairment (MCI; or prodromal AD) and dementia (mild, moderate or severe AD). This staging is somewhat at odds with the idea of cognition as a continuum with the same drawbacks as the biomarker definitions detailed above. Syndromal categorical staging such as this, is applicable to all members of a cohort as individuals across all biomarker groups will also have a corresponding cognitive profile too. Thus, regardless of the biomarker diagnosis, an individual will also slot into one of these three categories of cognition making it more difficult to differentiate AD diagnosis on cognition alone. Hence within the AD continuum, biomarkers are ubiquitous to the cognitive staging and diagnosis. Outside this however, the cognitive profiles can be somewhat conflated, as in the majority of cases individuals do not have longitudinal cognitive profiles and there are routinely cross-sectional profiles of individuals which may not present a univocal depiction of their cognitive abilities. The only way to circumvent this is to undertake large cohort-based studies at great time and expense which is not routinely feasible in primary care settings. Nevertheless, validated and standardised cognitive instruments employed within cohorts and subsequent studies across AD, do benefit from ample normative data which allows comparison based upon wider population-based data.

Regardless of biomarkers, when looking at the wider definitions within cognitive function and dysfunction, the focus is upon both individual performance and comparison to population normative data (Jack et al., 2018). Individuals who are termed cognitively unimpaired, perform *within* the expected range for that individual based upon all available information, such as prior testing performance if present but can also be based upon clinical judgement as well as, or in place of, cognitive performance. This grouping also allows for individuals performing outside the normal range of population-based norms for a range of tests and furthermore allows for individuals to have subtle serial cognitive deterioration or subjective cognitive impairment on serial cognitive testing. In essence a definition of cognitively unimpaired without AD biomarkers can mean a number of things depending upon previous testing, individual performance level and the test/s being employed to measure the individual's cognition. However, normative data does not always exist for all measures, thus limiting the applicability of this staging approach and again, giving further reason for individual measures having not to have been suggested within this framework.

When it comes to the next grouping, MCI, it becomes more nuanced. Individuals at this stage exhibit impaired cognitive performance *below* an expected range for that individual based upon either prior performance, clinical judgement and/or population-based norms. In addition, a decline in cognition from a baseline assessment must also be present, and can come from the individual or an observer who is able to report on a longitudinal change in the individual. The individual may also present primary symptoms that are non-amnestic, have some neurobehavioural changes and have very mild functional impairment related to the decline in cognitive abilities. These stages allows the incorporation of function, cognitive and neurobehavioural symptoms, however as seen with all three stages in this model, the impairment is very loosely defined without any stringent criteria relating to cognitive impairment/s. However, again amnestic impairment covers a broad range of memory domains and measures, the lack of specific suggestions hampers how this should be interpreted clinically.

The final stage is that of frank dementia, this can be broken down further into mild, moderate or severe dementia depending upon the level of impairment. Individuals at this stage have a substantial cognitive impairment that has progressed and continues to do so, as reported by performance on cognitive measures or reported by the individual or the observer. The level of impairment is also such that the cognitive impairment may be coupled with a prominent neurobehavioural change which has resulted in a significant functional impairment that clearly impacts daily life. With the sub-stage of dementia defined by this level of functional impairment exhibited by the subject. At this stage individuals have typically progressed in the disease and display a variety of impairments in cognition, function and some behavioural deficits, at this stage, cognition and function become more overlapped as

when deficient to a greater degree, the more homogenous these two domains become.

Overall, this model's approach is to simplify and broaden a scope of impairment and how this is indexed for an individual, is somewhat oversimplified, which leads to a lack of clarity at each stage in how to accurately measure the cognitive processes in question. Both in relation to which cognitive domain and which measure/s to use. In order to address this, it is argued to further divide the three stages of cognitively unimpaired, MCI and dementia, to give a more nuanced view of cognition and the decline in each of the domains of cognitive impairment as not all follow the same trajectory, both in AD and outside it. To try and bring the biological definition and cognitive staging together, criteria has been suggested to be combined. Also incorporating key elements of the FDA draft guidance for Early AD (FDA, 2018) is also key to make this more widely accessible and as such allows for potential beneficial implication to patients and carers alike, further along the research and development pipeline. The resulting amalgamation adapted from Jack et al 2018 is displayed in **Table 1.2.**

Table 1.2 Cognitive & Biomarker Profiles within AD Staging

Syndromal Cognitive Stage				
Biomarker Profile		Cognitively Unimpaired	MCI	Dementia
	$A^-T^-(N)^-$	Normal AD biomarkers & CU	Normal AD biomarker with MCI	Other Dementia
	$A^+T^-(N)^-$	Preclinical AD	Prodromal AD	AD Dementia
	$A^+T^+(N)^-$	Preclinical AD	Prodromal AD	AD Dementia
	$A^+T^+(N)^+$	Preclinical AD	Prodromal AD	AD Dementia
	$A^+T^-(N)^+$	AD pathology with concomitant neuropathology	Prodromal AD with concomitant neuropathology	AD Dementia with concomitant neuropathology

Focusing purely upon cognitive profiles of those within the Alzheimer's continuum, the proposed staging of cognition is a clinical one, but it still lacks specificity of cognitive measures (Jack et al., 2018). For someone to be on the Alzheimer's continuum they must at least have a positive amyloid biomarker. Whereas the previous model has three distinct categorisations, by including biomarkers this model gives an enunciated model of AD and cognition comprising of six stages. As this model is key to the core research aims of this project it is described in full below (from Jack et al., 2018).

Numeric clinical staging—Applicable only to individuals in the Alzheimer's continuum

Stage 1

- Performance within expected range on objective cognitive tests. Cognitive test performance may be compared to normative data of the investigators choice, with or without adjustment (the choice of the investigators) for age, sex, education, etc.[1]
- Does not report recent decline in cognition or new onset of neurobehavioral symptoms of concern.
- No evidence of recent cognitive decline or new neurobehavioral symptoms by report of an observer (e.g., study partner) or by longitudinal cognitive testing if available.

Stage 2

- Normal performance within expected range on objective cognitive tests.
- Transitional cognitive decline: Decline in previous level of cognitive function, which may involve any cognitive domain(s).
- May be documented through subjective report of cognitive decline that is of concern to the participant. Represents a change from individual baseline within past 1–3 years, and persistent for at least 6 months. May be corroborated by informant but not required.
- Or may be documented by evidence of subtle decline on longitudinal cognitive testing but not required.
- Or may be documented by both subjective report of decline and objective evidence on longitudinal testing.
- Although cognition is the core feature, mild neurobehavioral changes—for example, changes in mood, anxiety, or motivation—may coexist. In some individuals, the primary compliant may be neurobehavioral rather than cognitive. Neurobehavioral symptoms should have a clearly defined recent onset, which persists and isn't explained by life events.[2]
- No functional impact on daily life activities

Stage 3

- Performance in the impaired/abnormal range on objective cognitive tests.
- Evidence of decline from baseline, documented by the individual's report or by observer (e.g., study partner) report or by change on longitudinal cognitive testing or neurobehavioral behavioural assessments.
- May be characterized by cognitive presentations that are not primarily amnestic.[3]
- Performs daily life activities independently, but cognitive difficulty may result in detectable but mild functional impact on the more complex activities of daily life, that is, may take more time or be less efficient but still can complete, either self-reported or corroborated by a study partner.

Stage 4

Mild dementia

- Substantial progressive cognitive impairment affecting several domains, and/or neurobehavioral disturbance. Documented by the individual's report or by observer (e.g., study partner) report or by change on longitudinal cognitive testing.
- Clearly evident functional impact on daily life, affecting mainly instrumental activities. No longer fully independent/requires occasional assistance with daily life activities.

Stage 5

Moderate dementia

- Progressive cognitive impairment or neurobehavioral changes. Extensive functional impact on daily life with impairment in basic activities. No longer independent and requires frequent assistance with daily life activities.

Stage 6

Severe dementia

- Progressive cognitive impairment or neurobehavioral changes. Clinical interview may not be possible.
- Complete dependency due to severe functional impact on daily life, impairment in basic activities, including basic self-care.

This research framework was deliberately left vague by the authors in order to encourage wider adoption, however it can be argued that there needs to

be a clear agreement to implement the adjustments on cognitive measures where applicable, as it has been widely shown that age, sex, educational level and socio-economical status all impact cognition to varying degrees. Thus, in order for a measure to be utilised across multiple countries and centres it needs to have these adjustments in place to have generalisability to the whole population. If not taken into account these variables will mask any potential subtle, or in the case of education; not so subtle, cognitive impairments that are potentially associated with AD (discussed in detail below).

The key additions are in the three formative stages 1-3; with the latter three stages all comprising of increasing states of dementia, which are well characterised by the symptoms and clinical manifestations individuals at these stages exhibit. It is in these first three stages where further discussion is warranted. An extra 'transitional' stage (ie 2.5) between cognitively unimpaired and MCI gives a greater account for fluctuations in cognition that may be transient and can occur in this model without the subsequent diagnosis of MCI. When looking at the differences between the first two stages (1 & 2), there would be an absence of functional differences which may lead to the suggestion they are not distinguishable, however aside from the clinical manifestations, differences would be measurable in terms of cognitive decline between stage 1 and stage 2. The clinical meaningfulness of this is difficult to establish and the cognitive decline is normal within the normal ranges for the measures being implemented (Aisen, 2018). But again, this could be highly variable depending on which domain and measure is used for measurement of cognition here. The real onset of measurable cognitive decline is by definition subclinical and is currently thought to be present after the gradual amyloidosis over a decade prior to this. However, as Aisen and others have posited, this measurable subclinical decline may precede amyloid positivity (PET or CSF diagnosis) giving a window for the detection of symptoms very early in the disease course to allow for possible early intervention and treatment. Fundamentally what is missing from these descriptions is a measure/s and/or cognitive domains that should be measured in order to marry these descriptions to a clinical presentation across all six stages. Furthermore, subtle impairment requires consistent measurement and a fundamental understanding of how cognition and clinical presentation marries to that of pathological biomarkers. Something that is yet to fully achieved and is a key aspect of what this thesis hopes to achieve.

The overarching body of work for classification of AD, using the ATN biological framework and the clinical presentation of AD can best be seen through a graph of trajectories over the disease course. These are iteration of the graphs first derived by Clifford Jack and colleagues (2013). This most comprehensive current version is shown in **Figure 1.2.** taken from work by Palmqvist and colleagues (2019) and Hansson (2021). The composition of these staging criteria results in a disease trajectory best outlined by looking

at discrete pathological markers as described above. The graph also includes shaded areas which represent current knowledge of progressive amyloid and tau pathology. To note the final section pertaining to symptomatic AD covers the disease once symptoms are present up to death. As discussed previously, the nuance of impairments seen within cognitive and functional domains is ambiguous hampering the confluence between the clinical presentation of the disease and the pathological processes undertaken. However, this graph does allow for some broader relationships to be inferred between global cognition and Tau PET and MRI measured atrophy. The cognitive picture is far more nuanced than this, all of which is described in more detail further below.

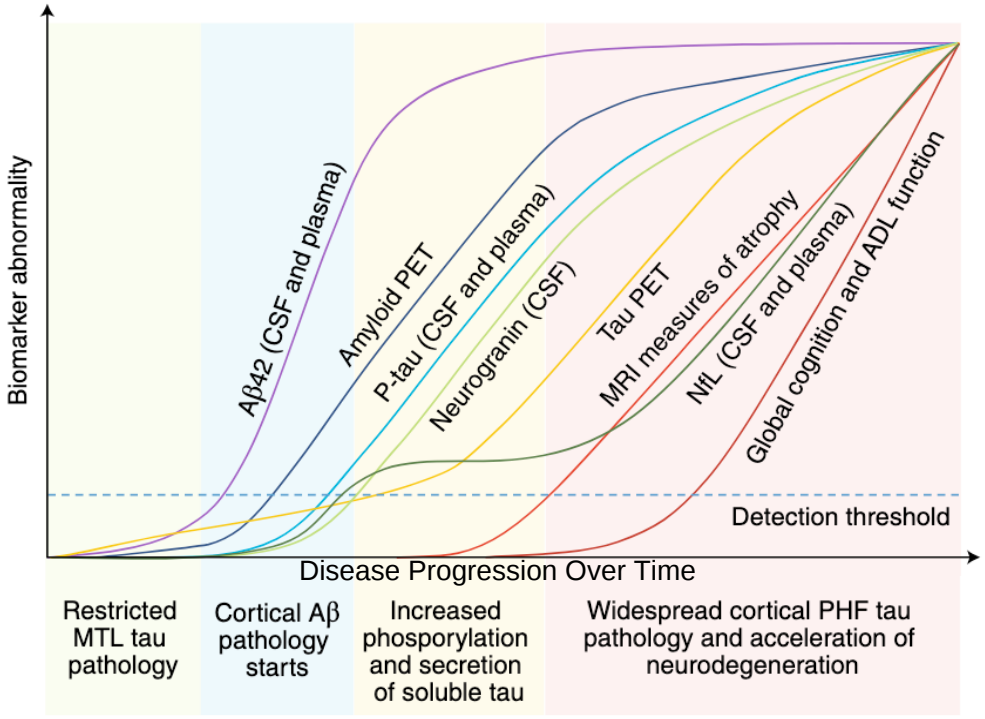


Figure 1.2. The trajectories of different fluid and imaging biomarkers in the AD continuum. Green, blue and yellow shaded areas depict pathology prior to symptom onset, the red shaded area depicts MCI through to severe AD (Stages 1-6). Graph adapted from Hansson, 2021 and based upon other prior studies (Palmqvist et al., 2019; Janelidze et al., 2020a&b; Barthelemy et al., 2020; Mattsson-Calgren et al., 2020).

The Amyloid Hypothesis & Disease Biomarkers

These measures and the classification methodologies described above are all founded upon the amyloid hypothesis of AD. This has been the cornerstone driving research and drug development alike for the past three decades. For over a century it has been well accepted that a progressive build-up of disparate forms of amyloid protein, in an array of organs, has been a causal factor for many devastating diseases. However, it was only in 1984 that the idea of amyloid-beta may have a central role in AD was put forward by George Glenner (Glenner & Wong, 1984). Over the past two decades this has been the cornerstone of AD research and has been dubbed the amyloid hypothesis of Alzheimer's Disease. In incipient sporadic cases it is proposed that amyloid beta monomers gradually begin to build up, slowly clumping together to form insoluble oligomers/fibrils then latterly larger amyloid plaques (Beyreuther & Masters, 1991; Hardy & Allsop, 1991; Selkoe, 1991; Hardy & Higgins, 1992). Despite recent conjecture this is the dominant model of AD pathogenesis and has been the lynchpin that has guided drug development over this

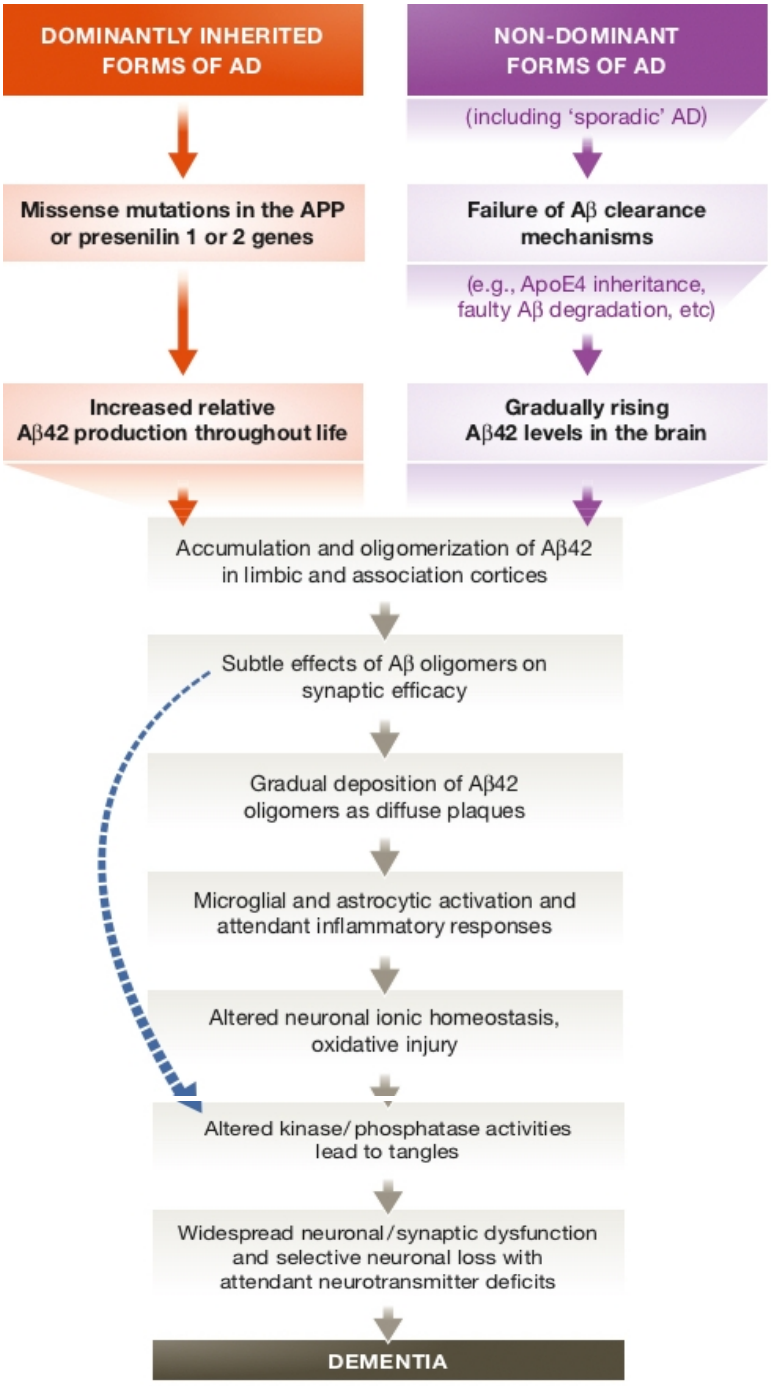


Figure 1.3. The cascade of pathological events that lead to AD proposed by the amyloid cascade hypothesis

time period (Cummings et al., 2018; Selkoe & Hardy, 2016). There are many reasons why this has been so revered and unwavering as a theory, however there are still many unknowns and unanswered questions, all of which will be discussed further in **Appendix 2**. When looking at non-dominant forms of AD, most common of which is 'sporadic' AD, there are a cascade of events that eventually lead to dementia. This amyloid cascade hypothesis has been developed consistently since its inception and the current consensus is shown below in **Figure 1.3** (Selkoe & Hardy, 2016).

In the normal processing of the amyloid precursor protein (APP), it is cleaved close to the membrane by an extracellular protease known as α-secretase. This liberates a soluble extracellular fragment, sAPPα. Alternatively, APP is cleaved by an aspartyl protease referred to as β-secretase (or β-site APP cleaving enzyme 1, BACE1) generating a soluble extracellular fragment (sAPPβ) and a cell-membrane-bound fragment (C99). C99 is cleaved within the membrane by an enzymatic complex formed of four proteins (presenilin, nicastrin, anterior pharynx-defective 1 and presenilin enhancer 2), known as γ-secretase. Presenilin is the catalytic subunit of γ-secretase and is encoded by either the PSEN1 or PSEN2 gene, these genes have deterministic importance to progression rates of AD (Suzuki et al., 1994; Duff et al., 1996; Scheuner et al., 1996). The γ-secretase cleavage releases an intracellular peptide known as amyloid intracellular domain (AICD) and the Aβ peptide. Aβ has many different lengths, the most abundant being of 40 amino acids and the less soluble of 42 amino acids. Aβ aggregates to form oligomers (two or more proteins), protofibrils (singular groups of proteins), fibrils (larger groups of the same protein) and ultimately plaques (sticky abundance of proteins), which are fundamental hallmarks of AD pathology.

Autosomal Dominant AD

Key findings that have helped to shape and reinforce the amyloid hypothesis come from further study of the genetically inherited form of the disease, autosomal dominant AD (ADAD) otherwise known as dominantly inherited AD (DIAD) (**Figure 1.3**). This genetic mutation accounts for approximately 1% of all AD cases (Bateman et al., 2012). Mutations in one of three key genes (APP, PSEN1 and PSEN2) involved in the processing of Aβ have been shown to lead to early onset AD with complete penetrance (Scheuner et al., 1996). These mutations to APP and its processing leads to an elevation of $A\beta_{42}$ & $A\beta_{43}$ peptides and begin the aforementioned cascade of pathology.

Importantly, these genetic mutations tend to have a predictable age of onset (Lopera et al., 1997), which are dependent upon mutation type and background family genetics, thus making this form of AD ideal for studying the initial pathogenesis of the disease (Wijsman et al., 2011). Within these mutations there is also increasing evidence to show an overlap with sporadic forms of AD (Cruchaga et al 2012). However, these mutations are incredibly rare and as such require increased awareness to locate these individuals,

many of them are in small community pockets due to the nature of the genetic underpinnings of this variant. The worldwide initiative to study this form of AD; DIAN and DIAN-TU (Dominantly inherited Alzheimer's Network/Trials Unit) has made great strides in the effort to learn more about the disease and treat individuals who are at risk of developing the disease due to their known genetics and family history. These trials are still ongoing but it is hoped that the currently available therapeutics will arrest any decline in cognition and function, or stop it from occurring altogether (Bateman et al., 2017). Batemans group published some formative results from this initiative which has reinforced the pathogenesis of AD. They found ADAD was associated with a number of pathological changes over decades, these results showed abnormal CSF biomarkers of AD; brain amyloid deposition and brain metabolism as well as significant progressive cognitive and functional impairment akin to that seen in sporadic AD (Bateman et al., 2012). The homogenous nature of these findings with those seen in late onset/sporadic AD, led further credence to the early pathogenesis of AD occurring 15-20 years prior to that or the onset of clinical symptoms.

Tau Pathology

Conversely to the focus placed upon amyloid within the disease, AD is also a tauopathy. This pathology has been conclusively shown to occur after the cascade of effects from amyloid deposition and the innate immune response within the cortex, with the hallmarks of this stage of pathology coming in the form of neurofibrillary tangles (NFT) (Felsky et al., 2019). This target is still yet to be fully probed, manipulated and results produced within a full clinical trial or research program. Nevertheless a growing body of evidence points towards a promising concept for the amelioration of AD pathology and symptoms when this target is engaged (Small & Duff, 2008; Karran & De Strooper, 2016; Herrup, 2015; Cummings et al., 2018; Yanamandra et al., 2013).

Tau's Function

Two key functions of tau are its ability to promote assembly and to preserve the structure of microtubules (Weingarten et al., 1975). Phosphorylation of tau is also critical and key for neurite outgrowth and axonal transport mechanisms, but it is clear this becomes aberrant during AD (Kowall & Kosik, 1987). This destabilisation and lack of neuronal growth/maturation is another downstream pathological effect, but is commonly thought to be key to the continued decline in an individual (DeKosky et al., 1990). Within AD the six isoforms of tau known to exist, have all been shown to be hyperphosphorylated and aggregated into paired helical filaments (PHF) (Grundke-Iqbal et al., 1986a; Grundke-Iqbal et al., 1986b; Iqbal et al., 1989; Iqbal et al., 1986; Lee et al., 1991; Goedert et al., 1992). These groupings of tau are according to their tubulin-binding domains as 3-repeat (3R) and 4-repeat (4R) tau proteins. In the normal brain and within PHF-tau equal amounts of each exist but changes in the 3R and 4R tau ratio can cause

abnormal tau accumulation (Harada et al., 2016). Whilst key conformational and truncational changes (Jicha et al., 1997; Jicha et al., 1999a; Jicha et al., 1999b; Novak et al., 1991; Gamblin et al., 2003; Cotman et al., 2005) have been observed to this protein, post hyperphosphorylation (Delobel et al., 2008), it can be argued that the strongest rationale for the dysfunction of tau in AD, is the distinct abnormal hyperphosphorylation of the protein itself (Grunke-Iqbal et al., 1986b; Alonso et al., 1994; Iqbal et al., 1986). As this toxicity of phosphorylated tau appears to be solely due to its abnormal hyperphosphorylation as when these proteins are dephosphorylated the diseased tau converts it into a normal version of the tau protein and behaves as such (Alonso et al., 1994; Li et al., 2007; Wang et al., 1995; Wang et al., 1996).

Within the AD cortex, tau is recovered in three forms, soluble, oligomeric and fibrils (Kopke et al., 1993; Iqbal et al., 1986; Bancher et al., 1989). Whereas in a non pathological cortex almost all forms of tau are soluble in nature. Normal tau is at comparative levels in healthy elderly and those with AD, however, levels of total tau (t-tau) are between four and eight times greater in those with AD due to the presence of these oligomeric and fibrillar conformations (Khatoon et al., 1992). Of this increase around 40% is oligomeric and sedimentary, with these oligomers being comprised of both hyperphosphorylated and non hyperphosphorylated tau (Kopke et al., 1993; Iqbal et al., 1986). Thus, showing the basis for the increase in this protein in the CSF seen within AD. However, the tau in NFT has been shown to be inert, with around 40% of the abnormally hyperphosphorylated tau in AD brain actually present in the cytosol and not polymerised into larger conformations such as NFT & PHF (Kopke et al., 1993; Iqbal et al., 1986; Bancher et al., 1989). This abnormally hyperphosphorylated tau (p-tau) inhibits assembly and disrupts microtubules (the polar opposite of its normal behaviour) (Alonso et al., 2004; Li et al., 2007; Wang et al., 1995) and this toxic behaviour also critically involves evoking normal tau in this process (Iqbal et al., 2010; Alonso et al., 2004; Alonso et al., 1996), as well as two other microtubule associated proteins (MAP) $MAP1_{a/b}$ & MAP2 (Alonso et al., 1997). This seconding of healthy tau by the neurotoxic p-tau, suggests tau is actually a key factor much earlier in the disease course of AD. Whilst not forming in large deposits in the way amyloid conglomerates, tau could be argued to spread in a smaller less noticeable way, with further research using newly validated imaging ligands key to this endeavour.

Measurement of Tau in vivo

CSF measurement of p-tau and t-tau has been the only available biomarker in this regard, but it is difficult to relate this directly to the pathology cascade within AD. As tau in many forms have been shown to be present in many other neurodegenerative diseases, this may be indexing other forms of neurodegeneration unrelated to any AD pathology (defined as NFT & PHF) and without the topography may relate to comorbid cell death or synaptic

dysfunction. More reliable and attributable to the T criteria (Jack et al., 2018) is p-tau which underpins the NFT & PHF formations in the cortex (Kopke et al., 1993; Iqbal et al., 1986; Bancher et al., 1989). However, with around 40% of this p-tau inert and not polymerised into these formations it is more difficult to directly correlate this p-tau in CSF measures to that of tangle load in vivo. The inception of the tau PET ligands, of which there are now a number of validated entities to choose from, have shown in numerous validation studies to provide a higher inherent specificity and sensitivity akin to the in vivo tau load (Brier et al., 2016; Josephs et al., 2016; Choi et al., 2018; Smith et al., 2018). As with Aβ, autopsy studies have shown that tau pathology accumulates in a distinct spatial topography throughout the disease (Braak & Braak, 1991; Thal et al., 2002). These have been previously constrained to histopathological studies owing to the lack of ligands for measuring tau pathology in vivo via PET.

As a result of existing amyloid biomarkers available in vivo, the distinct topography and its relationship with cognition has been well defined (Villian et al., 2012). Tau deposition is hypothesised to more closely correlate with cognitive decline than compared directly to Aβ load (Jack et al., 2018; Hansson, 2021; **Figure 1.2**) and as such engaging this target may improve cognitive symptoms. Target engagement with compounds affecting this protein's conformations are yet to produce any results from clinical trials so it is still unknown whether altering these conformations elucidates any amelioration or cessation to pathological, functional or cognitive decline. But given the evidence described above and the specific uniformity of decline with a number of global cognitive processes, this indicates some change should be expected.

Genetic Risk Factors & APOE

One of the major breakthroughs of the last few decades was the finding of the risk variant on the Apolipoprotein E (ApoE) phenotype with carriers of the ε4 allele having an increased risk of developing incipient AD earlier and having a sharper decline once the disease takes hold (Corder et al., 1993). This is the strongest and most common genetic risk factor for late onset AD (Corder et al., 1993; Bu, 2009; Huang et al., 2012). The human form of the ApoE gene exists in three polymorphic alleles; ε2, ε3 and ε4, which have been shown to have an approximate worldwide incidence of 8.4%, 77.9% and 13.7%, respectively (Farrer et al., 1997). However, this dramatically increases to ~40% with the ε4 allele in patients with AD, suggesting a major role in the disease process. ApoE's normal behaviour is as a major cholesterol carrier, regulating lipid homeostasis (Mahley & Rall, 2000), supporting lipid transport as well as membrane repair and synaptic plasticity within the cortex (Slezak & Pfrieger, 2003). Outside AD and the CNS $ApoE_{\varepsilon 4}$ is implicated in hyperlipidaemia and hypercholesterolemia (both leading to atherosclerosis), coronary heart disease and stroke (Mahley & Rall, 2000; Lahoz et al., 2001). The ε4 allele has also been shown to interact with cerebrovascular disease to impede clearance mechanisms which may be a

key exacerbating factor to the progression of the disease (Veitch et al., 2019). These implications are still relevant to AD as any disruption of blood flow will have downstream effects. The relationship between the dysfunctional variant and AD pathogenesis can be argued to impact homeostasis as well as increased dysfunction of synaptic plasticity, which is known to have downstream cognitive and function effects (Huang et al., 2010; Kanekiyo et al., 2014; Heneka et al., 2015). This major genetic phenotype is something to be cognisant of in any analysis as this significant variation between individuals can have strong implications for any subsequent findings.

Genome Wide Association Studies (GWAS) have also played a key part in uncovering new potential avenues of research and hypothesis of inception and dysfunction within AD. They have identified polymorphisms in or near several genes that are associated with AD risk (Harold et al., 2009; Naj et al., 2011; Hollingworth et al., 2011; Bertram et al., 2008). A meta analysis of GWAS in AD also produced a further 12 loci to investigate (Lambert et al., 2013). However, none have the sizable impact that APOE does on the disease course (Karch & Goate, 2015) however, there are still some key genetic variants that have been uncovered. The R47H variant of TREM2 has been shown to triple the risk of AD in GWAS studies (Guerreiro et al., 2013; Jonsson et al., 2013; Song et al., 2018), as this variant of TREM2 impairs the interaction between neurons and Aβ plaques (Song et al., 2018). The identification of novel genetic loci affecting sporadic AD risk is critical to the understanding of the underlying aetiology of AD. The identification of common variants that have small effects on AD risk is crucial as it creates a wider picture of the pathological processes that contribute and are involved in the disease. These variants identified through GWAS are in genes involved in lipid metabolism, the inflammatory response, and endocytosis all which are outside the main proteinopathies within AD.

Neurogranin

Along with CSF markers further compounds have been elucidated more recently with the advancement of proteomics methodologies and assays, one of these is neurogranin. This postsynaptic protein is known to be involved in regulation of calmodulin after neuronal excitation (Baudier et al., 1991; Diez-Guerra et al., 2010) as well as long term potentiation and related cognitive functions (Wu et al., 2002; Huang et al., 2004; Mons et al., 2001). Levels of neurogranin are predominantly measured through a CSF assay and within normal aging levels of the protein are highest in cortical areas associated with its primary function (Bogdanovic et al., 2002).

Synaptic dysfunction is widely believed to be intrinsic to cognitive dysfunction and has been widely thought of as a central pathological mechanism within AD (DeKosky et al., 1990). This synaptic dysfunction is believed to occur before neuronal degeneration and death and as such an

indicator to this is highly valuable to tracking progression (Dekosky et al., 1990; Davies et al., 1987; Bertoni-Freddari et al., 1997). Within AD, synaptic loss is also seen to be more strongly correlated with cognitive decline than either amyloid or tau pathology load (Masliah et al., 2001; Scheff et al., 2007; Sze et al., 1997). Because of the relationship between neurogranin function and synaptic function, it has been shown that neurogranin has normal age related decreases in levels across several brain regions including the hippocampus which is known to be involved in the early pathology of AD (Mons et al., 2001). As such, decreases in synapse numbers has been shown to correlate with worsening memory impairment within AD (Heinonen et al., 1995; Scheff et al., 2007). Neurogranin levels have also been shown to be lower within AD models in the hippocampus and frontal cortex (Davidsson et al., 1998; Reddy et al., 2005) and neurogranin levels have also recently been shown to be impaired in those with AD as well as individuals with MCI (Kvarstberg et al., 2015). This study also showed the levels of neurogranin correlate with cognitive decline in prodromal AD. Results also indicated that neurogranin predicts conversion from MCI to AD, with higher neurogranin levels predicting a faster rate of cognitive decline within confirmed amyloid-positive prodromal AD individuals. This was further replicated in other cohorts suggesting its additional suitability to CSF biomarker panels, in particular to indicate future cognitive decline (Hampel et al., 2018; Tarawneh et al., 2016; Lista & Hampel, 2017; Portelius et al., 2015). Further study into a longitudinal analysis of this protein is still needed to fully understand its interplay with the main pathological process of AD, but as such is included in the (N) criteria of the recent research guidelines (Jack et al., 2018) showing its suitability as a valuable biomarker of synaptic dysfunction/neurodegeneration.

Neurofilament light chain (NFL)

Whilst tau is an inherent measure of microtubule stability, wider axonal measures can also be gleaned from CSF (and soon plasma too). One such protein is neurofilament light chain (NFL), which is a structural component of the neural cytoskeleton, suggesting increased levels of this protein correspond to axonal degradation (Lee et al., 1993). This has been shown to have raised CSF levels within AD corresponding to white matter lesions/axonal degradation (Sjogren et al., 2001), with this finding again being replicated in subsequent cohort studies such as ADNI (Zetterberg et al., 2016). NFL levels are increased at the clinical stage of AD and are related to the cognitive deterioration and structural brain changes at this stage of the disease. This finding supports the assertation that degeneration of large-calibre axons is an important feature of AD neurodegeneration.

One key difference with NFL is that is can be accurately measured in the blood too, with the supposed holy grail of a blood based test for AD still a way off, this is one measure that shows increased values within plasma, corresponding with increases in CSF values as well ($\rho = 0.59$; Mattsson et al.,

2017). Mean plasma NFL values were shown to be increased in patients with MCI (42.8 ng/L) and AD (51.0 ng/L) compared with controls (34.7 ng/L) and had high diagnostic accuracy comparable to established CSF biomarkers. However, these small discrepancies between groups suggest further validation is needed in larger samples to explore the levels of variations across the disease spectrum. Overall, whereas neurogranin can be utilised early in the disease course, NFL is thought to be more accurate within the latter stages of AD, although in depth longitudinal analysis is still needed to fully elucidate the interplay between NFL and disease progression across the spectrum of AD.

Treatments: Memantine & Acetylcholinesterase Inhibitors

The last truly new molecular entity for AD was approved by the Food & Drug Administration (FDA) in 2003 and the European Medicines Agency in 2002 (FDA, 2003; EMEA, 2004). Memantine was licenced shortly after the advent of three somewhat diverse acetylcholinesterase inhibitors (AChEI) and was targeted at unresponsive individuals who were in the moderate to severe stages of AD. Memantine is a symptomatic treatment which does not alter the disease course in any way but may provide some improvement in some of the common symptoms exhibited by individuals at the latter stages of the disease (McShane et al., 2006). These include memory, attention and some aspects of language. Then the main efficacy endpoint for approval in AD is usually the ADAS-Cog (Rosen et al., 1984), however in the more severe spectrum, in which Memantine is approved in, the Severe Impairment Battery (SIB) was utilised (Schmitt, 1997). The SIB indexes seven domains and is scored between 0 and 100, when looking at pooled trial data Memantine shows a significant improvement of 2.97 points at six months (McShane et al., 2006).

Memantine is an NMDA receptor antagonist and has low affinity with glutamate receptors. The primary focus of symptomatic treatment in AD was the augmentation of cholinergic transmission, which this does not alter. Nevertheless, on top of Memantine, other possible therapeutic approaches have been hypothesised that are based upon neurotransmitter enhancement or modulation, these include serotoninergic, noradrenergic substances or neuropeptides and compounds acting on excitatory amino acid receptors, such as for glutamic acid (Emre & Qizilbash, 2001). The hypothesis that this mechanism of action might prove beneficial in AD is due to the abundance of L-glutamate in the CNS and it having been implicated in long term potentiation (LTP), learning and memory functions as well as neuronal plasticity (Sucher, 1996; Riedel et al., 2003; Thomas & Grossberg, 2009).

In the Cochrane review of both published and unpublished trial data (McShane et al., 2006) they concluded that whilst Memantine was well tolerated, it showed no significant clinical benefit in mild to moderate AD but pooled data in moderate to severe AD indicate a beneficial effect both on

cognition and activities of daily living, corroborating a manifestation in the clinical impression of a change on a subject level. However, this improvement whilst clinically detectable was small in effect (0.28 points on the seven-point CIBIC). Improvement in symptomatology was only present for a short period before a continuing cognitive and functional decline in line with the disease course. Concurrently to this, combination therapy with AChEI's has yielded conflicting results with some authors suggesting a significant benefit (Thomas & Grossberg, 2009). Whereas a more rounded review looking at the all the available data at the time showed that whilst there may be a small benefit at six months, there is no benefit to function nor any improvement in clinical impression of disease, which could be argued to be somewhat subjective anyway and is dependent upon which studies (the inclusion of the extended release trials) are included in the analysis (Farrimond et al., 2012). With a lack of clinical efficacy in the mild to moderate spectrum of AD (Farrimond et al., 2011; McShane et al., 2006), Memantine is only approved at the moderate to severe stages of AD (FDA, 2003; EMEA, 2004).

Prior to Memantine, clinicians can prescribe one of three AChEI's; Donepezil, Rivastigmine and Galantamine (all of which are now off patent). These all work in a very different way to Memantine and enhance the amount of time acetylcholine is present in the synapse by blocking a protease that degrades this neurotransmitter. Loss of cholinergic function and neurons in the forebrain has been highly associated with sporadic and incipient AD (Francis et al., 1999), therefore anything to alter this pathway should have therapeutic benefits. This class of compounds have been shown to improve memory function both in murine models, clinical trials and in clinic. Donepezil, or Aricept as it was branded, is the most commonly prescribed out of the three (≈60% in EMEA, Pariente, 2008), this was in part down to it being first to market but also the compound that showed the best side effect profile in clinical AD (Birks & Flicker, 2006; Birks & Harvey, 2018). Donepezil has consistently shown a highly favourable side effect profile, with its preferential binding ratio to acetylcholinesterase over butylcholinesterase, in comparison to the other two compounds (Rivastigmine and Galantamine). Resulting in significantly less adverse events (AEs) of dyspepsia, nausea, vomiting and diarrhoea (Rogers et al., 1998; Schiender & Farlow, 1995). Nevertheless, whilst the symptomatic benefits of treatment with Donepezil and other compounds in this class are clear, they are deemed to lack interaction with the neuropathology of AD (Ellul et al., 2007). Further to this, these compounds have been shown to only have transient effects on the symptomatology of AD which slows the cognitive and functional decline for a short period before continuing on its neurodegenerative course, this was one of the key rationales for continued drug development in AD and one key findings of the Cochrane review of AChEI's was that after such a period of around 6-12 months these four drugs yielded little clinical effect (Birks, 2006).

Another Cochrane review comparing the three compounds revealed that all three are efficacious in mild to moderate AD and whilst they all have slightly different mechanisms of action there is no evidence to suggest any singular compound is more efficacious than the other (Birks, 2006). However, when compared directly, Donepezil had a more favourable side effect profile at this stage of the disease. As it is a smaller single tablet of delivery, at more progressed stages of AD this carries significant benefits. This is due to Rivastigmine and Galantamine both requiring multiple tablets at higher dosages, which has been argued to be part of the discrepancy in side effect profiles. Whilst there are no differences in efficacy, these extraneous differences have led to the prescribing divergences as mentioned above.

Nevertheless, in the only comparative clinical trial that has been conducted between these compounds, Donepezil and Rivastigmine performed comparably on measures of cognition and behaviour, however Rivastigmine was argued to provide greater benefit in activities of daily living and global functioning as these measures were statistically significantly different between these two treatment groups (Bullock et al., 2005). Overall, there is strong evidence that acetylcholinesterase inhibition may offer continued therapeutic benefit for up to two years in patients with moderate AD and Donepezil still continues to be the most widely prescribed and researched of the three.

Treatments: Current Issues & Future Directions

Nevertheless, the fundamental issue with Memantine and the three AChEI's, is that they do not interact with the pathology of the disease. And with it being over fifteen years since there have been any new treatment options, there is a huge unmet clinical need for disease modifying compounds. This is not through a lack of effort or investment on the part of the pharmaceutical industry, but through a lack of understanding of the pathological process at the very earliest stages of AD, the insensitivity of the primary efficacy endpoints, poor population selection and in a number of cases poor target engagement.

As has been shown in both BACEi and antibody trials a significant reduction in $A\beta_{42}$ measured via CSF and PET imaging has not yet resulted in a consistent improvement or halting of decline in cognition (see **Appendix 1** for full details). The compounds are clearly interacting with the pathogenesis of the disease and altering and reducing the production of $A\beta$. In some cases, reducing the amyloid burden to a negative PET read on the individual's brain scans. However, the results are mixed with only two antibodies showing positive results in two separate trials. Many arguments have been made as to why this might be. It has been suggested that the previous trials have contained a substantially heterogeneous population, this was found to be

such an issue that over 26% of individuals in one of the trials were found to be amyloid negative on PET scans (Voss et al., 2016). When trying to measure a drug that alters the disease pathology, having over a quarter of your individuals without the specified disease is cataclysmically destructive for measuring efficacy. Something that is only compounded when it is in a global phase III trial. Along with MMSE ranges within the tested domain (MCI, Mild, Moderate or Severe), cognitive impairment to either a clinical or functional measureable degree is a key criterion to inclusion in these trials. It has only been within the last few years that trials now contain amyloid positivity as one of their inclusion criteria on any study within the AD field. This prior heterogeneity could have meant that comorbidities played a significant part in these trials and they were not measuring AD pathology but another form of dementia or cognitive impairment that was present, transient or not there altogether. The latter explanation of the absence of cognitive impairment can be argued to be down to another suggested causality for the trial failures; a lack of sensitivity and/or reliability of the chosen instrument for measuring the efficacy of these compounds.

One further criticism that is present in nearly all drug studies that fail to meet their primary efficacy endpoints is that of a lack of target engagement. This is either due to suboptimal dosing, in neurological conditions lack of brain penetration, poor interaction with the targeted molecules or receptor types or the drug isn't exposed enough to the target to make any interaction with the target site and is either metabolised or doesn't reach there at all in some cases. However, in both the studies of the BACEi's and the antibodies treatments for Aβ clear target engagement was seen with highly significant reductions across the board in all of these trials when looking at CSF Aβ decreases between the treatment groups and the placebo (Cummings et al., 2018; van Dyck, 2018). These reductions are only possible if the compounds in question are brain penetrant (an issue for the larger antibody molecules and not the inhibitors) so it is easy to rule out a lack of target engagement, suboptimal dosing and exposure durations for the absence of efficacy. As if these things were problematic the compounds would not reduce Aβ levels to any significant degree at all.

Further to the heterogeneity of the prior cohorts, due to disease incongruence, the rates of clinical decline in AD have been shown to be highly variable due to plethora of different environmental and genetic factors (Karch & Goate, 2015; Corder et al., 1993;
Saunders et al., 1993; Stern et al., 1994; O'Donoghue et al., 2018). The new ATN criteria (Jack et al., 2018) as well as the further disease staging suggestions by Frisoni and colleagues (2022) further emphasise these new suggestions of heterogeneity. If a subject is amyloid positive (A^+) then knowing where they are in the other two biomarkers can give a much clearer idea of the rate at which they are likely to decline. Combined with the insensitivity of the efficacy endpoints, the other main argument rests on the

fact that the disease begins 15-20 years prior to our current understanding of symptom onset, by which point it is hypothesised that the biological damage of Aβ has already peaked (Jack et al, 2018). As a result of this programs are now looking earlier, at the preclinical stage of the disease. Whereby it is thought the disease trajectories can be affected to a significant extent by amyloid altering compounds. However, this is not as simple as it appears to be, finding people who it is currently thought have none or minor symptoms without an available treatment to offer them is incredibly difficult.

Overall, a number of clear issues exist with current clinical trial design. This thesis will look to answer some of these. Primarily through providing a greater understanding of the measures used to index cognitive impairment within AD, but also by looking to bridge the gap between the clinical and biological presentations of AD.

Neuroimaging

As seen with the complicated heterogeneous nature of clinical trials in AD, pathologic confirmation is key for both target engagement and tracking of disease progression. Thus neuroimaging has been at the forefront of the advancements in research and clinical trial methodologies within AD over the past decade (Jagust et al., 2018). The inception of amyloid and tau PET ligands has allowed us to accurately phenotype individuals. MRI's are also being used to rule out pathologic comorbidities that may cloud any treatment effects as well as more recently to show early indicators of functional network changes. This section will explore the key instruments and methods that are utilised in clinic and research within AD.

MRI

Magnetic Resonance Imaging (MRI) has long been used to uncover brain abnormalities and within AD this is no different. MRI is unable to distinguish the amyloid or tau protein build ups but key indicators such as hippocampal volume, whole brain atrophy and cortical thickness can be gleaned from these scans. With MRI being the primary imaging modality for many years a significant amount of research has been conducted in understanding what MRI scans can indicate within AD. The primary focus of this has been upon hippocampal volume and brain atrophy rates (Jack et al., 1999). However, studies of AD patients have shown brain metabolism (Ossenkoppele et al. 2012), neuronal activation (Golby et al. 2005), resting-state functional connectivity (Nuttall et al. 2016) and overall brain structure (Jin et al. 2017) all deteriorate throughout the AD spectrum.

Numerous studies have shown smaller hippocampal volume within individuals diagnosed with AD compared with healthy controls (Shi et al., 2009; for review). However, at earlier stages of the disease the evidence is less well-formed, with conflicting evidence suggesting atrophy in prodromal AD/MCI due to AD is more variable between individuals. This may be however due to the absence of confirmed pathology in some of these individuals (as reported in some trials up to 30% of individuals were amyloid negative [Siemers et al., 2016]), due to the lack of validity of diagnostic measures available at the time. Nevertheless, hippocampal atrophy is thought to be a significant contributing factor to the cognitive and functional deficits exhibited by individuals across the disease spectrum (Shi et al., 2009). As such, MRI measures of atrophy were treated as surrogate markers for disease progression and showed that these rates of atrophy influenced the rate of cognitive and functional decline within MCI and AD groups (Jack et al., 2005; Shi et al., 2009; Jack et al., 1999). A meta-analysis of 42 studies across MCI and AD showed a progressive atrophy between these classifications with effect sizes of d=1-1.7 and the left and right hippocampi showed asymmetry that maintained throughout this progression, something that is argued to be a characteristic that could suggest the onset of the

illness (Shi et al., 2009). Across these studies, as with many MRI studies, there was a significant amount of heterogeneity; the strength of the MRI machine (1.5-4T), ApoE status, delineation methods of the hippocampi and acquisition protocols. These are the inherent difficulties when comparing imaging results from this method, hence coherent protocols across imaging centres are created for such standardisation purposes, such as those in ADNI (Alzheimer's Disease Neuroimaging Initiative) (Kruggel et al., 2010; Jack et al., 2008). Due to this co-operation and pooling of data and resources between centres such as these initiatives, collaborative cohorts such as these have become bastions for novel findings over the last decade.

The Default Mode Network (DMN) & fMRI

Additionally, whilst volumetric MRI is focused upon structural AD related pathology, its benefits can also be utilised by using a time course analysis to measure the function of specific regions and areas of the cortex. Functional MRI (fMRI) procedures have suggested a number of difference deficits in AD, with one of the most promising is the early association with dysfunction of the default mode network (DMN) at the very earliest stages of AD (Villeneuve et al., 2015; Palmvqist et al., 2017; Hahn et al., 2019). This area is a large-scale connected group of brain regions, primarily composed of the medial prefrontal cortex, posterior cingulate cortex/precuneus and angular gyrus. In order to understand the very beginning of the potential cascade it is therefore key to fully understand non pathological processes in healthy elderly individuals. To this extent several studies have found associations between increases in amyloid load and *increases* in brain connectivity (Sperling et al., 2003; Mormino et al., 2011; Lim et al., 2014). This has been argued to be possibly due to early compensatory mechanisms within the brain trying to adapt to this increase in pathology.

It is still unknown where the pathological cascade begins within the cortex, however further investigations into non demented individuals have shown that regions most prone to the earliest accumulation of Aβ are some of the areas involved in the DMN (Villeneuve et al. 2015; Gonneaud et al. 2017; Palmqvist et al. 2017). With abnormal levels of CSF $A\beta_{42}$ being detected prior to a positive PET signature (Bateman et al., 2012; Blennow et al., 2012; Fagan et al., 2009; Mattsson et al., 2014; Morris et al., 2010) in preclinical AD, these methods were used to study this in the earliest preclinical AD subject from two existing cohorts; BIOFINDER & ADNI. Palmqvist and colleagues (2017) showed that Aβ accumulation preferentially starts in the precuneus, medial orbitofrontal and posterior cingulate cortices, which are several of the core regions of the DMN. This suggests that Aβ starts to accumulate predominantly within parts of the DMN in preclinical AD which also effect brain connectivity at this early stage. They also indicated that the earliest Aβ accumulation is linked with hypoconnectivity within the DMN and between the DMN and frontoparietal network, however not with overall brain atrophy or glucose hypometabolism. Subsequent research from the same

group has further explored this, with amyloid uptake in these regions with resting-state fMRI functional connectivity in a large sample of non-demented elderly individuals (Hahn et al., 2019). Notably, the sample was comprised of only PET negative individuals (indexed by visual read), without clinically relevant global amyloid deposition. The findings were also independent of known confounding factors such as age, sex, ApoE status, presence of SCD and gray and/or white matter structural alterations. This further suggests that the very earliest Aβ accumulation significantly affects brain function. These findings have also been replicated in autosomal dominant AD, again strongly implicating the medial parietal cortex in early Aβ deposition as the primary measurable biochemical event in the development of AD (Gordon et al., 2018; McDade et al., 2018).

Overall, this altered functional connectivity can be strongly argued to represent a compensatory mechanism within the cortex as an inherent response in order to maintain cognitive function, despite the increasing presence of amyloid. All of this evidence indicates the need for future studies to determine whether changes in functional connectivity, may be clinically relevant to predicting individual cognitive and functional decline, as well as overarching disease progression.

PET

Positron Emission Topography (PET) scans are based upon nuclear medicine with radioactive ligands. The scanner is able to detect pairs of gamma rays emitted by these radio tracers/ligands which inherently bond to the previously designated molecular conformations. These ligands are predominantly labelled with fluorine-18 (^{18}F) or carbon-11 (^{11}C) both of which are radioactive and depending on their chemical combinations have half-lives that allow for enough time to be made, transported to the clinic and administered to the subject within a day. As such complex and precarious compounds that require a high degree of operational efficiency, they tend to be the most expensive of diagnostics. Despite the cost, the scientific benefit of being able to measure disease, in vivo carries great diagnostic accuracy and benefit. However, in most western countries PET is only employed within clinical trial paradigms, rather than primary care. This is primarily due to the cost but currently as no disease modifying compounds exist for AD many health service providers opt for CSF measurement over PET. Analysis can be voxel wise which is a unit of measurement on brain images or by using the Standardised Uptake Value Ratio (SUVR), which is the ratio of the concentration on the image to the concentration injected (which is another ratio of the body weight of the subject to the injected radioactivity level). Time since dosing can also be incorporated into the calculation, but the standardisation of this is not always optimum. One of the critical things to using this modality is to have an imaging protocol (as with MRI). This standardises the machine in question, usually through a ‘phantom scan’ and aligns the scanners produces images in a certain order after the radiotracer

is administered as well as pre-specifying analysis, combining of the images and artefact (errors or blur) removal. If this is not done in a uniform way this leads to messy data and incongruent images across the population being researched.

FDG PET

One of the earliest measurements to be developed for PET was flurodeoxyglucose or FDG. This operates by bonding to glucose uptake transporters in the blood giving an indication, not exclusively but in this case specifically within brain imaging, of which areas of the cortex are being used. This is primarily done over a period of time or a duration of a specific task by measuring the intensity of glucose metabolism of neurons, higher values indicating greater energy being expended. This gives a picture of distinct spatial topographies being utilised overall or under certain conditions.

Within AD, this is less frequently employed than the majority of other neuroimaging modalities as it is not as disease specific as protein ligands described below. However, a great deal of research was done within this modality prior to the inception of these AD specific ligands. One study by Kuhl and colleagues (1982) showed a 26% decrease in glucose metabolism at 78 years of age compared to those aged 18 years in 40 normal individuals. Whilst the n is comparatively small compared to cognition, the data is more reliable and therefore a smaller sample size is warranted, coupled with the cost of the scans, imaging studies arguably yield more data from less individuals. Nevertheless, across a meta analysis of 27 studies, FDG-PET has been shown to have a pooled sensitivity and specificity of 91% and 86% in discriminating probable AD from controls (Bloudek et al., 2011). However, when incorporating a comparison to MCI and either controls or AD, it is not as discriminant, with smaller more variable decreases shown in a range of brain areas (Mosconi et al., 2008). Herholz and colleagues (2002) also showed that in a sample of around 500 individuals FDG-PET has a diagnostic sensitivity of around 84% for probable prodromal AD (>24 MMSE).

When comparing across 119 studies of differing imaging modalities, a systematic review showed FDG-PET to have a superior diagnostic accuracy than the other currently available biomarkers at that time; MRI, CT, SPECT and CSF (Bloudek et al., 2011). This differentiation doesn’t lend itself to an ability to track disease progression well, with changes being defined within certain groupings not tracking with the pathology and only can be suggested that the cause of this is AD without confirmed pathology. Primarily again due to cost and PET availability, CSF indices remained the diagnostic method of choice with clinicians, trialists and many researchers until the advent of amyloid imaging.

Amyloid PET

Amyloid PET is fundamentally a specific diagnostic concerned about the levels of bonding a ligand has with the large conformations of Aβ (predominantly plaques). As the ligand bonds with this instead of glucose, it gives a different indication to the structural and functional characteristics of the individuals' cortex. In the case of amyloid PET, positivity is usually defined as the uptake of the ligand in cortical regions in relation to a reference region which is believed not to accumulate amyloid, which is most commonly the cerebellum. Prior to the first compound being used in humans, amyloid was well established as a key protein within AD (Glenner & Wong, 1984; Beyreuther & Masters, 1991; Hardy & Allsop, 1991; Selkoe, 1991; Hardy & Higgins, 1992). However, it was only with the inception of Pittsburgh Compound B (^{11}C-PiB) that gave the ability to measure amyloid in vivo (Mathis et al., 2002). ^{11}C-PiB has been shown to have high affinity and specificity for fibrillar conformations of Aβ (Mathis et al., 2002; Klunk et al., 2004). Discriminant validity of the SUVRs between AD individuals and controls yielded differences between 1.94 and 1.52 in the initial studies and was rapidly utilised throughout research centres worldwide. However due to its ≈20-minute half-life, this restricted its usage to sites who had their own cyclotron to synthesise this ligand. As other radiotracers were developed to sustain their radioactivity for longer periods, based upon ^{18}F to maximise the half-life duration, this allowed for a wider usage by the research and clinical community.

[^{18}F]-Florbetapir (Wong et al., 2010, Amyvid/Lilly), [^{18}F]-Florbetaben (Rowe et al., 2008, LMI) and [^{18}F]-flutemetamol (Rinne et al., 2012, GE) were subsequently commercially developed and are the three currently approved amyloid imaging probes for human use (Marcus et al., 2014). All three have been shown to have significant correlations with histological findings on autopsy, showing the warranted pathological bonding within humans (Clark et al., 2011; 2012 [both Florbetapir]; Sabri et al., 2015 [Florbetaben]; Salloway et al., 2017 [Flutemetamol]). These three compounds have been the primary agents utilised in research and clinical settings and were developed in order to sustain their radioactivity for longer durations whilst having comparable sensitivity and specificity values to those of ^{11}C-PiB. In a systematic review and meta-analysis of all studies undertaken with these three compounds, there were no marked differences between the diagnostic accuracy of these tracers (Morris et al., 2016). All three also performed better when looking at healthy controls to AD individuals than when incorporating MCI, however this may be due to the heterogeneity of the prior definitions of MCI, the lack of standardisation of the machinery and measurement of the images/SUVRs. When looking at visual reads rather than quantified ones, there were also strong discriminant values for the tracers over other modalities with no significant differences between the three. The meta-analysis also showed very high sensitivity and specificity for all three tracers which were above those of CSF but comparable with FDG-PET when

looking for AD individuals. The comparisons with prior meta-analyses (Zhang et al, 2014; Bloudek et al., 2011) showed this comparable performance with other modalities remains strong, but nevertheless given the underlying interaction with the pathology, amyloid PET remains the discriminant diagnostic tool of choice when available.

Differing cut points have been put forward to discriminate amyloid positive individuals from amyloid negative. As [^{11}C]-PiB is chemically different to that of the [^{18}F] family of ligands, this has a SUVR cut of 1.1-1.4 (Villeneuve et al., 2015), however with the [^{18}F] have been suggested to have between 1.42-1.56 units (Jack et al., 2017) but this is still varied between groups with ADNI suggesting a 1.11 SUVR cut off for Florbetapir (Landau et al., 2012; 2015; Jagust et al., 2015). In comparison, different groups use a visual read instead which standardises groupings in other ways with similar levels of specificity and sensitivity (Morris et al., 2016). The indicative positive/negative discrimination is indexed by whichever method the group conducting analysis prefers, however, whilst human reads will be fairly consistent, there is always going to be stronger consistency in quantification if a standardised value is taken. However, the disagreements between cut offs across research groups for SUVR values therefore mean visual reads have become in some ways more consistent and more widely utilised.

Quantitative analysis has been argued to be superior to visual reads in patients in whom the importance of detecting smaller amounts of Aβ is greater, such as in the early disease stages and for monitoring the effect of drugs that interact with amyloid, which are inherently produced to alter accumulation of Aβ and as such greater granularity in measurement is needed (Mckhann et al., 2011; Barthel et al., 2011; Ikonomovic et al., 2008; Schmidt et al., 2015). Nevertheless, visual analysis is normally performed using a binary scale while quantitative analysis usually involves sensitivity and specificity analysis without pre-specified cut-off values in most cases. Due to this, these data can be subject to over-fitting of the ROC model which can be argued to result in inflated sensitivity and specificity indices (Bacskai et al., 2007; Altman & Bland, 1994). Conversely, visual interpretation depends on the reader's experience and this can lead to a lack of clear cut-off values between normal and pathological categorised scans yielding unnecessary variation in the dataset. Which method is employed is not without its own flaws but both are utilised in different scenarios depending upon, research group preference, subject disease demographics and scan setting (clinical vs research).

Tau PET

All of these issues with interpretation of results still ring true when it comes to tau ligands, however, the weight of evidence within these tracers is still in its infancy. Tau PET tracers, now in their second generation, have recently been shown to track AD pathology by comparing scans to post-mortem brain

tissue (Villemagne et al., 2018). With a number of competing compounds in clinical trials, there is yet to be a clear preference in the field as most of the data is in its infancy (Okamura et al., 2018). Nevertheless, data thus far has yielded strong evidence that a number of compounds track disease topography and bond with the right targets. Currently there are ten first- and second-generation tracers in development, all with only research level of approvals thus far, none are commercially available for wider use (Villemagne et al., 2018).

The key difference between tau tracers and amyloid bonding compounds is the need to discriminate between diseases. This is because tau is a key player in a number of other neurological conditions not only in AD, whereas $A\beta_{42}$ is present in others but is only the first and most abundant hallmark of the disease in AD (Jack et al., 2018). The spatial distribution of tau deposits has been shown to be different for each individual tauopathy and is strongly related to the clinical phenotype of these diseases (Villemagne et al., 2015; Harada et al., 2016; Vogel et al., 2019; Schwarz et al., 2018). Contrastingly to the diffuse and widespread distribution of amyloid bonding ligands across the cortices, the retention of tau ligands is mainly observed in the inferior temporal and parietal cortices of AD patients (Okamura et al., 2018). However, within the normal aging process PHFs have also been shown to accumulate in the medial temporal lobe (Scholl et al., 2016). Therefore, a certain amount of tracer retention in this area is needed in order to classify a subject as tau positive (T^+). But as tau is a downstream effect of AD this cannot be discriminately diagnostic in the way amyloid tracers are for AD (Jack et al., 2018).

Of all the currently available ligands [^{18}F]-flortaucipir has been the most widely studied and validated (Xia et al., 2013; Chien et al., 2013; Chien et al., 2014). Whilst only approved for research use, thus far, it is still developed in a commercial setting. The topography of the tracer is in line with expected AD pathology and has been shown post mortem to follow the NFT Braak staging in AD (Marquie et al., 2019; Harada et al., 2018; Braak & Braak, 1991; Schwarz et al., 2018). However, a significant amount of off target binding has been shown to occur (Ikonomovic et al., 2016; Lowe et al., 2016; Harada et al., 2016). Some studies have suggested that a very significant amount of binding (≈35-50%) is due to tracer interaction with MAO-A and MAO-B, as reductions to this degree are founds when an inhibitor of MAO-B is administered (Ng et al., 2017). However, this has been disputed and evidence for this off target binding of [18F]-Flortaucipir is mixed (Hansen et al., 2018; Smith et al., 2018). Off target binding with MAO-B is also an issue for some of the other first generation tau ligands [^{18}F]-THK5351 & [^{11}C]-PBB3] (Jang et al., 2018; Villenmagne et al., 2018; Vermeiren et al., 2017; Okamura et al., 2018). This lack of selectivity somewhat inhibits these tracers from wide clinical use as off-site binding clouds and impairs any signal gleaned from on-site binding. These tracers, such as [^{18}F]-flortaucipir,

have also shown discrepant data between preclinical and clinical binding profiles (Marquie et al., 2015; Lowe et al., 2016) as well as between ante-mortem and post-mortem findings (Marquie et al., 2017a; Marquie et al., 2017b). Importantly, these inconsistencies have been shown to only apply to straight 4R tau filaments found in other tauopathies and not to the 3R or 4R (repeat isoforms) of PHF-tau found in AD. But whilst the coherent AD profile of binding for tau itself with these ligands has been demonstrated, these discrepancies in findings can actually result from tracer binding to an alternative target and increase the likelihood of a false positive/engender lower specificity, giving further cause for the unreliability of such ligands.This has led to further development from other groups to improve the pharmacokinetics and binding properties of the available ligands. Several are in development still however, some have shown distinct improvements in off-target binding with [^{18}F]RO-948 showed high affinity for NFT and excellent selectivity over other protein formations in AD (Honer et al., 2018). Preclinical and in human data also indicated lower binding affinity to MAO-A&B than that of first-generation tracers [^{18}F]-Flortaucipir and [^{18}F]-THK5351 (Wong et al., 2018; Honer et al., 2018; Gobbi et al., 2017). By contrast to the first-generation ligands, pre-clinical studies show that [^{18}F]-PI2620 (Stephens et al., 2017) binds not only to PHF-tau and 4R tau but also to 3R tau. Proof of concept studies are underway to ascertain whether [^{18}F]-PI2620 also binds to 3R tau in vivo. Both [^{18}F]-MK-6240 and [^{18}F]-PI2620, are currently the only tau ligands to have shown minimal evidence of off-target binding thus far, although both show positive results further proof of concept studies are needed to fully validate these ligands in AD (Walji et al., 2016; Stephens et al., 2017; Goedert, 2018; Okamura et al., 2018).

Overall, tau PET ligands are still in their infancy with imaging to autopsy studies needed to confirm the topographical distribution of the ligand binding truly reflect the requisite tau deposits in the cortex. It is also important to establish accurate quantification of tracer binding given the prospective utilisation of these ligands. Currently there are limitations to our knowledge around the progression, speed and cortical direction of tau pathology during normal aging and AD, as such these ongoing longitudinal studies are key to help clarify the progression of PHF formation. It will be a critical point to assess how the different spatial progressions map onto the clinical presentation of AD. In particular if these cortical “subtypes” present different cognitively. With tau antibody approaches similar to those undertaken for Aβ soon to undergo the clinical development process in AD, fully understanding the properties of these ligands will be critical to accurately analysing the pharmacological properties of these new compounds (Sigurdsson, 2018).

Summation of Current Imaging Methods

Numerous neuroimaging methods are now available for both diagnostic purposes and measuring AD cross-sectionally and longitudinally. With the inception of pathology-based ligands it is easier than ever to track and

measure this specific AD related proteinopathies in vivo. This has led to the improvement of diagnostic criteria (Jack et al., 2018) and increased our understanding of the disease, however there are still a number of unanswered questions. These ligands have become the primary sources of disease measurement where available. But primarily due to cost, primary care needs and ligand availability, they are not routinely employed outside clinical trials. However, in this regard they are indispensable. The ability to not only track decline and measure target engagement carries obvious benefits for new compounds in development. These modalities will be crucial to our further understanding of AD and potential therapies. Previous measures such as FDG-PET and fMRI are not redundant but play a smaller role now in this regard but can provide useful measures of AD when other alternatives are unavailable.

Cognition in AD

Simply put, cognition is thinking, it's all of the mental processes that underlie the majority of human thought. From reasoning and decision making to sensory interpretation of stimuli to memory and information retention, all of these are cognitive processes. The senescence of cognition is as with other parts of the body. However, in Alzheimer's disease this is far more pronounced and a more rapid decline (Jack et al., 2013). Cognition is the point within AD when the pathology manifests itself as noticeable measurable deficits in a subject. This section will address the cognitive aspects of the disease, how they are measured and what more can be done to better address this condition, especially in the very earliest stages of the disease.

Measuring Cognition & AD Disease Staging

When it comes to using a cognitive measure within a disease such as AD, two key factors are important to consider, how a measure tracks with the decline of the disease and secondly, how likely is it to be a good indicator of the presence or absence of the disorder from a single administration (predictive utility). These two things are discernibly diverse and whilst not mutually exclusive they often do not occur for an individual test of cognition used within AD (Soobiah et al., 2019). Longitudinal sampling as well as standardisation of scoring are both ways to increase the ability of detecting impairment and decline respectively for these two areas. Increasing the number of time points for an individual subject increases the available information and therefore the more noticeable deviations from their 'normal' or baseline performance are. Nevertheless, this isn't without its complications and extraneous factors and also relies upon the inherent measurement stability of the scale in question. All of this amounts to an individual measure/scale/test to require a large quantity of normative data both at individual time points on healthy individuals, individuals with specific disease confirmations and longitudinal sampling within samples. Thus, these complexities and mass of data points needed yield very little deviation, from "well-validated" or widely utilised tests which have been employed for significant periods of time (decades), when choosing the best measure for a research, trial or clinic setting.

Tracking the decline of the disease is often related to the pathology, but in AD the complicated disease pathology has led to a complicated interplay between biomarkers and cognition. This has meant treating cognition more like an inherent biomarker itself and tracking its decline over the ≈20-year disease course in relation to the functional impairment exhibited by individuals due to their cognitive impairments (Jack et al., 2018; Albert, 2011; Jack et al., 2016). However as shown in these studies and discussed at length earlier amyloid is shown to be the first large pathological signature and as such relating this pathology to cognitive processes is key.

The term MCI or mild cognitive impairment, is not unique to AD but is widely used as the diagnostic to determine the early signs of disease progression that have manifested themselves in a detectable manner in everyday conditions. MCI is defined as being below one standard deviation worse than the normal performance for an individual's age (Jack et al., 2018; McKhan et al., 2011). Depending on which measure is being looked at, this can (and should) also take into account age, gender and education level. These factors will be explored in more detail later. MCI is commonly inter-spliced with prodromal AD in Europe, with both definitions expound subtle cognitive worsening from a normal level. However, prodromal AD specifically refers to a confirmed diagnosis of AD through biomarker pathology (Dubois & Albert, 2004). MCI does not always infer a diagnosis of AD and can have many other causative factors (Peterson et al., 1999), such as diet (Lourida et al., 2013) or a vitamin B12 deficiency (Malouf & Evans, 2008). Much research was conducted around the concept of MCI in the late 1990s to 2010, classifying subtypes, amnestic vs non-amnestic both outside dementia and AD and within it. However, as discussed at length earlier, with the improvements in disease understanding and biomarkers that are hypothesised to track pathological decline in vivo, the diagnostic terminology has evolved into prodromal and preclinical AD (McKhann et al., 2011; Dubois et al., 2014). MCI still refers to the earliest of cognitive change but it does not engender the same diagnostic classification as it was in the late 1990s. Previous incarnations of this terminology have been Age Associated Memory Impairment (AAMI, Crook et al., 1986) and "benign senescent forgetfulness" (Krul et al., 1962). Whatever the terminology, this represents the very initial progression of a decline in cognition into AD and as such, warrants pervasive levels of research. Although there are many instruments to measure this which will be discussed in detail later, first the focus will be on the domains of interest in AD and how, and if, this maps to the disease progression.

Most measures index one or more cognitive domains, these are more distinct abilities or processes that include episodic memory (ability to remember locations, times and places for example), processing speed, attention and planning (also contained within a wider, higher order group of cognitive processes, termed executive functions). How a measure is determined to index individual domains is down to the component task paradigms. Say a measure of how many words an individual can remember from a list inherently indexes an individual's episodic memory due to the nature of the stimuli they are being asked to recall. As such whilst construct validity and testing paradigms may vary slightly across different measures characterising by domains allows for cross comparison without restrictions to an individual measure.

Amyloid Burden & Cognitive Domains

A number of recent papers have reviewed the literature around this and produced meta-analyses of a large number of cohort studies that have tracked progression across time in both healthy individuals and those with various stages of AD (Mortamais et al., 2017; Backman et al., 2005; Hedden et al., 2013). With the new diagnostic criteria for research at the earliest stages of AD (preclinical & prodromal) it is key to ascertain how best to index the -1 standard deviations below the normal performance for someone who comparable to the demographics of the individual being assessed. What is not clear is if that is the first indicator of cognitive decline or if there are other earlier indicative domains or impairments in cognition exhibited by amyloid positive individuals.

Cognitive measures currently used to describe AD focus upon contrasting differences between AD and non-AD (healthy elderly) individuals, nevertheless this is inappropriate when looking at the earliest indicators of a decline in cognition at the prodromal and preclinical stages of the disease. The very earliest cognitive changes, should they exist, are likely to be subtle and as such necessitate highly sensitive measures that index the inherent brain regions affected by the disease course (Mortamais et al., 2017). As shown in the latest research criteria (Jack et al., 2018), amyloid deposition is the first indication of disease pathology in this model of the disease. As such, relating amyloid deposition to cognition at the earliest stages is key to understanding the cognitive changes within AD itself.

Episodic Memory

Amyloid deposition has been shown to occur decades before the onset of clinical presentation, with poorer performance on tasks involving episodic memory being frequently associated with higher amyloid burden, leading to a strong hypothesis for this being one of the first indicators of cognitive decline within the disease (Mortamais et al., 2017). A recent meta-analysis of over 1,200 individuals from 16 cohorts reaffirmed this finding using mostly cross-sectional studies (Hedden et al., 2013). On an individual study level, results are divergent to this, with heterogeneous findings, this variation has in part been due to the cognitive measures employed and the method of amyloid burden estimation. One such method is by using the most well validated amyloid tracer, Pittsburgh Compound B (PiB), this showed significance on a measure of verbal list learning (California Verbal Learning Test [CVLT]) only when looking at their relationship with amyloid burden as measured by the PiB uptake index (Perrotin et al., 2012). Whereas outside this, a study using staging-based visual reads found no relationship (Song et al., 2015). This is likely to be down to the inconsistency of visual reads on scans such as these, enforcing the need for a quantifiable measurement of amyloid burden applied consistently throughout an individual cohort. Another similar episodic memory task, the WMS-LM (story/paragraph recall), has yielded more consistent results with amyloid burden. Both an 18-month

study by Doraiswamy and colleagues (2012) and a cross sectional study by Reisa Sperling and her group (2013) showed increased amyloid burden resulted in poorer performance on this subtest of the WMS.

Contrastingly, longitudinal studies following individuals from the very earliest stages of the disease, using a range of amyloid tracers, have yielded consistent results with an increasing amyloid burden, significantly greater decline in a range of episodic memory measures (WMS-LM, CVLT, FCSRT [Free and Cued Selective Reminding Task], CBB [Cogstate Brief Battery], HVLT [Hopkins Verbal Learning Test], VRM [Verbal Recognition Memory] and VRT [Visual Retention Test]) has been shown (Lim et al., 2013; Lim et al., 2015; Mormino et al., 2014; Pietrzak et al., 2015; Yotter et al., 2013; Villemagne et al., 2013; Stonnington et al., 2014; Farrell et al., 2017; Rabin et al., 2018; Donohue et al.,2017; Buckley et al., 2017). These findings were predominantly across 6 globally diverse cohorts suggesting the robustness of this outcome regardless of cultural differences, amyloid deposition measurement and episodic memory measure employed. These episodic memory deficits also coincide with individuals showing significantly more amyloid deposition in the temporal lobe (Stonnington et al., 2014; Yotter et al., 2013), thus indicating a neurobiological interplay between the AD pathology and cognitive processes impaired. However, as demonstrated by this broad finding, many different measures are used to index this domain, with all having different nuances important to take into account when measuring cross-sectionally. All aforementioned measures differ in terms of their construct and testing paradigm, making cross comparison between measures difficult and inducing unneeded variance into wider adoption of these findings.

The Dallas Lifespan Brain Study also showed when dichotomising their cohort by amyloid status yielded fewer effects on individual cognitive domains than a continuous SUVR measurement. However, amyloid positive individuals increasing baseline SUVR values led to an increase in the decline in episodic memory, as measured by HVLT and VRM composite. This study also showed a “dose” response in standard uptake value ratio (SUVR) increases (SUVR values of 1.0, 1.2, 1.4 & 1.6) and cognition, these findings remained the same even after controlling for baseline amyloid burden, clearly indicating an interaction between pathology and cognition (Farrell et al., 2017).

In a recent meta-analysis (Baker et al., 2017) of this phenomenon pooled a total of 30 cross-sectional (N = 5005) and 14 longitudinal (N=2584) cohorts. Cross sectional Aβ related cognitive impairment was observed for global cognition (MMSE; d=0.32), visuospatial function (d=0.25), processing speed (d=0.18), episodic memory and executive function (both d’s=0.15). Declines observed related to Aβ were also found for global cognition (d=0.30), semantic memory (d=0.28), visuospatial function (d=0.25), and episodic memory (d=0.24). However, the findings also showed that Aβ deposition

related impairment was moderated by a number of factors such as age, amyloid index, inclusion of control variables such as gender and educational level for both cross sectional and longitudinal analysis. This finding shows that when looking cross sectionally it can be argued that cognitive impairment is much more widespread than a single domain, however, the irrefutable evidence longitudinally is for significant early decline in episodic memory in relation to amyloid load.

Semantic Memory

Whilst these robust deficits shown in episodic memory in relation to amyloid burden, semantic memory is less well known, with this primarily being due to the lack of variation to the tasks employed which are inherently rudimentary (category fluency and naming). Cross sectional evidence across multiple cohorts has stoutly shown no association between semantic memory and amyloid load (Johnson et al., 2014; Perrotin et al., 2012; Song et al., 2015; Doraiswamy et al., 2012; Sperling et al., 2014). However, within the AIBL (Australian Imaging, Biomarkers and Lifestyle) cohort small impairments on the Boston naming test (BNT) have been associated with increasing amyloid load longitudinally (Pietrzak et al., 2015; Ellis et al., 2013). However, this finding has not been shown at every AIBL analysis paper indicating this small finding within a single group requires further study outside this cohort in a longitudinal analysis to expound upon any possible relationship with Aβ load. A meta-analysis of both longitudinal and cross-sectional cohort studies (N=2584 [longitudinal]; N=5005 [cross sectional]) by Baker et al (2017) did show semantic memory to have the largest effect size (-0.28) of all individual cognitive domains when indexing differences between AD and non-AD individuals. Global cognition remained the strongest indicator of difference between groups from this analysis with -0.32 and -0.3 cross sectionally and longitudinally respectively.

Anatomically it is currently thought that the inception point of the AD amyloid pathology is in the trans-entorhinal cortex/perirhinal cortex, which is an area that has been shown to be responsible for *semantic* memory processing (Hirni et al., 2013; Braak et al., 2006; Braak & Braak, 1991). The pathology is then thought to spread to the entorhinal cortex and hippocampus which are key for *episodic* memory processing (Hirni et al., 2013). Thus, from a theoretical standpoint impairment in tasks of semantic memory should be seen to be preeminent in the course of AD. However, a plethora of prospective cohort studies in preclinical AD have shown the exact opposite, with episodic memory being the antecedent cognitive domain of decline (Doraiswamy et al., 2012; Ellis et al., 2013; Farrell et al., 2018; Hedden et al., 2013 [Meta-Analysis], Lim et al., 2014; Lim et al., 2015; Mormino et al., 2014; Roe et al., 2013; Villemange et al., 2013).

There is nevertheless some evidence that suggests this may not be the case, with one 9-year long cohort study and a 14-year follow up study, indicating

the initial cognitive domain of decline is semantic memory (Amieva et al., 2008; Amieva et al., 2005). Both of these studies have shown this finding using the Isaacs Set Test (IST) (Isaacs & Akhtar, 1972). Whilst not being widely employed in cohort studies, this congruency between measure and finding, leads to an argument of the validity of this domain or more strongly the test itself. Further failings elsewhere within other semantic memory measures, indicate inherent implicit flaws in the construct and praxis that are dubitable to this cognitive domain. The duplication of findings for this task is in the inherent characteristics of naming and categorical recall that are argued to be reliant upon concept formation which, is deemed part of the executive function umbrella (Dimitrov et al., 1999). This leads to the conclusion that better more specific measures and wider research of semantic memory is needed to properly index this domain early in the pathogenesis AD. Concurrently as with episodic memory, semantic memory has been shown to be significantly lower up to a decade prior to symptom onset in AD and likewise also shows a descendent inflection in parallel, with an acceleration a few years prior to diagnosis, along with hippocampal shrinkage (Ritchie et al., 2016). Whilst the evidence of semantic memory's relationship to amyloid is murky, the findings within specific measures warrants further more nuanced research.

Working Memory & Executive Function

Looking at the working memory domain, it is further compounded by the inability to accurately discriminate this from other memory processes on measures that engage this cognitive function (Collette & Van der Linden, 2002). Nevertheless, small differences have been found when dichotomising the AIBL cohort on the one back and one card learning measures from the CBB over 18 months and 3 years (Lim et al., 2015; Hollands et al., 2015). However, the majority of studies have found the absence of a robust relationship between amyloid load and this measure (Johnson et al., 2014; Amariglio et al., 2012; Lim et al., 2014; Pietrzak et al., 2015). Until further studies are undertaken using more coherent working memory tasks, showing a clear and distinct relationship between performance and amyloid burden it is suggested that these cognitive processes are not deficient at the earliest stages of AD.

In contrast to episodic memory, there has been an abundance of studies that have found an absence of an association between executive function and amyloid burden (Mortamais et al., 2017). This finding remains regardless of testing paradigm, within this broad domain, or experimental design (Johnson et al., 2014; Perrotin et al., 2012; Song et al., 2015; Amariglio et al., 2012; Sperling et al., 2013; Lim et al., 2014; Roe et al., 2013; Ellis et al., 2013; Baker et al., 2017; Dubios et al., 2018). A small number of studies have shown some associations using executive function measures within AD cohorts; Doraiswamy and colleagues (2012) showed differences between the digit symbol substitution test (DSST) and amyloid burden. Further to this,

another association is between an executive composite measure (comprising category switching, letter fluency and a one back paradigm from the CBB) and amyloid PET SUVR (Pietrak et al., 2015). Nevertheless, both of the measures employed in these two aforementioned studies are not domain specific and have strong episodic memory components, which are argued to be the drivers of these findings. A further study by Doherty and colleagues (2015) is one of only two studies to show significant differences between amyloid positive and negative individuals. They found greater age-related decline in the stroop interference trial and the Trial Making Test (TMT) A and B when their cohort was dichotomised by amyloid status. Both of these measures have a strong attentional component which could be argued to influence this finding, thus arguably giving further credibility to the absence of earlier executive impairment within AD. However, it is broadly agreed from these findings that executive function declines at a later stage of the disease to that of episodic and semantic memory.

Imaging & Cognition

In addition to the concordance with amyloid deposition other structural and functional neuroimaging studies have looked into the relationships with cognition. The primary focus of these studies was centred around the structural changes to the hippocampus and related temporoparietal structures. When looking at these changes cross-sectionally a number of studies have found no significant associations with composite memory scores in preclinical stages of AD (Besson et al., 2015; Toledo et al., 2015; Wirth et al., 2013). However, in an eighteen-month longitudinal study, Seidenberg and colleagues (2013) found that the RAVLT had a steeper slope of decline across this time period which was significant, with smaller bilateral hippocampi (as measured with vMRI) at baseline as well as across this period of decline. The outcome measures from the RAVLT were the sum of the initial trials (1-5) and also the index of delayed recall. Thus indicating a relationship between immediate and delayed verbal recall and volumetric decline in the hippocampus. However, no other studies have shown comparative findings and whilst the number of individuals is small (n =78), the theoretical background of the finding, explains why a task dependent upon new memory recall is impaired with a change in hippocampal size.

Looking at brain wide cross sectional data changes in the temporal lobe and posterior cingulate cortex cortical thickness as well as brain atrophy, have shown significant associations with poorer performance on composite scores of memory and executive function (Toledo et al., 2015; Wirth et al., 2013; Dore et al., 2013). These regions have also been found to have significant associations in early AD between increased hypometabolisms and poorer performance on composite executive function and memory scores (Wirth et al., 2013; Toledo et al., 2015). However, a study by Besson and colleagues (2015) found the absence of an association between increased hypometabolism in these brain regions and poorer composite memory and

executive function scores. The key difference between these studies was a mean age group difference of around ten years, with the Benson study comprised of individuals who were much younger (65-67 mean age) potentially suggesting that this variation is age related.

Overall combining measures of cognitive functions with imaging and pathological markers engenders the strongest possible ability to indicate deterministic relationships. Whilst important discrepancies can be elucidated from cross sectional analysis, using cohort studies that are measuring individuals longitudinally is central to a better understanding of disease progression, both in relation to amyloid load and for greater understanding of domain specific impairments. These impairments should also be held in the context of the scales underpinning their assessment, with domains such as semantic memory being assessed with a dearth of highly specific measures. Current longitudinal evidence suggests the initial decline in episodic memory is seen across many different measures, however more nuanced measures of semantic memory are needed to fully explore this domain within AD. Cross-sectional indications are more varied, however strong evidence suggests a general decrease in global measures are driven by attention, executive function and memory which is contradictory to the longitudinal findings.

Tau Pathology & Cognition

As broad Tau pathology is not thought to occur to a detectable degree until amyloid plaques are abundant, this does not engender the earliest stages of the disease. However, some interesting findings have been shown. And with the growing number of trial failures within amyloid targeting compounds, greater focus has now been placed upon the relationship between cognition and tau pathology (Giacobini & Gold, 2013; Pedersen & Sigurdsson, 2015).

As discussed previously, the spread of pathology is widely agreed to originate in the trans-entorhinal cortex/perirhinal cortex then spreading to the entorhinal cortex and hippocampus spreading to the frontal, then parietal lobes before engulfing the whole cortex (Hirni et al., 2013; Braak et al., 2006; Braak & Braak, 1991) and it is in these latter stages where tau has been implicated to drive the cognitive and functional decline to a degree. Detailed in an overarching manner on the Jack graphs (Jack et al., 2016; 2014; 2011 & **Figure 1.2** above, Hansson, 2021), the initial cognitive impairment occurs after the build-up of amyloid has already predominantly happened and levels of tau are also starting to peak. Greater cognitive and subsequent functional decline also coincides with progressed neurodegeneration, which can also be attributed to increases in tau pathology, due to the stagewise approach with ATN (Soldan et al., 2017; Jack et al., 2017).

Studies have also shown that tau PET has better sensitivity than Aβ PET for detecting early cognitive changes in preclinical AD (Ossenkoppele et al., 2019). However, as this methodology is still in its infancy longitudinal studies with ligands with no off-target binding are yet to read out. Nevertheless, due to well validated CSF assays the progression of tau pathology and cognitive decline is well documented (Samgard et al., 2010). Some studies suggest that high CSF t-tau, but not CSF p-tau, levels correlate positively with more clinical symptoms as well as the degree and speed of cognitive decline (Wallin et al., 2006; Stefani et al., 2006; Samgard et al., 2010; Veitch et al., 2018; Brier et al., 2016). Whereas CSF p-tau levels have been shown to increase during the earlier stages of cognitive decline and progression to AD (Andersson et al., 2008; Hansson, 2021) suggesting that p-tau may be useful as a longitudinal marker of the neurodegenerative process earlier in the disease, whereas t-tau is more akin to latter stages of the disease where greater neurodegeneration occurs.

The very latter stages of the disease (MMSE <12), when AD dementia occurs, are defined by widespread impairment, as once the disease progresses to full AD dementia, cognitive and functional deficits become abundant, more debilitating and extensive (McKhann et al., 1984; APA, 2000). At this point the cognitive and functional deficits make assessment more challenging and as such global measures are routinely utilised to minimise patient burden. These late stages of the disease are when pathology overwhelms the cortex, atrophy is more progressed and widespread and ultimately leads to the end of life. In order to maximise efforts to prevent people from reaching this, the earliest intervention is needed, as such with impairments in cognition being the initial outward manifestation of AD finding the earliest possible signs of these deficits is critical to this.

Factors Influencing Cognition

In order to find such deficits an individual has to be measured against themselves or a large population of healthy individuals (normative data) to be able to tell if they have a specific deficit. The majority of people don’t have a baseline measure prior to the existence of a condition, therefore neuropsychologists utilise population normative data most frequently to compare each person’s individual scores to. However, with an overwhelmingly heterogeneous population a vast array of factors need to be taken into account in order to give an accurate presentation of a individuals’ performance, especially for comparison to normative data, as well as trying to index the state of their own decline. Given enough individuals the variation should mimic that of a gaussian distribution, allowing for the ability of measurement of an individual against a normative population.

Cognitive Reserve/Education

This cognitive heterogeneity becomes less of an issue when factors influencing this variance are taken into account. One of these key factors is

termed cognitive reserve or cognitive resilience. This is the notion that some individuals with AD are able to withstand the amyloid burden placed upon the cortex during AD pathogenesis by maintaining their cognitive level (Katzman 1993; Stern 2012) and as such delaying or reducing the risk of developing AD symptoms (Bennett et al. 2003; Stern 2009). Many factors are thought to comprise and input to cognitive reserve including intelligence quotient (IQ), education, engagement in complex occupations, physical exercise and cognitively stimulating activities (Bennett et al. 2003; Rentz et al. 2017; Scarmeas et al. 2009; Scarmeas et al. 2003; Stern 2009; Stern et al. 1994; Wilson et al. 2003b; Wilson et al. 2003a; Wilson et al. 2007). Using Aβ PET research has shown that amyloid burden has a small but persistent effect on cognition and that cognitive reserve moderates that effect (Rentz et al., 2017; Kemppainen et al. 2008; Rentz et al. 2010; Roe et al. 2008; Giovacchini et al., 2019). When looking at tau pathology the association between cognitive reserve and tau in predicting MMSE has been found to be significant (Rentz et al., 2017; Hoenig et al., 2017). The interaction of tau on this measure of global cognition compared to Aβ may be related to the greater proximal association of tau to cognitive impairment (Delacourte et al. 2002; Nelson et al. 2012), but confirms that tau takes over precedent after Aβ deposition along the AD trajectory. However, these effects are confounded by many healthy controls scoring at ceiling on the MMSE, as this is not particularly suitable at the earliest stage of the disease. Nevertheless, evidence to date indicates that cognitive reserve may be protective to a degree against the early onset of sporadic AD processes and as such enable some individuals to remain cognitively stable despite elevated tau and Aβ burden prior to decline into AD (Soldan et al., 2017; Pettigrew & Soldan, 2019).

Cognitive reserve is commonly measured using a proxy such as educational attainment. Facal et al., (2018) have shown the importance cognitive reserve (education levels) play in the decline from MCI to AD. Key to any analysis within cognitive reserve is how to split education level. This can be looked in a number of ways; chunking into formal education years under the presumption everyone finished school, treating the values in a continuous manner, by grouping the top and bottom quartiles or by chunking this into five-year bins. These approaches are likely to yield differing results based upon sample selection. Within the current cohort this variable has been gathered as a one continuous in nature and will initially be treated as such.

APOE

The apolipoprotein E gene (APOE) has been shown to influence cognition by increasing speed of decline and baseline cognition (Seo et al., 2016; Li et al., 2017; Liu et al., 2013; Raber et al., 2000). It has been widely shown to be the most important gene for driving sporadic AD (Sleegers & Van Duijn, 2001; Frisoni et al., 2022). Each individual has two copies of the allele which is ε2, ε3 or ε4. Those that carry one or more copies of the ε4 variant have an

increased risk for developing AD and reduces the age of onset of AD symptoms by around 12 years (Belloy et al., 2019; Corder et al., 1993; Roses, 1996; Myers et al., 1996; Slooter et al., 1998). Two thirds of those with amyloid positive MCI carry the ε4 risk allele (Mattsson et al., 2018), with the prevalence dropping to 38% of those with clinically diagnosed AD and 14% of cognitively unimpaired individuals (Yamazaki et al., 2019). Whilst carriers of the risk allele for APOE have a greater risk for AD, it is does not follow an autosomal dominant pattern in the same way as the variants in APP and PSEN. Neverthless it is still strongly associated with a familial history of dementia in general (Jansen et al., 2019). Given the earlier onset of cognitive impairment and differential trajectory of decline, cognition APOE status is a key factor to take into account when conducting any analysis of cognition within all stages of AD.

Critically there is a move from the field to diagnose those carriers of the risk allele to be diagnosed as a separate clinical and biological phenotype of AD (Frisoni et al., 2022). As a number of clinical and epidemiological studies suggest that this risk allele infers a distinct clinicopathological entity (Burnham et al., 2020; Toledo et al., 2019; Schmechel et al., 1993).

Gender

Across the last few decades of research AD is something that has always been found to disproportionately affect women more than men; with two thirds of the estimated AD population in the USA being women (AA, 2019). Sex differences have been observed in normal subtle age-related cognitive decline (Beeri et al., 2006). Cross-sectional analysis has shown that women perform better on verbal memory tasks whereas men perform better on visuospatial tasks (Proust-Lima et al., 2008; van Exel et al., 2001; van Hooren et al., 2007). However, longitudinal analysis has been inconsistent with some studies reporting greater annual rates of cognitive decline in men (Wiederholt et al., 1993), women (Proust-Lima et al., 2008), or no sex differences at all (Barnes et al., 2003). From a pathological standpoint baseline hippocampal volume and APOE status has been shown to be predictive of conversion from MCI to AD in women, but not in men (Spampinato et al., 2016; Caldwell et al., 2017). Greater hippocampal atrophy and cognitive decline in women were driven by interactions between sex and amyloid load on measures of memory and executive function, and between sex and t-tau on executive function only. All of which suggests an increased susceptibility of women more so than men to the clinical effects of AD pathology. In a review of the literature on sex differences in AD, Mielke and colleagues (2014) found that there was strong evidence that social and economic factor across the current elderly generation may have influenced the gender prevalence of AD. Factors such as higher occupational and educational attainment for men was the case many decades ago whereas now the opposite is beginning to be true, at least in part for educational attainment (Ryan & Siebans, 2012). With clear current gender differences

within the current most susceptible generation for AD, gender is needed to be taken into account with any analysis as this could have confounding effects on any study results analysed.

Age

Cortical atrophy and enlargement of the ventricles is a part of the benign senescence that comes during normal aging. Subtle cognitive decline is observed longitudinally across the lifespan in the elderly (Hickman et al., 2000; Nichols & Basu, 1994) with normal aging contributing to this very subtle decline. However, recent research suggests that the picture is far more complex than this, with older adults exhibiting both losses and gains in cognitive abilities as they age (Spreng & Turner, 2019). This shift in cognitive architecture is thought to parallel with changes in cortical functional network architecture. These observations manifest in greater functional connectivity across lateral prefrontal regions and the DMN, implicating detectable alterations in cognitive control, memory and semantic processing as part of the normal aging process. This research needs to be probed further as integrated theory of cognitive aging in its infancy. However, age remains a key mitigating factor and must be accounted for in any analysis.

Overall, these nuances have been shown to influence cognitive performance to a level that could have implications for efficacy analysis both in research and clinical settings. Accounting for these variables within forthcoming analysis will be critical. As factors such as aging and ApoE status have been found to add significant heterogeneity to a population.

Testing instruments

One of the key distinctions to make at the outset, is that cognitive testing, other than when it is domain specific, primarily falls into two distinct categories short cognitive tests and longer test batteries (which are usually comprised of domain specific tests). Shorter cognitive measures are primarily designed to condense these longer test batteries into shorter, quicker to administer, subject friendly measures. The upsides are speed of administration and patient burden. The downsides are a lack of scoring range, lack of psychometric validity and applicability to the disease in question. These short measures commonly provide snapshots into each of the domains and a global impression of cognition. The recognition and assessment of individuals thought to have MCI or more advanced forms of dementia, is done using these short cognitive tests and functional questionnaires (Arevalo-Rodriquez et al., 2013; Moyer, 2014) in the majority of settings (primary care, community dwelling or secondary care). Examples of this are the Mini Mental State Examination (MMSE, Folstein et al., 1975), Montreal Cognitive Assessment (MoCA, Nasreddine et al., 2005), Addenbrookes Cognitive Examination (ACE-III, Mathuranath et al., 2000). These measures are routinely employed in elderly populations and clinic settings due to speed of administration and perceived propriety in indexing

the constituent cognitive domains. Much like the Wechsler intelligent quotient (IQ) scales, the short cognitive tests use constituent domains to present an overall global measure for the test itself.

*For a detailed technical description of each of the measures being used for analysis within this thesis please see **Chapter 2**.*

MMSE

The Mini Mental State Exam (MMSE) is the most widely utilised test, both in clinic and research, to screen for dementia. It is a thirty-item questionnaire comprised of very rudimentary components that mapped onto Awareness/Orientation, Language, Attention, Working Memory, Delayed and Immediate Memory and Executive function cognitive domains because of their inherent paradigms. Currently the MMSE is still under copyright but it has been widely utilised in western medicine as a quick screen for basic cognitive impairment, as it only takes approximately 5-10 minutes to administer and score. It is widely used across the world for this reason. The eight domains of the MMSE were designed to cover all key cognitive processes whilst excluding questions around behaviour which when designed were seen to be separate issues that would cloud measurement of cognition.

Several studies have indicated that sociocultural variables, age and education could affect individual scores (Bleecker et al., 1988; Brayne & Calloway, 1990; Crum et al., 1993). Therefore, local standards have been developed for each population and cultural context being evaluated, with most populations akin to this having validation studies conducted (Diniz et al., 2007; Kulisevsky et al., 2009; Shiroky et al., 2007; Trenkle et al., 2007; Kang et al., 1997). For example, some Asian countries having unique variation across the repetition aspects of the MMSE and some specific adjustments being made for nomenclature for orientation to place items and education adjustments for individuals with little formal education (Shim et al., 2017; Ng, 2007; Murphy et al., 2019). The psychometric limitations of the MMSE such as learning effects and large ceiling and floor scoring have been shown to vastly limit their diagnostic accuracy in AD (Tombaugh et al., 1992; Sperling et al., 2012; Mitchell, 2009; Spencer et al., 2013).

Cut off scores are something that varies between research groups with non-concordance seen between clinical trials throughout the early stages of AD (Cummings et al., 2018; Chapman et al., 2016). The Alzheimer's Association (2019) currently defines the cut offs as the following; 25-30 MCI/Normal, 24-21 Mild, 20-13 Moderate and <12 Severe. However, these are subject to regional and cultural variations as detailed above. Within a sample of >23,000 US-based individuals MMSE cut offs were highly inaccurate at diagnostic classification in normal cognition, MCI and AD dementia (Chapman et al., 2016). This study also demonstrated consistently low PPVs (64% in some cases) and AUCs (<0.75), across multiple cut points on the global

score, suggesting whilst target populations may utilise the MMSE as an inclusion criterion, using this alone suggests that subject selection may be biased and highly varied across cognition and function potentially masking study results.

The majority of clinical trials within AD utilise the MMSE as part of their inclusion criteria, this is primarily in order to engender homogeneity of global function and disease level quickly within the cohort (Chapman et al., 2016). However, no significant treatment effects have been shown on the MMSE across every study undertaken in the last two decades (Lasser et al., 2015; Cummings et al., 2018; Siemers et al., 2016; Sims et al., 2017; Egan et al., 2019; Sevigny et al., 2016). Whether this is due to lack of sensitivity as an endpoint or simply a complete failure of the compounds to alter cognition, is unanswered due to lack of an approved compound to alter AD progression. Longitudinally looking at the placebo groups it does tend to track with declines in function but lacks the sensitivity of individual domains. It is also argued that the continued use of the MMSE, albeit brief and inexpensive, for inclusion into AD trials may lead to inaccuracy of efficacy and other study findings not because of a lack of efficacy of the compounds, but due inappropriate subject populations.

As a tool the MMSE is not particularly adept at standalone progression predication of an individual. A recent Cochrane review found no evidence supporting a substantial role of MMSE as a stand-alone single-administration test in the identification of MCI patients who could develop dementia (Arevalo-Rodriguez et al., 2015). Its predictive utility for wider dementia syndromes is much better. As another Cochrane review demonstrated, with specificity and sensitivity measures around 85% and 90% respectively (Creavin et al., 2016). But even within this review and meta-analysis there were multiple cut points demonstrating the variations between definitions in primary care, research and clinic and also indicating a need for validated cut off points akin to other markers of disease or function.

The MMSE is widely utilised and useful for obtaining a quick snapshot in clinical settings. It appears to be more useful as a 30-point staging tool whereby doctors no matter their location can routinely gain a quick overall picture of function and rough current level of ability from a score on this test. However, as would be expected of a short cognitive test, it lacks deeper domain specificity when dealing with constituent cognitive processes. As such whilst an adequate tool for primary care use, other measures are needed to give a more in-depth representation of an individuals’ cognitive ability both cross sectionally and longitudinally to better track decline. Using this as a standalone measure is highly unadvisable due to its documented poor discriminate ability within the AD continuum as well as poor psychometric properties such as large floor and ceiling effects as well as strong learning biases.

Clinical Dementia Rating Scale (CDR)

The clinical dementia rating scale (CDR) was developed in the 1980s as a staging tool for dementia and is still the main measure used for clinical trial outcomes 40 years later (Hughes et al., 1982; Morris, 1993; Lowe et al., 2012). It comprises 6 domains, which span broad cognitive abilities (mainly measured through an interview rather than completed by the individual) and functional day to day activities. The semi structured interview takes around 45 minutes for each person (individual and significant other/study partner) and the clinician raters each of the six domains either 0, 0.5, 1, 2 or 3, plus giving the individuals an overall global score on the same scale (CDR-GS). Scores for each domain are totalled which provides studies with the primary outcome measure termed the sum of boxes (CDR-SB). The three cognitive domains are very broad and are very basic verbal assessments of cognition (memory, orientation and judgement & problem solving), which is often why supplementary domain specific cognitive measures are assessed in conjunction with the CDR. The functional domains consist of community affairs, home & hobbies and personal care, with a heavy emphasis on the latter domain. This component of the CDR gives broad details on everyday life of an individuals, in much greater depth than a standard question and answer measure. Currently, in conjunction with the MMSE the CDR is widely employed within clinical trials for staging of dementia severity, based upon the six-domain framework of the measure. Whilst this gives a broad view of functional level, its applicability to cognitive performance is questionable.

Comparing the CDR and MMSE is important to understand how they compare diagnostically. The level of agreement between the MMSE and CDR scores has been shown to improve across the AD spectrum, further suggesting its suitability for picking up marked deficits rather than subtle cognitive changes. Using previously validated MMSE cut offs for CDR global scores (Perneczky et al., 2006) values were the worst for individuals with MCI $\kappa = 0.15$, slight agreement going up to $\kappa = 0.48$, moderate agreement for severe dementia. This suggests that whilst both scales index pronounced levels of impairment, they do not index the same constituent processes.

The CDR is widely utilised in many different countries, it has been cross-validated against other interview style measures through the CERAD initiative and is available in over fifteen different languages (Morris et al., 1989). In a large sample of community-dwelling elderly, akin to a population-based sample, results showed the CDR has strong internal consistency (Cronbach's α 0.83–0.84) and inter rater reliability of 0.95 for the global score, with test-retest reliability of $\kappa = 0.80$ (Nyunt et al., 2013). A further study of inter rater reliability also showed moderate to high overall kappa scores across clinical trial expert raters (Rockwood et al., 2000). Criterion validity for both CDR-GS and individual domains scores has been shown by correlation with the MMSE, BNT and measures of survival (Fillenbaum et al.,

1996). Importantly the CDR-GS is indexing comparative cognitive impairments in individuals across a period of three decades indicating the general stability for the assessment of dementia (Williams et al., 2009).

However, whilst the CDR-GS has great reliability it has inherent issues for measuring change over time. The CDR-SB score can range from zero to eighteen and as such is often the endpoint of choice when measuring individuals longitudinally. It is argued that the CDR takes into account cognitive processes through its interviews, as executive functions play a large role in the ability for activities of daily living (problem solving, home & hobbies from the CDR), as well as measuring overall memory function with the memory box domain. These factors, as well as the in-depth nature of the scale, differentiate from the short form questionnaires of activities of daily living (ADL).

The CDR-SB index has been shown to have excellent two-year internal responsiveness, for disease progression, indicating it is a prime candidate as a sole primary endpoint in disease modifying trials (Coley et al., 2011; Williams et al., 2013; Cedarbaum et al., 2013). However, further exploration of the usefulness of this as an endpoint earlier in the disease is needed as it is still unclear how some of the early changes are clinically relevant. This index is also well documented to follow the progression of AD pathology, with higher levels of pathology (both Aβ and tau) showing increased rates of decline on the CDR-SB score (Samtani et al., 2014; Weiner et al., 2017; Veitch et al., 2019).

Overall, the CDR is a validated and a tool capable of indexing the disease staging of a subject across the AD spectrum. Where it does fall down as a measure, is in the earlier stages of the disease (prodromal/preclinical AD). Whereby it doesn’t have the domain specificity to uncover subtle cognitive decline within single or multiple cognitive domains. Nevertheless, this is a worthwhile measure and provides clinical and functional information on a subject through the gleaning of information across the administration period.

ADAS-Cog

For the past three decades the eleven-item version of the Alzheimer’s Disease Assessment Scale - Cognitive Subscale (ADAS-Cog11) has been ubiquitous with clinical trials in AD across the disease spectrum (Podhorna et al., 2015). However, with the shift in trial populations towards preclinical and prodromal AD the ADAS-Cog11 has been shown to have inadequate sensitivity to detect changes at this early stage of the disease (Doraiswamy et al., 2001). This is due to many individuals with MCI scoring at ceiling on eight of the domains of the ADAS-Cog 11 (Winblad et al., 2008; Grundman et al., 2004; Hobart et al., 2013; Llano et al., 2011; Graham et al., 2004; Ueckert et al.,2014; Raghavan et al., 2013; Pyo et al., 2006; Zanotta et al., 2014). As a result, new measures were added to the form to give the ADAS-

Cog-13. This comprises of Word Recall, Naming Objects and Fingers, Commands, Constructional Praxis, Ideational Praxis, Orientation, Word Recognition, Language, Digit Cancellation, Delayed Word Recall and a Maze test (Rosen et al., 1984; Mohs et al., 1997). The full ADAS as well as the cognitive subscale was originally designed to assess the severity of dysfunction in mild to severe AD (Rosen et al., 1984). The full ADAS takes around forty-five minutes to administer and is scored from 0 to 150 (Cog scale 0-70 [11], 0-85 [13], 0-90 [14]) by summing errors made on each task, as such higher scores indicate greater impairment. There have been many variations made of this measure however overall in a review of the numerous variations of the ADAS-Cog results suggested that the original ADAS-Cog is not an optimal outcome measure for pre-dementia studies (Kueper et al., 2018). Nevertheless, with the ubiquitous prominence of the ADAS-Cog in AD clinical trials, replacement outcome measures require much validation. Thirty-one modified versions of the ADAS-Cog were found within the literature review. Modification approaches that appear most beneficial include altering scoring methodology or adding tests of memory, executive function, and/or daily functioning. Although modifications improve the performance of the ADAS-Cog, this is at the cost of introducing heterogeneity that may limit between-study comparison.

All three acetylcholinesterase inhibitors that were approved in the 1990's showed improvements (lower scores) on the ADAS-Cog on treatment (Birks, 2006). As a result of this and the absence of new positive results since then, the ADAS-Cog has stayed as a primary efficacy measure of choice to this day in the vast majority of phase III clinical trials in AD. As suggested by the authors at its inception (Rosen et al., 1984) and by a recent review of the variants of the task (Kueper et al., 2018) this is not fit for purpose within the early stages of the disease due to its poor psychometric properties and lack of published validity data (Soobiah et al., 2019). However, with this measure being utilised in a steadfast manner across clinical trials, researchers have turned to other nuanced ways of improving this index without removing it from the efficacy analysis as a primary or secondary endpoint. These have come in the form of cognitive composite scales.

The Validity of These Measures

The biggest question surrounding these measures is are they fit for purpose within the earlier stages of AD. As discussed above the MMSE, CDR and ADAS-Cog are stalwart measures for AD. They measure a variety of cognitive domains and day to day function (Lezak, 2004). However, whilst the performance and accuracy of these measures has been shown in mild and progressed AD (Kueper et al., 2018; Schmand et al., 2011; Vellas et al., 2008), it is far more questionable for measuring subtle impairment earlier in the disease (Mura et al., 2014; Podhorna et al., 2016; Mortamais et al., 2017). Failures for adequate measurement in these early stages of AD, is

probably largely due to the floor and ceiling effects for these measures within these populations (Mura et al., 2014; Baker et al., 2017; Duke Han et al., 2017; Karin et al., 2014; Cano et al., 2010). These tests are nearly always administered in combination, as such the testing length of these assessments typically spans several hours leading to fatigue effects on performance. So not only do the tests take too long and don't measure the early stage of the disease, they do not fully translate to day-to-day function and by proxy therefore limit their clinical relevance to the clinical presentation of AD (Rockwood et al., 2007; Royall et al., 2007).

As these tests often comprise assessment of a variety of cognitive domains, such as episodic memory, semantic memory, attention, language and executive function. The questions become:

- how well do these measures accurately index these domains they are designed to measure (content validity)?
- How well do measure the disease process itself within AD (criterion validity)?
- How does it perform across different regions and demographic groups (cross-cultural validity)?
- Does the outcome measures/output scores from a measure, reflect the domain being measured (construct validity)?
- Can measures with the same underlying construct be used interchangeably to provide the same outcome (concurrent validity)?
- Even when measures share the same domain of interest, do they measure the same domain (convergent validity)?
- And finally, how do they relate to everyday performance/a clinical presentation (interpretability)?

These questions form the basis of psychometrics when designing cognitive measures. However, a recent review of cognitive measures used within clinical trials in AD, designed to assess treatment response/efficacy of compounds, showed only 50% of the measures used as the primary outcome measure in a given trial, had published information about their validity (Soobiah et al., 2019). Whilst it is important to state these measures have been broadly translated and used in multiple countries and in multiple clinical trials, the cross-cultural validity, is broadly unknown. The CDR in particular only has published information on how it performs in relation to other measures of the same construct (convergent validity). Whilst it does have strong internal consistency and inter-rater reliability information published, given that the only published validation efforts made are to align to prior measures, not AD itself, there is a strong argument to be made it is not fit for purpose in early AD. The picture for the MMSE and the ADAS-Cog is similar. These two measures only have published validation efforts in relation to how the measures perform when comparing between measures at the same time (concurrent validity). They both have test re-test and inter-rater reliability metrics which are fairly strong, however the fundamental question

remains the same, how can researchers be sure they are measuring the disease course or even the right domains accurately within AD itself.

This fundamentally underlines the issue of the validity of these measures and the issues facing the field. A fundamental aim of this thesis is to improve this understanding of these commonly used measures within AD by understanding their validity as measures with the early stages of the disease. But also, the clinical meaningfulness of them by linking them to disease stage and pathology at the core of AD.

Cognitive Composites & Index/Discrepancy Scores

With the plethora of testing instruments available covering different domains, and the immovability of the current key outcome measures in clinical development, the field has somewhat shifted towards composite endpoints. The FDA (2017; 2018) have also indicated that they have a favourable opinion on having validated composite measures as part of the efficacy package within clinical trials. Made up of different domain measures from the widely validated tests and other domain specific measures, composites have come to play a key role both in clinical trials and academic research with two key composite measures being developed for preclinical and prodromal AD (Sabbagh et al., 2019).

Preclinical Alzheimer's Cognitive Composite (PACC)

The preclinical Alzheimer's cognitive composite (PACC) was developed (Donohue et al., 2014) using data from three observational cohort studies (AIBL, ADNI, and ADCS Prevention Instrument Study) in Aβ positive and negative individuals. It is comprised of the MMSE (Folstein et al., 1975), Logical Memory (LM) Delayed Story Recall (Wechsler & Stone, 1987), the Digit-Symbol Substitution Test (DSST; Wechsler, 1981) and recall from the Free and Cued Selective Reminding Test (FCSRT; Grober et al., 2009). These measures were selected on the basis of studies showing changes in these cognitive measures many years prior to functional decline being exhibited (Bateman et al., 2012; Amieva et al., 2008; Salmon et al., 2013). Other measures have been proposed to be added such as semantic memory but are yet to be as well validated as the initial composite (Papp et al., 2017). In preclinical AD, a large number of individuals require screening to find those with preclinical AD, as by definition none are exhibiting impairment. These cognitive measures have a number of advantages over biomarkers or imaging methodologies, such as cost and patient burden as well as being closely related to the core symptomology of AD and akin to disease progression. But most importantly for clinical trials they have proven sensitivity at the latter stages of the disease to treatment effects.

Studies on the level of amyloid load in the brain (in essence between preclinical AD and prodromal AD) have shown large and significant effect sizes (d=0.85) between cognitive decline after three years in these groups

(Bransby et al., 2019). This was found regardless of MMSE inclusion or not. The debate around the inclusion of the MMSE, as discussed earlier, has shown evidence of a lack of sensitivity earlier in the disease course of AD and lack of strong psychometric properties (Lim et al., 2016). However, this was not the case in this most recent study on this composite. Critically across the three initial cohorts studied with this composite there was significant separation between the cognitive decline between Aβ+ and Aβ- individuals as early as 12 months (Donohue et al., 2014). However, as all early AD trials have A^+ positivity as stringent inclusion criteria, this analysis only goes to show the sensitivity compared to cognitively normal individuals and as such the subsequent effect size findings between differential early stages of AD (prodromal vs preclinical) suggest strong suitability for inclusion as an efficacy endpoint in preclinical clinical trials.

Alzheimer's Disease Composite (ADCOMS)

The ADCOMS composite has demonstrated increased sensitivity to pAD/aMCI than the constituent measures that comprise it (Edgar et al., 2016; Swanson et al., 2019; Hendrix et al., 2019; Logovinsky et al., 2019; Tahami et al., 2019; Bajaj et al., 2019). Whilst some of these phase II trials have shown promise, the largest trial to show positive results was the phase II trial of Lecanemab (Swanson et al., 2019). The ADCOMS outperformed all other cognitive or functional variables within the dataset. It can be argued that this is showing a true treatment effect due to its increased sensitivity shown outside of this trial.

With the near total failure of AD phase III trials over the last two decades and the lack of validity of the ADAS-Cog to find treatment benefit, the FDA has finally sought to relax their guidance. In their most recent paper, their guidance was to shift towards the notion that a change in cognition is clinically meaningful. However, this is dependent upon concordant biomarker change (reflecting underlying AD pathological changes) as well as the disease stage of the participants. Incorporated into this shift, is the assertation that the clinical meaningfulness can be established through a composite endpoint, that reflects the change in cognition across multiple domains and is pervasive in nature (Sabbagh et al., 2019; Edgar et al., 2018; FDA, 2013; 2018). However, the key criteria still remain components of the MMSE, ADAS-Cog & CDR.

The ADCOMS was developed for use in prodromal/MCI to mild AD (Albert et al., 2011; Dubois et al., 2010). The analysis of the pAD/MCI due to AD data set, comprised of 4 study placebo groups from studies completed between 2004-2010 n=1160 (Wang et al., 2016). However, CSF biomarkers were only available for a subsection of this cohort (n=405). The mild AD analysis comprised of three datasets gathered within the same timeframe as the pAD/MCI due to AD data, n=469. This was then validated against a further dataset from each population, pAD/MCI due to AD n=784 (Peterson et al.,

2005), mild AD n=236 (Rogers et al., 1998), however, neither study comprised an amyloid positivity assessment. This could be argued to impinge on the efficacy of this composite as there could, based upon recent trial data, be up to 30% of individuals in the combined cohort that are amyloid negative (Egan et al., 2018).

The development of this composite was produced through a method of partial least squares (PLS) regression. This method allowed for the description of a model for linear decline, allowing the functional and cognitive declines to be linearly modelled. Thus, giving AD a singular trajectory rather than multiple discordant ones across many domains. By alleviating the effects of any measures that concurrently measure identical components (such as orientation measures as PLS downweighs these in the model), this allows for more model accuracy and far lower chance of coincidental correlations. To ensure only suitable measures were included in the composite measure a threshold was set at 0.8 for the variable importance of projection (VIP) as per Wold's criterion for predictor deletion, akin to an effect size (Wang et al., 2016). This led to 12 measures comprising the composite. Each one was weighted based upon their relevance to the explained variance in the model of decline from the PLS regression analysis. This linear decline was also only calculated across 12 months which could be argued to be far too short a timeframe. This is because disease progression within early-stage AD occurs across multiple years (Jack et al., 2018) and is not always uniform in nature (van Maurik et al., 2019; Ferretti et al., 2020).

A number of criticisms have been given to the ADCOMS since its inception, the arbitrary weights given to the variables have drawn considerable criticism from a number of research groups (Schinder & Goldberg, 2020). Also, as recent research has demonstrated individuals with amyloid positivity and tau positivity decline at a more rapid rate (Mattsson et al., 2020), something that was not incorporated into the initial methodology. Further to this, with the amyloid heterogeneity of the original development cohorts now also confounded by presumed tau heterogeneity, this can be argued to be a factor that could significantly influence the decline of the cohort and therefore inducing questions into the construct validity of the weights and the measures included within the composite. Further criticisms stem from the duration for disease modification. In clinical trials for AD this is commonly 18-24 months whereas the ADCOMS was based upon a 12-month decline. To account for the magnitude of expected treatment effect and by proxy, power calculation and sample size, the studied timeframe has to be long enough to see separation in line with the disease course. Despite all of the aforementioned flaws with the ADCOMS it remains more sensitive to decline than either the ADAS-Cog or the CDR. And in the absence of a positive alternative, has been widely employed within phase III trials of AD, even though, it can be argued to be still undergoing validation it remains the best hope of finding any efficacy there is.

Discrepancy Scores within AD

Following on from the manipulation of existing endpoints, one aspect of the short cognitive tests commonly used across all stages of AD, is the discrepancies between constituent cognitive domains. This methodology is widely utilised in IQ testing and looks to find an individual profile of a subject across a large number of domains (Wechsler, 2008; 1981).

This paradigm within preclinical AD was first undertaken by a group is the US in the early noughties (Jacobsen et al., 2002). They looked at a group of 20 healthy elderly individuals and 20 individuals with what at the time was termed preclinical AD. These analyses were taken from data of the individuals who subsequently met diagnostic criteria early AD when they were still considered healthy controls at the time of the test administration. All diagnosis for all individuals were based upon independent annual examinations from two senior neurologists with the individuals classified into the preclinical AD group participating as control individuals for an average of 4.6 years prior to a subsequent change in diagnosis. The clear differences now are the diagnostic criteria and availability of Aβ indices which may lead to Aβ- individuals within the sample. However, with the longitudinal follow of these individuals it was clear that they all went on to have dementia which suggest the presence of pathology, albeit likely some non-AD varieties.

The two groups in this study were matched on age, education and gender and were assessed on a cognitive battery comprising of the Dementia Rating Scale (DRS; Mattis et al., 1976), CVLT (Delis et al., 1987) long delay free recall, Boston Naming Test (BNT; Kaplan et al., 1983), the Block Design subtest from the WISC-R (Wechsler, 1974) and the WAIS-R vocabulary sub-test (Weschler, 1981). The measures index semantic and episodic memory, shown earlier to be key indicators of early disease pathology which was unknown at the time of selection here. Further to this the two components of the Wechsler batteries are short yet key indicators of cognitive processes known to be at their peak towards the latter stages of maturation of cognition (Spreng & Turner, 2019). This cognitive battery maps on very neatly to our current understanding of preclinical AD and pathology. The results from this small study showed statistically meaningful significant differences in the asymmetry/discrepancy score between these cognitive domains and subject group. Whereas the DRS and CVLT scores also were significantly different between the groups, the size of the effect was not as large as that of the discrepancy measure. Thus, suggesting that as seen with other measures profiling IQ, discrepancy analysis may indicate some of the earliest changes within preclinical AD.

Whilst these findings come from a small sample, their relation to current understanding should not be under-valued, as such replicating this study within new criteria and corresponding biomarker and imaging indices is

highly warranted. Incorporating an analysis such as this but with the most widely used measures is also a further strong avenue of potential research.

Addressing Gaps in Our Current Understanding

Overall, it is widely agreed that initial impairments at the earliest stages of the disease are best indexed by domain specific cognitive measures. However, there is no agreement on which measure to use. Whilst typical AD manifests itself through amnestic impairment, that is still a highly varied cognitive process. And as the earliest domains of impairment are semantic and episodic memory, which are not best measured using global indices. These global measures need to be further looked at in greater detail to better understand what they precisely measure.

Further to this, global cognitive measures are often utilised in conjunction with other staging measures. These measures bare slight relevance to cognitive domains of impairment, but primarily focus upon presenting an overall picture of cognition and function, depending on which measure is employed. As validating new instruments is prohibitive, a better understanding of these measures is required.

Consequently, more nuanced ways of using existing measures are being sought. As seen with the ADCOMS composite, this is hoped to give greater sensitivity to the measurement of cognitive impairments due to AD. To better characterise and elucidate the earliest stages of this impairment, pathology must corroborate and be related to cognitive impairment. This improvement in a confluence, between the clinical and cognitive presentation of AD and the pathological markers of AD, will also help rule out co-morbidities known to cause such impairments. These biomarkers and imaging modalities play a critical role in diagnosis and progression of AD. However, non-invasive, cheap and quick to administer cognitive measures, currently take precedence. But they have not changed for nearly 3 decades. Nevertheless, until a better understanding of pathology and its relationship to cognitive impairment and clinical presentation of AD these measures cannot be replaced. It is therefore critical to develop a better understanding of these measures to explain variability, highlight areas for improvement and bridge the diagnostic gap that exists between clinical presentation and pathology.

This project will look to build upon our current understand of these cognitive impairments. It will explore the hypothesis that the discrepancy measures may indicate early disease symptomologies prior to individual cognitive domain decline. It will also look to explore the potential sources of variance within cohorts and how the relationships between measures of verbal and episodic memory may not always be concordant. These hypotheses will be explored within highly phenotyped cohorts with biomarker, functional,

cognitive and imaging measures associated with each subject allowing for wider comparisons and implications to be gleaned from these analyses.

Each chapter and study will focus on one (or more) part/s of validity for these key scales in question. **Chapter 3** will look to ascertain the criterion and convergent validity of a new measure within AD, a discrepancy score computed from commonly used measures. Whilst also exploring the cross-cultural validity of the full battery of tests. **Chapter 4** will delve into great detail on the concurrent, convergent and construct validity of amnestic memory measures within 2 different AD-related populations. And finally, **Chapter 5** will focus on criterion validity of these measures and the overall interpretability of these measures in relation to marrying the biology and clinical phenotypes of AD.

Chapter II: Methodological Considerations

Whilst academic studies rightly or wrongly stray away from the stalwart AD measures (CDR, MMSE & ADAS-Cog), they tend to explore other measures with greater construct validity and disease course specificity. This is a luxury that is normally not applicable to clinical trials. Currently, despite many efforts, all phase III trials remain cemented in these perennial measures. To this end, the goal of the proposed analysis and methodology is to look at potential causes of variation within these measures, how they can be comprised in a different way and to use machine learning methods to predict biomarker grouping from cognitive and clinical measures. This will be undertaken to help improve and understand these measures. The hypothesis is this better understanding will help improve trial designs, disease characterisation and diagnosis in primary care. This chapter will give full details of the cohorts used for all subsequent analyses, as well as the methodology and rationale for each analysis.

Dataset Composition

The main dataset under investigation is taken from the screening and baseline dataset of two large global phase III trials of Elenbecestat (E2609-G000-301 or MissionAD1 & E2609-G000-302 or MissionAD2) in early AD. Elenbecestat is an oral BACE-1 inhibitor that has been shown to reduce the Aβ level in the cerebrospinal fluid (CSF) (Majid et al., 2016). The population consisted of individuals with a diagnosis of MCI due to AD and no more than 25% diagnosed as early stage mild dementia due to AD. The cohort consists of 2212 randomised individuals and a cohort of 9758 screened individuals. As these studies were terminated early, only 959 individuals had reached one-year of the study, hence the focus of these analysis will be cross-sectional based upon baseline and screening data. The author was of the study team across both studies and had a key role in the recruitment of individuals to this cohort, advising sites and discussing ways in which to aid individual cognitive pre-screening and study uptake on a site by site basis.

The screening paradigms were broken down into five tiers of individuals aged 50-85. This was to reduce participant burden by not subjecting them to

assessments that were more invasive towards the end of the screening paradigms.
Figure 2.1 Flow diagram of the screening tiers. Bolded assessments had to be passed to move onto subsequent tiers (or randomisation).

Individuals had to complete and meet criteria at each tier of screening before moving onto the next tier of screening. Tier 1 consisted of participant demography and medical history as well as meeting cognitive/clinical assessment criteria: Mini Mental State Examination (MMSE; total score of 24 or greater), International Shopping List Task (ISLT; scored against normative data, criteria of 1 standard deviation below the age and sex related mean score on either the immediate or delayed recall component), Clinical Dementia Rating Scale (CDR; global score of 0.5 & memory box score of 0.5 or greater) and the modified Hachinski ischemic score of less than 5. The Cogstate Brief Battery (CBB) was also administered in the 20 to 30 minutes between the ISLT total recall and delayed recall tests as a common distractor task. All individuals will be confirmed as meeting the diagnostic criteria for either MCI due to AD or the early stages of mild AD, and that they do not have other medical conditions which may interfere with study participation. All of these criteria had to be met before the participant could move onto the second tier of screening. Full details of each of the key cognitive assessments are described in the next section of this chapter.

Tier 2 of screening consisted of the Columbia Suicide Severity Rating Scale (C-SSRS), EQ-5D-5L, QOL-AD and the Zarit Burden Interview (ZBI). Other than a positive answer to suicidal ideation on the C-SSRS, there were no other criteria for these scales that had to be met to proceed to Tier 3.

Tier 3 assessments include a complete physical examination with dermatologic review, a full neurologic exam, measurement of vital signs (body temperature, sitting heart rate, and sitting BP), measurement of height and weight, and a single 12-lead ECG. Blood samples were taken for measures of clinical chemistry, haematology, thyroid function, vitamin B12, and a viral screen. Blood samples were also taken for pharmacogenomic (PGx) analyses, PD, exploratory biomarker assays, and for immunologic assessments. A urine sample was also taken in order for urinalysis and a serum pregnancy test for females of child-bearing potential. If these samples showed, Absolute lymphocyte count (ALC) below Lower Limit of Normal (LLN) or below 800, TSH above the normal range, abnormally low Vitamin B12, immunoglobulin (Ig) deficiency or other immunodeficiency disorders, viral hepatitis, TB, shingles, herpes simplex virus, INR ≥ 1.7, bilirubin $\geq 1.5 \times$ ULN; albumin < LLN; ascites or hepatic encephalopathy then an individual was excluded from further procedures.

Tier 4 comprises an MRI scan for brain abnormalities which may exclude study participation. Additional scanning sequences for vMRI and fMRI

assessments were run immediately following the eligibility MRI sequences in all individuals. Contraindications and pathological findings on the eligibility MRI included but weren't limited to; an area of superficial siderosis, evidence of cerebral vasogenic edema, evidence of cerebral contusion, encephalomalacia, aneurysms, vascular malformations, or infective lesions, evidence of multiple lacunar infarcts or stroke involving a major vascular territory, severe small vessel, or white matter disease, space occupying lesions, brain tumours (however, lesions diagnosed as meningiomas or arachnoid cysts and less than 1 cm at their greatest diameter need not be exclusionary criteria).

Within the cohort in question, these aforementioned MRI findings were treated as exclusionary if they appeared on an MRI scan for a participant. However, the most common machine available to sites was a 1.5T scanner meaning some of the smaller details may not have been visible to the central readers. As previously elucidated in more recent research post data collection, some of these cortical microinfarcts may not have been visible and as such the root cause of individuals' cognitive impairment in the absence of (or indeed in conjunction with) amyloid pathology. This will be discussed further in this chapter below.

Individuals who have met all eligibility criteria from tiers 1-4, were then assessed for biomarkers of brain amyloid pathology with either amyloid PET (Screening amyloid PET) or cerebrospinal fluid Aβ(1-42) (Screening CSF) or both. There were 4051 individuals who had assessment of amyloid pathology undertaken in total. These individuals all met criteria for MCI due to AD or Mild AD with cognitive impairment of varying degrees with many other potential causes of cognitive impairment ruled out apart from confirmation of amyloid (prior to these assessments taking place). From these 4051 individuals, 2212 were randomised to Elenbecestat or placebo and undertook further assessments prior to dosing to their assigned treatment arm. The MMSE, CDR, ADAS-cog14, functional assessment questionnaire (FAQ) and Neuropsychiatric Inventory ten item questionnaire (NPI-10) were all completed prior to this first dosing.

Whilst this cohort is one that has been highly characterised and relatively homogenous in nature, it is also taken from 644 clinical sites across 29 different countries spanning all regions of the world. This can be argued to be fundamentally representative of the global AD population. This unique cohort requires substantial considerations when dealing with data prior to any analysis which will be discussed in relation to each analysis chapter below.

The second cohort under investigation is one taken from memory clinics across the USA, Canada & the United Kingdom. This cohort had no screening criteria, but presented to a memory clinic to be screened for part of a clinical

trial. These individuals either presented due to believing they had memory impairment, had no memory impairment but a significant other believed they did or were being monitored by the site in question longitudinally. The recruitment of these subjects and how they came to each clinic varied greatly by country and clinic. As site/clinic and individual level information on recruitment methods was not captured this was not available to report.

Each site was trained on the administration of the WMS-LM and by the author on the alternate version of this measure. The alternate form of the WMS-LM was based upon the original alternate version developed by Morris and colleagues (1997) found on linguistic and psychometric comparative principles detailed in the original paper. The alternate version of the WMS was developed to be country specific, with specific context around normal activities differing between the 3 country specific versions. For example, in the US one story was about a football game, whereas in the United Kingdom it was around a rugby game. All three short and long form stories were adapted using the same principles and relevant to each country in question. The 196 subjects undertook the alternate version first then underwent testing on the WMS-LM. To note these individuals had no known clinical comorbidities or other known reasons for any cognitive impairment they may be experiencing. The order of administration was set due to the necessity of the screening procedures at all memory clinics. The training occurred by the author virtually and data was captured via pencil and paper administration with upload to a central repository and data transcription from those forms was made into a singular spreadsheet.

Table 2.1 Cohort Composition For MissionAD (MAD) Cohort and Secondary Cohort From Chapter 4 (VLL). Green for clinical and cognitive measures, red for imaging, yellow for other biomarkers.

Variable	MAD Screening	MAD Baseline	VLL Cohort
Demographics & Maximums	**9758**	**2208**	**196**
ISLT	8562	-	-
CBB	8497	-	-
MMSE	9114	2208	-
CDR	6551	2208	-
ADAS-cog14	-	2193	-
FAQ	-	2039	-
NPI-10	-	2039	-
EQ-5D-5L	-	2039	-
QOL-AD	-	2039	-
ZBI	-	2039	-
Sleep/Dream Questionnaire	-	2039	-
WMS-LM	-	-	196
Morris Paragraphs	-	-	196
vMRI	1782	1279	-
Amyloid PET	3728	1281	-
Tau PET	-	30	-
CSF Ab42, 40, t-tau & p-tau, ng	323	78	-
Plasma Ab 1-x	-	1597	-
Plasma NFL	-	943	-

Cognitive Measures

Cognitive instruments and their outcome measures will form a fundamental cornerstone of all analyses within this thesis. All chapters will contain one or more of the cognitive instruments described here with full details of all measures used across all analyses are be outlined below.

Clinical Dementia Rating (CDR) Scale

The CDR (Hughes et al., 1982; Morris, 1993) is a multi-domain global *functional* scale used to denote the presence and severity of dementia and is widely used across AD clinical trials (Lowe et al., 2012). The CDR is considered the gold standard for staging dementia severity, primarily focused upon functional impairment but it was designed to give an overall holistic view of the individuals' disease severity. The scale is formed of six domains and is administered through a semi-structured interview with the patient and then the caregiver, they are scored either 0, 0.5, 1, 2 or 3 and comprise a global score (CDR-GS) and a total score termed sum of boxes (CDR-SB). The domains are memory, orientation, judgement and problem solving, community affairs, home and hobbies, and personal care. As the global score can only be one of five outcomes, variation and suitability for tracking pathology is limited, however it is primarily employed within the framework of a staging approach, similar to a simplified MMSE score. The CDR is widely used in many different countries, it has been validated through the CERAD initiative and is available in over fifteen different languages (Morris et al., 1989). Specifically, the CDR has been standardised for worldwide use and as such has demonstrated strong inter-rater reliability and convergent validity (Morris, 1997). The CDR-SB score can range from zero to eighteen and as such is often the measure of choice when assessing individuals longitudinally.

Alzheimer's Disease Assessment Scale Cognitive Subscale – 14 item version (ADAS-Cog14)

The ADAS-Cog is a subscale of the broader AD assessment initiative that was developed by consortia in 1984 (Rosen et al., 1984). The ADAS was designed to measure both cognitive and non-cognitive aspects of mild to severe AD. There are 21 tasks in total but the focus of the bolus of work is on the cognitive aspects of the scale which were initially designed to have 11 items. This has since been adapted to many different variations (see **Chapter 1** for a discussion on this). The version most widely used and in the vast majority of clinical trials in early stages of AD (and in MissionAD 1&2), is the 14-item version. This has a total score out of 90 whereby a higher score indicates a greater level of impairment which in essence is error scoring. **Table 2.2.** details each subscale within the 14-item version and how it is scored.

Table 2.2 ADAS-Cog14 Subscale descriptions and scoring conventions. Adapted from ADAS-Cog Scoring & Administration Manual (FDA, 2012).

Task	Description	Scoring
Word Recall	A list of 10 words is read by the participant, and then the participant is asked to verbally recall as many of the words as possible. Three trials of reading and recalling are performed.	Mean number of words not recalled across the three trials; scoring range is 0 to 10.
Commands	The participant is asked to perform commands that involve one to five steps. For example, the two-step command is to "Point to the ceiling, then to the floor."	Scored from 0 to 5 based on the largest number of steps that are correctly performed (score is 0 if five step command is correctly performed).
Constructional Praxis	The participant is shown four geometric forms (circle, two overlapping rectangles, rhombus, cube) and asked to copy them on a piece of paper.	Scored from 0 to 5 based on the number of correctly drawn forms.
Delayed Word Recall	The participant is asked to recall as many words as possible from the 10 words presented during the WORD RECALL task. There is one trial of recall. This task should follow immediately after completion of the COMMANDS and CONSTRUCTIONAL PRAXIS items.	Scored 0-10 based on the number of words not recalled
Naming Objects & Fingers	The participant is asked to name the fingers of their dominant hand as well as twelve objects: flower (plastic), bed (doll house furniture), whistle, pencil, rattle, mask, scissors, comb, wallet, harmonica, stethoscope, and tongs.	The number of fingers and objects correctly named; scoping range is 0 to 4.
Ideational Praxis	The participant is asked to pretend to send a letter to themselves: fold letter, put letter in envelope, seal envelope, address envelope, and put a stamp on the envelope.	Scored from 0 to 5 based on difficulty of performing the five components.

Orientation	The participant is asked the date, month, year, day of the week, season, time of day, place, and person.	The number of correct responses; scoring range is 0 to 8.
Word Recognition	The participant reads twelve words aloud, and then these twelve words are randomly shuffled with twelve new words, and the participant is asked whether they have previously seen each of the twenty-four words. Three trials are performed.	Mean number of correct responses across the three trials; scoring range is 0 to 12.
Remembering Test Instructions	The rater provides an assessment according to the number of times that the participant needed to be reminded of instructions for the Word Recognition task.	The rater provides a score from 1 to 5
Comprehension	This task also relies on the ten minutes of open-ended conversation. The rater provides an assessment of how well the participant can understand speech.	The rater provides a score from 0 to 5.
Word Finding Difficulty	During the aforementioned open-ended conversation, the rater assesses how much difficulty the participant has in finding desired words.	The rater provides a score from 0 to 5.
Spoken Language Ability	After the administration of the Word Recall task (Q1) ten minutes of open-ended conversation occur between the rater and participant, before the remainder of the tasks are presented. These ten minutes of conversation are used to assess language ability.	Quality of speech is given a global rating by the rater that ranges from 0 to 5.
Maze Task	The participant is required to find the path to the exit of a drawn maze from a start point. This task includes an example/practice maze to familiarize the participant with the task and the test maze, which is scored.	Scores calculated on the number of seconds to complete the task and/or when two errors are made and the task is stopped. Thresholds

		relating to the score of 0-5.
Number Cancellation	A sheet of jumbled numbers is presented to the participant and they are informed (with a demonstration) that they must cross out all of a certain number and to keep going until instructed to stop. This tasks last for 45 or 60 seconds.	Scores are calculated using the number of targets hit minus the number of errors minus number of times reminded of the task instructions. Scores between 0-5 with threshold scores for each point (0-5).
Total Score	Summation of all domain scores. If any are missing or incomplete then this results in the ADAS-Cog not being totalled	Total score of 0-90

Mini Mental State Examination (MMSE)

The Mini Mental State Exam (MMSE) is the most widely used test, both in clinic and research, to screen for dementia. It is a thirty-item questionnaire comprised of very rudimentary components that include; orientation to time, orientation to place, registration, attention & calculation, recall, language, repetition and complex commands (Folstein et al., 1975). These can be mapped onto Awareness/Orientation, Language, Attention, Working Memory, Delayed and Immediate Memory and Executive function cognitive domains because of their inherent paradigms. Currently the MMSE is still under copyright but it has been widely used in western medicine as a quick screen for basic cognitive impairment, as it only takes approximately 5-10 minutes to administer and score. It was also officially translated into ten languages in 2010 (French, German, Dutch, Spanish for the US, Spanish for Latin America, European Spanish, Hindi, Russian, Italian and Simplified Chinese) but was widely used across the world prior to this. As described by Folstein and colleagues (1975) the eight domains of the MMSE were designed to “thoroughly cover the cognitive realm” whilst excluding questions around mood, abnormal mental experiences and the form of thinking, all of which were deemed to not be “cognitive aspects of mental functions”. The other key reason for the inception of this measures was reducing the time for administration due to “elderly patients cooperate well for only short periods”. The measure itself has barely changed since its inception in 1975.

International Shopping List (ISLT)

The ISLT is an episodic memory test used to find impairment likely due to AD or MCI due to AD (Thompson et al., 2011). It is fundamentally a verbal list learning test with individuals trying to memorise 12 common food items that are read out loud to them by a rater. These items are language and country specific, with over 90 specific groups of words within the test. The words are

chosen from a larger word bank of 30 words that has undergone a formal validation process for that language and country. The test itself is made up of an initial immediate recall condition, which is made up of 3 trials. A 20-30 minute delay then occurs (with non-verbal cognitive testing allowed), that is then followed by 1 trial of delayed recall of the word list read out at the start of the test. The ISLT is administered using a computer program, enabling the reading of the shopping list items to the participant by trained staff at a consistent pace. The measure itself contains 2 key outcome measures of immediate recall (sum of all 3 trials) and delayed recall (raw score of the 1 trial). The scores are automatically computed using Cogstate's automated and secure data processing server and is also scored against age- and gender-based norms, providing unbiased and immediate results.

Cogstate Brief Battery (CBB)
The Cogstate Brief Battery (CBB) is a brief, computer administered cognitive test battery that takes around 10 minutes to complete. It consists of four cognitive tasks that measure psychomotor function, attention, working memory and memory. The administration, scoring and reporting is automated and standardised. Each task is constructed using playing cards as stimuli with the participant required to answer only "yes" or "no" on each trial in accord with a simple rule. These four tasks have been widely validated as a battery of cognitive tests (Maruff et al., 2013; Darby et al 2012; Lim et al., 2012; 2015; Stricker et al., 2020).

The first test is called Detection (DET). It is a simple reaction time test shown to measure psychomotor function. In this task, the participant must attend to the card in the centre of the screen and respond to the question "has the card turned over?" Individuals were instructed to press the "Yes" button as soon as the card turns face up. The face of the card is always the same generic joker card. The task ends after 35 correct trials have been recorded. The primary performance measure for this task was reaction time in milliseconds (speed), which was normalised using a logarithmic base 10 (log 10) transformation.

The second test is Identification (IDN). This is a task of choice reaction time shown to measure visual attention. In this task, the participant must attend to the card in the centre of the screen, and respond to the question "Is the card red?". Individuals were required to press the "Yes" button if it is and the "No" button it is not. The face of the cards displayed were either red or black joker cards in equivalent numbers in random order. These cards are different to the generic joker card used in the DET task. The task ends after 30 correct trials.
The primary performance measure for this task was reaction time in milliseconds (speed), which was normalised using a log 10 transformation.

The third is a One Card Learning (OCL) test. This is a continuous visual recognition learning task that assesses visual learning within a pattern separation model (Yassa et al. 2010). This task the participant must attend to the card in the centre of the screen and respond to the question “have you seen this card before in this task?” If the answer was yes, individuals are instructed to press the “Yes” button, and the “No” button if the answer was no. Normal playing cards were displayed (without joker cards). In this task, six cards are drawn at random from the deck and are repeated throughout the task. These four cards are interspersed with distractors (non-repeating cards). The task ends after 80 trials. The primary performance measure for this task was the proportion of correct answers (accuracy), which was normalised using an arcsine square-root transformation.

The final test is a One-Back (ONB) test of working memory and attention. Similar in presentation to the OCL task, individuals must attend to the card in the centre of the screen and respond to the question “is this card the same as that on the immediately previous trial?” If the answer was yes, individuals were instructed to press the “Yes” button, and the “No” button if the answer was no. The task ends after 30 correct trials. The primary performance measure for this task was the proportion of correct answers (accuracy), which was normalised using an arcsine square-root transformation.

The overall battery can be given an overall score by averaging the four individual *z*-scores produce by the automated scoring system. There are also other composite measures within the battery grouping the first two and last two measures. These form an attention and memory composite respectively. However, the main outcome measure is the CBB average *z*-score across all 4 measures.

Weschler Memory Scale – Logical Memory (WMS-LM)

The WMS-LM is part of the larger WMS battery. The logical memory is a measure of story recall. The WMS-LM had three different stories, one short form and two long form stories. For individuals that were aged under 65 both long form stories were used. This began with a trained rater reading out the first long form story in full, the individual then had to recall as much of the story as they could. This process was then repeated for the second long form story. Between 20 and 30 minutes then elapsed, without any intermediary cognitive testing, and individuals were asked to recall as much of the first story as they could remember. If no details could be recalled a set prompt was given and an opportunity was given for recall of that story. This process was then repeated in full for the second story. Scores are taken from a sum of both immediate recalls and both delayed recalls to give representative scores for each condition for each individual. For those over 65, the first short form story was replaced by a short form story totalling half the number of sentences (3) of the long form story (6). This short form story was

repeated twice during the immediate recall condition. The rater reads the story, asked for as much information as could be recalled, before repeating the same story again, then again followed by an immediate recall of the same story. The same 20-30 minute delay occurred before a single recall of the short form story (however, no prompt was allowed for this short form story), followed by a single recall of the long form story (this was allowed the same set prompt as given in the younger age group condition). Scores from both components are measured against a standardised scoring index from the WMS. Therefore, scores from either the immediate or delayed components can be used to find impairment, such as that for MCI (1 standard deviation below the mean).

Cohort Heterogeneity

All trials are based on the foundation and core assumption, that included within the cohort being studied, is a homogeneous population of individuals with a common stage of AD. The extrapolation of this, is that the placebo group should decline in a projected manner and by proxy, if the treated group has slower decline, this signals treatment efficacy. In essence, much of the perceived effect of any given drug is influenced by the performance of the placebo. This is why selection criteria of individuals is so critical to ascertaining efficacy and understanding the true decline due to AD pathology. Recent memory clinic-based studies have shown, that broad comorbidities are present in normal elderly, MCI and AD individuals (de Jager et al., 2018). Across both cohorts within this thesis, clinical AD had a larger prevalence of multiple other conditions and nonimpaired individuals predominantly harboured complex pathologies (Ferreira et al., 2020). However, it should be noted that not all individuals had confirmed amyloid pathology. Further to this, post mortem studies have indicated around 90% of confirmed AD cases have mixed pathology by the end of their lives (James et al., 2016). This finding is not surprising due to the vascular pathology, accumulation of comorbid proteins and the interplay of the degeneration of the cortices present as AD progresses.

The homogeneity of an individual cohort therefore has to balance the issue of natural population sampling, which select individuals who have comorbidities, against the need to accurately measure individual disease related impairments. This quandary is evident in most clinical trials within not just AD but neurodegenerative disorders in general. Clinical trial cohorts are among the most highly characterised, with highly stringent criteria above and beyond many academic datasets, leading to comparatively superior homogeneity. However, some limitations still persist.

Within trial cohorts such as the one under investigation there are stringent inclusion and exclusion criteria. These criteria comprise cognitive, functional and medical history questions. Any disorder or comorbidity that is likely to or

has been shown to impact cognition, such as geriatric depression (measured using the Geriatric Depression Scale [GDS]) is exclusionary, this is the foundation to progress a potential participant for more detailed imaging work ups, such as PET and MRI within AD trials. These modalities further discriminate which neurological disorder a participant may have and helps triage suitable individuals for selection into trial cohorts. This method of stepwise screening procedures was followed for the cohort under investigation, giving a highly homogeneous sample for randomisation into the MissionAD studies.

The balance in cohort selection is maintained by in depth neurological exams and MRI scans excluding clinically significant findings that may be related to other pathologies, but some concurrent pathology may be overlooked or not picked up. These are any minor small vessel disease state findings that can influence cognitive function. One such MRI finding is cortical microinfarcts from vascular pathology within the cortex. They are found in patients with vascular dementia (62%), Alzheimer's disease (43%), and demented patients (33%) compared with nondemented older individuals (24%) (Brundel et al., 2012). To note, recent findings have shown that these vascular pathologies are only visible in 27% of cases at 3T (Van Veluw et al., 2015a; 2015b) and the prevalence of these cortical microinfarcts has been found in 18% cases in a memory clinic setting (Ferro et al., 2019). As minute foci, they are shown to cause neuronal loss, gliosis, pallor, or more cystic lesions and as such, these small cortical microinfarcts in large numbers could cause cognitive impairment (Skrobot et al., 2016; Brundel et al., 2012; Soontornniyomkij et al., 2010; Smith & Beaudin, 2018; Smith et al., 2012). Furthermore, as they are found in all brain regions, possibly more so in the cerebral cortex, the impact of their cognitive impairment could be highly diverse.

Most commonly clinical trial sites tend to have access to scanners at 1.5T-3T, which will not elucidate these findings based upon these aforementioned studies. Even with central readers, the image clarity is such that it will not be optimum due to the dearth of field strength available. As such, undetected, this pathology may induce unwelcome heterogeneity into cohorts such as those in a clinical trial setting. These cortical microinfarcts could be resultant of other neuropathology and be the hallmark of a divergent disease course (Brundel et al., 2012; Soontornniyomkij et al., 2010; Kalaria, 2016; Jellinger, 2005; 2007; 2013; Skrobot et al., 2016). However, this should be overcome in part by amyloid assessments. Dependent upon the stage of the disease these cortical microinfarcts should only be relevant to later stage disease individuals who have the amyloid pathology as well as these and may only relate to an inflation in screen fails where the participant has cognitive decline due to cortical microinfarcts and result in amyloid negativity when this is assessed.

Microbleeds are a further complication that could be the effect of a cognitive impairment not due to amyloid. However, they are seen in 18-29% of confirmed AD cases (Loitfelder et al., 2012). These are again picked up to a lesser or greater degree based upon the field strength of the scanner in question but are caught in a far higher instance of cases and do not rely upon 7T field strength for detection (Greenberg et al., 2009). Along with convexal subarachnoid haemorrhages and cortical superficial siderosis, microbleeds are commonly caused by cerebral amyloid angiopathy (CAA), causing leakage and inflict cognitive impairment (Beitzke et al., 2015). As this is an alternative disease pathology, this rules individuals out of participation in clinical trials for AD but is worth noting as concordant findings that may not have been as well discriminated in prior cohorts.

White matter hyperintensities are the final current concomitant finding on MRIs that have been shown to influence cognitive function. They have been shown to greatly effect processing speed of tasks, argued to be dependent upon their focalisation in the cortex (Duering et al., 2014). Regardless of aetiology of the white matter hyperintensities, they have been shown to affect the clinical expression of AD (Puzo et al., 2019; Jagust et al., 2019) and contribute to cognitive heterogeneity in the early stages of disease (Lee et al., 2016; Yoon et al., 2013; Delano-Wood et al., 2009; Libon et al., 2010).

Further to these cortical microinfarcts, other vascular changes, such as those that can be detected on MRI, affect most patients with AD, the threshold of significance is something that is an unknown, as well as the interplay between these vascular pathologies and the degree of presence of amyloid. These can be present without the latter proteinopathy and as such can confound any cognitive impairment. With around 30-40% of individuals with significant cognitive impairment not meeting the threshold for amyloid positivity (Roberts et al., 2021), this could be argued to be a significant factor underlying the cognitive impairment seen. However, due to the absence of conclusive relationship between these MRI findings, due to their lack of discoverability and the high rates of comorbidities in this age group, it is difficult to conclude on a rationale for cognitive impairment in an amyloid negative population.

Despite the lack of knowledge of the source of the cognitive impairment, this could be argued to be conclusively different to that of AD, yet share some fundamental constructs. This is something to consider when understanding outcome analysis with multi-country clinical trial cohorts, as whilst each one will have a standardised imaging protocol, the field strength is variable depending upon machine availability in each site location. Thus, having the potential, in some individuals, to lead to undiagnosed vascular pathologies contributing to amyloid/amyloid negativity rates and individuals could display similar cognitive impairment profiles but distinct uncaptured neuropathologies. Individuals may have tau positivity but not amyloid, this

has been shown to a greater degree and could reflect AD or alternative pathology (Jack et al., 2018). For the premise of this research and in line with ATN research guidelines on AD, individuals who has confirmed amyloid pathology will be the only individuals who are designated as AD positive.

Discrepancy Score Analysis – Chapter III

Encompassing the need to use composite measures within AD, the aim of the initial analysis chapter is to understand the concordance between memory impairment and that of a discrepancy measure based upon intra-individual change. This will be investigated across the AD spectrum in a stage-wise model based on CDR-SB scores similar to the Jutten et al (2020) paper. This will be examined through a number of measures with particular focus on constituent component scores and discrepancy measures computed from the cognitive battery from the initial stages of the program; MMSE and CBB, with comparison drawn against the ISLT measures.

As has been shown across multiple studies, amyloid positivity rates vary dramatically based upon, age, apoe status and demographic factors. One alternative methodology to use these constituent component scores from the key cognitive measures (whilst maintaining differentials between domains) is to look to the field of IQ. The WAIS commonly employs discrepancy measures to give a picture of a participant's intelligent quotient or subset of skills. This technique has been sparsely used outside of IQ. There is some evidence that these discrepant measurements could be beneficial at identifying individuals who have AD (Jacobsen et al., 2002). Discrepancy scores will be computed from the cognitive battery within the first cohort and will be sought to give an understanding of a participant's cognitive profile.

Combing measures allows for a reduced impact of range restrictions, improved temporal reliability and alleviates the issue of statistical multiplicity within domains. Clearly there are insurmountable issues with using current measures alone, within earlier stages of the disease. So, looking within measures, within domains and differential scores to the global/overall scores is widely needed in AD. There is inherent debate in these composite measures is in the fact that picking apart measures, which have shown prior efficacy, has been argued to be a form of manipulation of the efficacy measures (Jin et al., 2018). Contrary to this, it is argued that this has given greater nuance to our understanding of the disease, it's trajectories and the cognitive impairments this entails across the differing stages of AD. Nevertheless, the measurements have not changed at all. It is difficult to argue to maintain the usage of a measure, which was once used to measure a population that is heterogeneous now, as the difference in the population characteristics to one that would be recruited now are striking. However, the regulatory authorities and indeed, physicians maintain these

measures due to their prior approval history, quickness and ease of interpretation.

Fundamentally it can be argued that the MMSE, CDR & ADAS-Cog all function as composite measures in and of themselves. As they have measures for individual cognitive and functional domains but also give an overall score, they operate in much the same way. However, they were not all developed based upon statistically combining tests to yield a single measure, but were developed to index multiple domains broadly yet shallowly. Much like IQ an MMSE, CDR and ADAS-Cog give a global/total score for participant, which is prodigious for classification but as with most things, the devil is in the details. At different stages of AD different cognitive function become impaired (Jutten et al., 2020). As Mortamais and colleagues (2017) have shown, semantic and episodic memory components are the first to become impaired and deficits sensitive to decline at a later stage (such as executive functions) do not decline in the early stages of AD. However, all of the aforementioned measures to a greater or lesser degree compress all cognitive domains into these global scores. By selecting component measures of indices that show greater sensitivity to decline whilst removing identical and/or confounding variables, this allows for increased sensitivity not only to the disease course but also the amelioration of the modelled level of disease impairment.

It can also allow for single timepoint measurement when indexing a decline from a prior level. By utilising an intra-individual measure based upon constituent cognitive functions, that are shown to have very little age-related change, a comparison can be drawn against those that do have age related change such as memory. Therefore, a performance whereby a language performance below those of processing speed or executive function indices would indicate a performance below what is expected. The measure of fluid ability as with prior work will comprise commonly used cognitive domain measures for executive function, attention and reaction time. This will be compared to a crystallised intelligence score which will be computed from the five language domains of the MMSE (totalling 8) and z-scored against the full cohort. The discrepancy score will be in the form of a z-score for each subject and will be calculated using the following equation: fluid composite – crystallised composite = discrepancy score.

The full screening cohort will be used for this analysis which totals nearly 10,000 individuals. All will have all presented to a memory clinic, primary care physician or clinical trial site with some concern about their cognitive function, primarily memory issues. However, a large minority will not have any memory impairments (~40%). In order to compare these measures to those typically used for screening for symptomatic AD, the relationships between these measures will be analysed, using simple correlation and regression methods, as well as group level comparisons based upon CDR

staging. These methods have been chosen to be used as this is a novel approach within this setting and as such, there is no basic information about these measures within a screening population for AD. Thus, in order to discover how a discrepancy score performs in relation to that of common memory measures, comparisons will need to be made in different forms, both on an individual and group level.

Comparisons Of Verbal Memory Measures – Chapter IV

AD is incredibly heterogeneous, even when looking at those who are amyloid positive, 30% of cognitive normal individuals will never go on to develop AD dementia (Jack et al., 2019). There are also large differences between clinically and biologically defined AD. With biological AD (aka amyloid positivity) three times more prevalent at any age than the clinical syndrome. This is in line with the understanding of the 15-20 years preceding the onset of symptoms within AD (Jack et al., 2018). However, whilst there is a degree of alignment, the argument persists that amyloid is not a prerequisite for developing AD dementia. It is only once tau pathology becomes more abundant does any clinical decline occur. This current understanding is far from crystal clear. Nevertheless, what is understood is that pathological amyloid persists for decades prior to symptom onset, with pathological tau accumulating subsequent to the presence of amyloid pathology (Frisoni, et al. 2022).

The sources of heterogeneity are being widely studied from a biological perspective, however, little consistent insight is offered into cognitive heterogeneity within AD (outside that related to biology). This is primarily due to the lack of agreement into which measures to use. All of which has resulted in a very siloed approach when it comes to clinical trials and academic cohort studies. Resulting in over 50% of measures used for finding treatment effects in clinical trials have incomplete psychometric validity published on them (Soobiah et al., 2019). The premise of a lack of comparative studies being if the study is using one measure to index episodic memory for an individual why put them through a second similar/identical measure. Whilst most researchers will not expect a 99% concordance for results between any two measures, there is very much a fundamental trust that because two measurement paradigms are similar, they will produce similar results. This chapter will look to ascertain this validity of this statement using the two different cohorts. Firstly, using two measures of story recall from a new screening population who may or may not have a memory impairment. This will look to understand variability within a population presenting to memory clinics for screening with unconfirmed pathology or diagnosis. The second will look at the confirmed amyloid positive individuals within the large cohort discussed earlier. This analysis will look specifically at those with confirmed clinical AND pathological AD and how two measures of verbal list learning differ in this population. In both

analysis these measures are assumed to be highly concordant and these analyses will look to uncover if they indeed are.

For diagnosis there is little, if any, agreement on which measure best indexes AD (amnestic) related cognitive impairment. This is shown abundantly in the ambiguity in all diagnostic classifications since 1984 (see **Chapter 1** for full discussion). Because of this discordance in test selection, these different measures are assumed to be closely aligned in terms of the cognitive domains they are indexing. Cognitive variability within AD is something that is commonly found on the MMSE (Duchesne et al., 2005), with 2-3 point variations between scores found with those diagnosed with MCI. However, when looking at domain specific measures, such as those used to measure memory impairment, these variations with repeated testing are often more stable (Patton, et al., 2005; Kueper et al., 2018).

One commonly employed method of comparison is that developed by Bland and Altman (1978). Whereby two measures are analysed to see how closely aligned the two measurements are to the same thing (or in this case the individual). The means and differences between the two measures on an individual case by case basis are computed and graphically represented. Limits of agreement are then calculated using standard deviations to give a representation of the two measures under investigation. The closer the mean score is to zero the better. And the closer the limits of agreement are to one another the more comparable the measures are. This simple yet comprehensive technique allows for concordant interpretation of both group and individual level data.

This has been studied within prodromal AD/MCI using the broad Cantab battery and other common AD measures; ADAS-Cog, FCSRT, CDR, MMSE (Abbott et al., 2019). This was analysed using the phase III trial of gantenuramab (SCarlet roAD), whereby repeated timepoints of the same measures were looked at to understand the relationship between these domain specific measures and those commonly used within trials of AD. Relationships between the measures were poor (r=0.1-0.4), however when looking at the consistency of these domain specific measures the limits of agreement for most measures, as well as the Cantab memory composite, were small. Thus, indicating good reliability across measurement timepoints for the battery.

As the measures are being compared to one another as a whole rather than individual component comparisons, some typical psychometric analysis methods (such as Cronbach’s alpha) are not fit for these analyses. One method that is similar though, is that of equivalence testing. This uses inferential confidence intervals to test for statistical equivalence between two measures as an adjunct to null hypothesis significance testing (NHST). This was shown within psychometric testing of healthy adults comparing two

forms of the Controlled Oral Word Association Test (Ross et al., 2007). Typical descriptive confidence intervals empirically can't capture the same amount of variance as NHST to 5% threshold (Tyron, 2001; Nickerson, 2000). Therefore, this method uses inferential confidence intervals to better capture the variance between two measures, which is the core aim of this analysis. However, whilst this methodology is more sensitive to group differences than NHST, the Bland & Altman method also allows for visualisation of individual level differences too. And when coupled with other methods such as NHST gives a broader picture of measure performance. It is also preferable to equivalence testing on the basis of prior use within AD and coupling with correlation analysis as well as NHST will give the broadest level of understanding of the performance of these two pairs of measures within the populations under investigation. To note the same analysis will be performed for both studies within this chapter to allow for a comparison of results.

Classification Methodologies – Chapter V

Whilst AD clinical diagnosis is based upon clinical judgement and doesn't always align with pathology, the significant presence of amyloid is irrefutable evidence of the biological hallmark of this neurogenerative disorder (Mckhann et al., 1984). Amyloid positivity assessment is often not done in primary care due to cost, time and experience of physicians (James et al., 2020). Utilising new techniques, such as machine learning to find ways to use common measures used in AD settings could help alleviate misdiagnosis, participant burden and allow individuals access to potential treatment and drug trials in a timelier manner. Even by improving the classification rate of amyloid positivity in a clinical trial context, can results in massive cost savings as well as increasing the speed of recruitment into these trials, which is beneficial for the companies, individuals and AD field alike.

A lot of work has been put into machine learning techniques using these more complex methods to increase the classification accuracy of diagnosis. Being able to predict a biological phenotype of AD from a clinical presentation is something that is broadly sought to help align the two diagnostic frameworks and aid overall diagnosis. This is primarily in settings where routine PET imaging, blood tests and CSF lumbar puncture are not widely available due to cost, physician knowledge or healthcare system offerings. However, predicting amyloid positivity is still as challenging as ever, regardless of the testing measure employed, no clinical or cognitive measures have been shown to correctly categorise amyloid positive individuals prior to PET or CSF measurements. In recent years with the spread of machine learning techniques into fundamental neurosciences, researchers have begun to apply this methodology to clinical trials as well as academic cohorts. Building models to predict this classification is one method of repurposing existing data to aid future trial design.

Prior research has shown there is no one size fits all approach for choosing a classification methodology. Machine learning methods are more sophisticated statistically but standard machine learning methods provide predictions and associations without necessarily having biologically or clinically grounded casual insights of the outcomes. This can lead to unbalanced models and skewed positive or negative classification values. Many have been used, with methods used expanding with the growing use of machine learning techniques and artificial intelligence paradigms. In a recent review of 60 papers within AD using some form a machine learning for classification, Tanveer and colleagues (2020) showed that support vector machines (SVM) is used in the majority of papers (mainly looking for differences between CU and AD individuals) and in 83% of papers when imaging modalities were the main focus. This paper showed a preponderance for method selection at the behest of the authors rather than supportive literature within AD or dementia. As the aforementioned studies all show a range of methods that have been previously used, with a number having consistent strong levels of accuracy in classification within AD.

However, the literature is far from clear cut. Bansal and colleagues (2018) conducted a literature search of four machine learning algorithms within dementia diagnosis and found decision trees and naïve Bayes classifiers to have >99% classification accuracy before and after attribute selection (reduction). Bayes' theorem is fundamental for inferential statistics and many machine learning models. Bayesian reasoning is a logical approach to updating the probability of hypotheses in the light of new evidence, and it therefore highly applicable to many scientific hypothesis' (Berry & Stangl, 1996). Whereas the aforementioned methods look towards regression and variance, this takes a probabilistic approach to the likelihood of groupings. This classification method has also shown strong classification properties in neurodegenerative disorders and dementia in particular (Bansal et al., 2018; Khan & Usman, 2019). However, its utilisation within AD specifically is still unclear.

Further to this, Bucholc et al, (2019) showed additional support for SVM, using ADNI-2 data. Across 6 machine learning methods, SVM had superior classification when given a wide range of functional, imaging and cognitive variables. Concurrent research by Khan & Usman (2019) showed again Bayes based classifiers improve clinical AD diagnostic accuracy up to 96.4%. However, a comparison of ten traditional and non-parametric classifiers for the prediction of dementia showed a different pattern (Maroco et al., 2011). PCA-LDA showed leading sensitivity and specificity. Further support for this methodology within AD comes from imaging results showing high classification of SPECT and PET images utilising PCA and LDA, whilst SVM showed high classification values after the PCA was employed on the image identification (Salas-Gonzalez et al., 2010; Lopez et al., 2009). SVM, neural networks and decision tree methodologies have shown further favourable

classification properties, but whilst having high specificity (SVM Me= 0.9 AUC) they had low specificity when compared to other methods (SVM Me=0.3). In a recent comparison paper Grassi and colleagues (2018) also showed superior accuracy of SVM for diagnostic classification based upon cognitive, functional, demographic and imaging variables. However, whilst their AUC of 0.96 was supported by a bootstrapping AUC of 0.92, their sample size was very limited for a study of this size (n=184) impairing its wider generalisability. Something this analysis will not be limited by.

Conversely, machine learning methods haven't always shown superior accuracy in diagnosis or classification more broadly. A recent ADNI paper looking to predict AD from genetic data showed a best of 72% classification performance from the Random Forest method (Oriol et al., 2019). This can be argued to be down to the input into the model (only genetic markers rather than more detailed variables previously shown) rather than the models themselves. A further paper from the DIAN cohort in preclinical AD, also showed poor accuracies (AUC=0.74) utilising PiB-PET features alone in an SVM and k-mean clustering model (Castillo-Barnes et al., 2020). A further ADNI paper (Ezzarti et al., 2019) also looked at these models and found SVM and k-nearest neighbour models performed worse when all features were included. They found model performance is partially dependent on feature selection and characteristics of the dataset. Furthermore, this method gives an alternative yet unlimited hierarchical way to treat the data within the cohort. As shown in the Sampson and colleagues' paper (2011) SVM have strong utility within classification problems with many variables (greater than the number in this cohort) and whilst this is not identical it is synonymous in its approach, other comparative analyses as discussed previously have shown SVM to have high classification results within AD (Bucholc et al., 2019).

Depending upon the number of variables in question, classification can either be used in full by the model of a factor analysis/principal components analysis or can be conducted to select only the appropriate variables. This is the traditional method of ascertaining classification and latent factor within datasets which is fundamentally the hypothesis of this analysis seeks to determine. Comparisons between PCA then LDA and PLS has shown model parsimony in classification performance, however, the key differential is in the feature reduction aspects of these methods. PLS is particularly well suited to analysing a large array of related predictor variables (Kettaneh et al., 2005; Carrascal et al., 2009) which is not specifically true of this dataset. Again, it is best suited with a sample size not large enough compared to the number of independent variables. PLS-DA was the method used within the initial ADCOMS development paper (Wang et al., 2016), which has been widely successful within clinical trials.

As within other biological disciplines, machine learning in AD has been predominantly used to index imaging and genetic modalities for classification purposes or to improve signal detection. Computer Aided Diagnosis (CAD) has aided in classification studies but has not been widely adopted, models for SUVR values in PET however, have shown promise, but the majority of clinical trials still use central readers and comparisons between the two methodologies has yielded highly congruent results. One recent study of importance was from the Coalition Against Major Diseases (CAMD), who developed an unsupervised machine learning model called a Conditional Restricted Boltzmann Machine (CRBM) to simulate detailed patient trajectories (Fisher et al., 2019). This work demonstrated the ability of a machine learning model to accurately and consistently predict changes in total ADAS-Cog score for 18-months of decline. This neural network treated the decline as a latent factor and predicted with high accuracy individual patient decline and is an example of how this can be applied to future clinical trials. However, given the cross-sectional nature of this proposed analysis this method will not be employed.

Overall, given the breadth of its use and fairly consistent top of the class classification values, the method to be employed within this analysis will be SVM. This machine learning based approach negates the need to treat outliers with removal and allows for misclassification within the model. Nevertheless, whilst SVM is a sophisticated classifier, it is somewhat prone to overfitting. This can be overcome using soft margin classifiers to allow for multiple misclassifications for individual variables. But in order for any model to be found, there is a need for a training and testing datasets. As this dataset comprises such a large number of individuals, the cohort will be split into 2 equal datasets. After the randomised data split, the cohorts will be limited to the key variables in questions which will be derived from the results of the prior analysis within this thesis. As with all classification problems the model bias/variance trade off will be monitored closely and looked at in detail for each result. Within the dataset the classifications for all variables are labelled. However, in order to train the model/algorithm semi-supervised learning methodologies will be employed, using tuning parameters within the SVM model function. Initially the SVM models will be built/trained on a labelled dataset to employ supervised learning algorithms before testing and validating this on subsequent unlabelled datasets.

Chapter Summary

In summary this chapter has covered the broad outline of the techniques and measures to be employed in all main analyses within the thesis. It has also covered in detail the nature of the dataset under investigation, the screening procedures for all individuals that has led to these data points and how these tiered procedures impact the homogeneity of the cohort in question. Moving forward the next chapters will all begin the analyses, starting by exploring

the relationship and feasibility of a discrepancy score within this clinical trial cohort.

Chapter III – Discrepancy Scores Within AD

Chapter Outline

Finding cognitive impairment in AD is typically done cross-sectionally. This method has many flaws as measuring neurodegeneration/AD requires a change from normal function. Discrepancy scores can find a normal function of an individual, cross-sectionally, by comparing current performance to stable cognitive functions such as language ability. This chapter will explore this concept within a very large cohort comprised of individuals with AD and explore if these measures can be derived from short global cognitive assessments. The study will examine the how this measure behaves throughout AD (criterion validity) and how it compares to that of the established memory measure within this cohort (convergent validity). This chapter will also explore how well the full battery of measures perform across regions (cross-cultural validity).

Introduction

Memory impairment is a cornerstone of the diagnosis of Mild Cognitive Impairment (MCI) likely due to Alzheimer's Disease (AD). Currently, this is understood to be the first outward cognitive symptom of this disease. Finding impairment in subjects likely to have MCI or AD is either done cross sectionally or longitudinally. Longitudinal measurement is the gold standard but something that is rarely available when screening subjects prior to clinical trials. Cross sectional measurement is commonplace and is compared to normative samples or within individual comparisons using contrast or discrepancy scores. The latter of which will be explored further in this chapter.

Fundamental to clinical trial success is the reliance on individual trajectories but they look for changes based upon reported functional change from a baseline score. These baselines can be highly changeable on a day-to-day basis, thus one score may not be representative of a normal level of function for an individual. Utilising a discrepancy measure for an individual allows for a greater level of precision on an individual's level at a given timepoint, which is always in relation to their function and their potential change in function. But is importantly, measured in a different way to the commonly used measures such as CDR or ADAS-Cog, which are both based upon overall levels/scores that may mask any changes across singular cognitive domains.

Tests of cognition are primarily designed to index a singular cognitive function or domain. Their construct validity, is often questionable, as most of these "domain specific" measures place demands on other cognitive domains at the same time (Soobiah et al., 2019). As an example, a measure

wanting to index processing speed consists of something as simple as eliciting a response/s from an individual to collect a measure processing speed, but requires attention as well as comprehension of verbal or written instructions. To better understand the domain specificity of a measure, this is sometimes done concurrently with neuroimaging. This helps understand the locality and brevity of cortical exertion undergone during a cognitive test. Whilst each region or network has a preponderance for a singular cognitive function or domain, there is a dominance for certain brain regions to be involved in specific tasks. For example, the Dorsolateral pre-frontal cortex (DLPFC) is highly involved in decision making, goal-directed behaviours and higher order thinking, termed executive functions (Elliot, 2003). With smaller regions such as the right inferior frontal gyrus have been strongly implicated in inhibition (Hampshire et al., 2010). These regional distinctions are important when looking at neurodegenerative disorders as in many they tend to follow specific patterns of degeneration and dysfunction. Understanding this better can help lead to earlier detection and aid treatment decisions earlier in their disease process.

Specifically, within AD the entorhinal cortex then the hippocampus are the areas thought to be initially affected by beta-amyloid and subsequently tau aggregation. These build ups of pathological proteins then lead to biological deficiencies in signalling pathways in these areas (Huijbers et al., 2014; Maass et al., 2018; Jack et al., 2018). The downstream effects of these dysfunctions result in impairment in the cognitive domains associated with these cortical regions. Within typical AD, longitudinal studies have robustly shown that episodic memory declines first, followed by executive function and processing speed later in the disease course (Nathan et al., 2017; Mortamais et al., 2017; Boraxbekk et al., 2015; Grober et al., 2008; Mistridis et al., 2015; Schmid et al., 2013). The deposition of these abnormal proteins is thought to commonly occur decades before the start of the decline of these cognitive functions. Current methods of identifying the early stages of AD, predominantly rely on biomarker measurements that are currently expensive, invasive and not always accessible by all communities and regions (Sperling, Jack, & Aisen, 2011). An alternative to biomarkers (in the absence of a blood test) to diagnose AD, is to develop an earlier cognitive marker for clinical diagnosis. This would measure underlying cognitive dysfunction and accurately predict future cognitive and functional decline, as a result of AD pathology (Mortamais et al., 2017). The current guidance from regulators and academic trends has seen cognitive endpoints recommended to be personally relevant and related to an individual’s premorbid level of function (Rentz & Weintraub, 2000; Weintraub et al., 2018; FDA, 2018). These key criteria are the foundations of an individual discrepancy score.

Discrepancy Scores Within AD – Fluid & Crystallised Composites/Differences

Broad and time-consuming neuropsychological batteries, such as the RBANS, WAIS and CERAD are often composed of a number of different measures that index diverse cognitive domains. These are primarily administered as a whole battery and give an in-depth outline of a subject's level of cognition, across domains. As part of these batteries an overall score is computed as well as domain scores of similar cognitive functions, but importantly these batteries also compute difference scores or discrepancy measures for each subject. Discrepancy measures or contrast scores, compare specific cognitive functions against each other. These are usually in the form of standardised *z*-scores in order for the comparisons to be drawn.

Recent evidence has suggested that a discrepancy between language skills and other cognitive domains (most notably episodic memory) & executive functions is found in early AD (McDonough & Popp, 2020). The dichotomy of these skills is broken down into fluid ability (memory, executive function and general cognitive skills) and crystallised ability (language skills). Indexing fluid ability gives an approximation of the efficiency of a cognitive domain/s and encompasses a raft of constituent cognitive processes (Wechsler, 1944, 1997). Tasks such as these are implicitly constructed to minimise usage of prior learned skills or knowledge (Johnson et al., 2004; Kaszniak, 1986; Lezak, 1995; Wechsler, 1944). Measures designed to index crystallised ability are created to explicitly index exactly those prior skills and knowledge learned through experience and education. This is routinely measured through vocabulary (Ekstrom et al., 1976; Wechsler, 1944; Zachary & Shipley, 1986) or less commonly, word pronunciation tasks (Blair & Spreen, 1989).

Results from this recent research (McDonough & Popp, 2020) indicated that greater amyloid deposition in early AD-related regions (precunus, temporal cortex and posterior & anterior cingulate) are associated with a larger discrepancy score. This study also showed amyloid deposition was significant and independently associated with a greater positive ability discrepancy score and these relationships were significant in both positive and negative discrepancy groups for this amyloid deposition. The rationale for comparing and contrasting these two domains within AD is that on an individual level, crystallised (language) abilities should enforce a ceiling on fluid (broad cognitive) abilities and thus measuring the discrepancy on an individual level could give a better indication of a decline of that subject over a comparison to a control population (Cattell, 1971; Kaufman et al., 2009). Those control group comparisons when appropriately administered and adjusted for age, education and other necessary demographic factors, is not always something that is available. Moreover, when this is available it is commonly not culturally specific, ecologically valid and doesn't account for educational attainment in most cases. The alternative within individual comparison using discrepancy scores have shown recent promise within AD.

Within the majority of cognitive neuroscience, it is expected to have educational attainment in one form or another as a covariate. This is done to account for underlying level of general cognitive ability. However, this is somewhat a nominal figure that rarely pertains to individual function and more a grouping based upon educational attainment level (e.g. High School, University or Post-Graduate). Educational attainment has long been mooted as a protective factor in the development of AD (Stern et al., 1994). This is commonly referred to as a cognitive reserve that is normally indexed via years of education or via a NART/IQ measure. However, these measures have a prolonged assessment timeframe and are not part of any clinical trial within AD. As such years of education is captured as a proxy measure. Years of education has been shown to correlate strongly with average ability discrepancy within a study of the Harvard Aging Brain cohort (McDonough & Popp, 2020). Therefore, it is expected that fluid ability declines prior to that of crystallised ability (O'Carroll & Gilleard, 1986; Wechsler, 1944). This also fits succinctly into the cognitive reserve hypothesis whereby an individual with larger crystallised abilities (or high IQ's/language performance) delay, disrupt or halt cognitive decline in AD (Stern et al., 1994; Stern, 2006). This differential decline leads to large discrepancies between the two abilities, a rate which has been shown to be indicative of the rate of AD symptom progression (Albert et al., 2011; Bastin & Salmon, 2014; McKhann et al., 2011; Schmid, Taylor, Foldi, Berres, & Monsch, 2013). Any change to the discrepancy score over time is suggested to be indicative of cognitive decline. Whereas, measuring this cross-sectionally may allow for an earlier signal of pathological change in AD, if subjects have a negative discrepancy score (ie lower fluid score compared to a crystallised score).

In comparison with single domain impairment, such as memory, the utilisation of this discrepancy measure has been sparse. However, a growing body of literature suggest this discrepancy differs in those with AD to cognitively normal subjects (McDonough & Popp, 2020) and increases in discrepancy as the disease progresses (Dierckx et al., 2008; Lezak, 1995; McCarthy et al., 2005). Importantly, and in line with the current direction of clinical trial populations, discrepancy measures have also been found to differ in preclinical AD subjects (McDonough et al., 2016). This study showed the difference between these groups with a score below zero, to be associated with greater level of amyloid beta as measured by PET. A score of close to zero or above, is indicative of healthy aging. Further to this, a lower (close to zero and positive) ability discrepancy score also has been associated with psychosocial factors such as, greater social and physical activity (O'Shea et al., 2018). That, in turn, has been associated with preserved functional integrity throughout the aging process and with greater levels of cognitive reserve (Christensen & Mackinnon, 1993; Fratiglioni, Paillard-Borg, & Winblad, 2004; Karp et al., 2006).

Discrepancy Scores Within AD – Validity Of This As A Measure

The findings from studies in early AD looking into discrepancies between fluid and crystallised abilities, still have some pitfalls. These are worthy of discussion and further research before widely implementing them into research and clinical frameworks. As previously discussed, an ability discrepancy relies on a composite of different fluid abilities, bringing together multiple cognitive domains which decline at differential rates within AD can be argued to mask any actual deficit if not analysed as constituent cognitive domains. However, within AD, subtle cognitive declines are not expected to be detectable until very late in the preclinical AD stage (Sperling et al., 2011) or early MCI (Jack et al., 2013). Detecting individual cognitive abnormalities prior to the onset of MCI is the next step in early diagnosis for AD but, by its very definition it is not clear how effective a cognitive marker ultimately would be to detect preclinical AD. However, due to the lack of congruency with AD biomarkers in the initial stages of the disease, the differential rates of decline and lack of concordance across memory measures, these discrepancy measures can be argued to make this new methodology worthy of further exploration.

A common issue with the prior studies on discrepancy measures, is that individual cross-sectional performance covaries with lifelong values of IQ/g/overall intelligence, making the distinction inseparable within analyses of this nature. For example, individuals starting out with a lower level of cognition in any domain might appear to be on the path of cognitive decline, despite no actual change in ability since their youth and thus no evidence of abnormal cognitive decline related to their own ability. Even when accounting for educational attainment in analysis the problem still remains. The primary benefit of a discrepancy score over a single domain measure is that it helps to control for individual differences in overall ability more robustly than proxy measures such as educational attainment. Subsequently serving as a measure for within-individual changes in cognition when no longitudinal assessment is available. Therefore, an advantage of the fluid/crystallised discrepancy measure, over individual domain scores is that this would allow for the ability to control for individual differences in overall ability and thus serving as a proxy for within-individual changes in cognition, something current measures struggle to do. The implication is that it possible, longitudinal studies might have shown different levels of sensitivity of their measures, had they accounted for individual differences/variances due to crystallised ability.

Critical to the validity of these discrepancy scores is the trajectory of the relationship between these two abilities within healthy aging. The dominant theory in this space is the cognitive dedifferentiation hypothesis (Baltes et al., 1980; Reinert, 1970), whereby cognitive abilities become more interrelated in old age. This is thought to be in part due to an increased

reliance on common underlying processes. A recent longitudinal study of this hypothesis has indicated the stability of these abilities in old age (Salthouse, 2010), whereas other research has shown weaker correlations in older adults compared to younger adults (Cunningham et al., 1975; Eisdorfer et al., 1959; Rabbitt, 1993) or younger adults having greater fluid than crystallised abilities whereas older adults have greater crystallised than fluid abilities (Park et al., 2002; Salthouse, 2010). These findings suggest that healthy aging also might be accompanied by an ability discrepancy. Contrastingly, other research has found stronger correlations between fluid and crystallised abilities in old age in relation to younger adults (Cunningham, 1980; Lindenberger & Baltes, 1997; McHugh & Owens, 1954; Reinert, 1970). In essence the age of some of these studies precludes the majority of our understanding of AD by some distance. To this end, these cohorts would have likely contained a raft of subjects with undiagnosed pathology and some subjects would arguably have had preclinical AD. This makes it somewhat difficult to draw any comprehensive conclusions as to the stability of these measures as it is not clear whether fluid abilities longitudinally decline to a faster extent than crystallised abilities in healthy aging. Nevertheless, recent work by McDonough & Popp (2020) showed age was independently associated with ability discrepancy after accounting for cortical thickness and Aβ deposition. Thus, using recent methods and measures, showing the reliability of a measure such as this across the lifespan within the context of current AD research. Furthermore, as the magnitude of these prior studies found discrepancies in healthy aging are small it is argued that any discrepancies that might occur in healthy aging would be much smaller than the differences found in AD, yielding a potential differential early marker of cognition.

The other key component to the validity of the discrepancy score is in the similarities of fluid abilities in early AD. Recently research has looked into this within a cardio-vascular disease population with some subjects having diagnosed MCI. This study found that 74/104 subjects had an ability discrepancy (outside +/- 1 SD) using a composite score of fluid measures, including memory (Takaiwa et al., 2018). They also compared a stable crystallised ability measure to a range of individual (fluid) domains and found the most frequent discrepancies in MCI subjects to occur in immediate verbal memory (66%) and attention (60%) from the RBANS. Whilst these findings are novel, the sample in question had no AD biomarkers and were recruited from a cardio-vascular disease population. Which whilst having links to the onset of AD, should be treated as a separate entity when compared to those at the early stages of AD without any comorbid CV conditions.

Discrepancy Scores Within AD – Individual Test Comparisons

Discrepancy scores as a definition covers a wide base of measures. As such alternative discrepancy measures to those already discussed have also shown some promise within early AD. But before exploring the other studies,

it is necessary to mention the differences in classification of AD now, compared to when these studies were published (see **Chapter 1** for full discussion). As such any discussions and findings should take the lack of biomarker confirmation into account with any interpretation of the following results. It is also worth noting the following studies explore different aspects of potential discrepancy scores in AD using many measures and all of the work is carried out by the group who developed the DKEFS executive function battery. Importantly they found there to be differences between verbal and visuospatial abilities in what was termed “preclinical AD” (Jacobson et al., 2002). In particular with this initial research by Jacobsen and colleagues (2002), they looked at a group of 20 healthy elderly subjects and 20 subjects with what at the time was termed preclinical AD (likely now termed early/prodromal AD). These analyses were taken from data of the subjects who subsequently met diagnostic criteria early AD when they were still considered healthy controls at the time of the test administration. All diagnosis for all subjects were based upon independent annual examinations from two senior neurologists with the subjects classified into the preclinical AD group participating as control subjects for an average of 4.6 years prior to a subsequent change in diagnosis. The clear differences now are the diagnostic criteria and availability of Aβ indices which may lead to Aβ- subjects within the sample. However, with the longitudinal follow-up of these subjects, it was clear that they all went on to have dementia which suggest the presence of neurodegeneration (N) at a minimum.

These two studied groups were matched on age, education and gender, and were assessed on a cognitive battery comprising of the Dementia Rating Scale (DRS; Mattis et al., 1976), CVLT (Delis et al., 1987) long delay free recall, Boston Naming Test (BNT; Kaplan et al., 1983), the Block Design subtest from the WISC-R (Wechsler, 1974) and the WAIS-R vocabulary subtest (Weschler, 1981). The measures index semantic and episodic memory, shown earlier to be key indicators of early disease pathology which was not as well-known at the time. Further to this, the two components of the Wechsler batteries are short yet key indicators of cognitive processes known to be at their peak towards the latter stages of maturation of cognition (Spreng & Turner, 2019). Nevertheless, this cognitive battery maps on very neatly to our current understanding of preclinical AD and pathology. The discrepancy score here was made up of fluid (Block Design, CVLT) and crystallised (Vocab, BNT) components transformed to the same scale. The results from this small study showed statistically meaningful significant differences in the discrepancy (asymmetry) score between these cognitive domains and subject group. Whereas the DRS and CVLT scores also were significantly different between the groups, the size of the effect was not as large as that of the discrepancy measure. Thus suggesting, that as seen with other measures profiling IQ, discrepancy analysis may indicate some of the earliest changes within preclinical AD.

This was further expanded upon within studies of cognitively unimpaired adults with genetic risk for AD (Fine et al., 2008; Houston et al., 2005; Jacobson et al., 2005). All three studies utilised a similar small cohort who were found to be cognitively unimpaired on the dementia rating scale (Matis, 1988). There are a number of issues with using this to rule out impairment due to AD. The initial normative data against which the cognitive impairment is made against is one without AD biomarker confirmation meaning some of the cognitively unimpaired subjects will be preclinical AD and a significant proportion of the impaired subjects will have impaired due to erroneous reasons outside of AD (Coblentz et al., 1973; Lucas et al., 1998). Also importantly the DRS shows greater sensitivity to large cognitive changes in those patients with severe dementia (Woodard et al., 1996). The use of the DRS as a measure of global cognition is argued to be akin but more in depth than the MMSE as it indexes similar domains to that of the MMSE and other common neuropsychological measures (such as WAIS, WMS & WCST; Brown et al., 1999). But due to its larger scoring range gives a more nuanced breakdown of a subject's cognitive performance. It has shown a significant correlation with MMSE of between r=0.78 - 0.82 (Woodard et al., 1996). Further to this, the aforementioned discrepancy studies all looked at different measures to comprise the discrepancy scores. Fine and colleagues (2008) found differences in APOE positive and negative groups in the cognitive switching discrepancies within the stroop task. They also found the discrepancy superior to APOE status in predicting cognitive decline in the sample of 24 subjects. Nevertheless, Housten and colleagues (2005) found there to be a higher proportion of discrepancies within the verbal and design fluency measures from the DKEFS. However, again there was no AD biomarker confirmation within the cohort of 52 subjects allowing for limited relations to early AD detection. Jacobson et al. (2005) also showed group differences for APOE e4 carriers compared to non-carriers in a slightly smaller sample of 42 on a discrepancy measure between digit and spatial span measures.

Whilst it is well documented that general cognitive domains decline in early AD and more rapidly as the condition progresses, there is also some suggestion of this discrepancy score to be dysfunctional in later stage dementia (Strite, et al., 1997). However, whilst the study looked into discrepancy scores, the subjects who were thought to have a diagnosis of AD (McKhann et a., 1984) were excluded and only subjects with dementia were included and were also categorised as Mild AD with scores of 22-30. Whilst it is important to note the finding as some of the study population may have met current diagnostic criteria, as with the prior Jacobsen group studies, the diagnostic classification of the study renders little comparative findings as a result of their study population.

Key to assessing the intra-individual difference between amyloid groups is also to control for genotype within all stages of AD. As it has been strongly

demonstrated, APOE carriers decline at a quicker rate than non-carriers (Corder et al.,1993; Farrer et al., 1997; Tilvis et al., 2004; Rawle et al., 2018) and there is also growing evidence of an absence of decline in those individuals who carry the protective e2 variant of this key AD gene (Li et al., 2020; Zalocusky et al., 2019). Some have even gone as far to suggest a differentiation of AD disease type of the basis of APOE genotype (Frisoni et a., 2022). As shown in Jacobsen and colleagues (2005) these discrepancy scores have also been shown to be asymmetric within these genotype subgroups albeit with a relatively small sample size, this is hypothesised to be due to differing biomarker profiles for these individuals within the spectrum of AD.

Overall taken together, these findings show the possibility that many types of cognitive discrepancies may exist across the AD process, but not all may be equally sensitive to AD biomarkers. However, whilst all studies showed an increased prevalence of asymmetric cognitive profiles in the AD group compared to controls (Jacobson et al., 2002; Fine et al., 2008; Houston et al., 2005; Jacobson et al., 2005) these studies had very low sample sizes and uncharacterised cohorts. This can be argued to lead to spurious findings given that larger samples and more meaningful effects are needed for wider generalisation to clinical and research settings.

Proxies of Crystallised Ability – Language & Vocabulary Measures

Taking a step back from the prior literature, the overall intent of the discrepancy measure is to describe an intra-individual change based upon their pre morbid level of function. Vocabulary and semantic skills are commonly seen as the initial areas impacted with AD pathology but are not the first area to show impairment, this instead is primarily seen within the domain of episodic memory, which is normally an area impacted later in the pathological process of AD (Mortamais et al., 2017). The lack of impairment seen in AD subjects in semantic tasks could be due to the lack of subject specific construct validity as inter-individual differences in semantic processing and acuity are highly variable. As such, a one size fits all test/measure may not accurately capture this impairment. Both on an individual level or extrapolating across a cohort study. By measuring verbal (semantic) acuity against fluid cognition may present an earlier way to index AD pathology. Something that is gravely needed both in daily practice, academia and clinical trial recruitment.

Verbal acuity has been shown to be incredibly stable across the lifetime prior to functional and cognitive decline. This allows for a strong comparison to be made from an objective measure, rather than a partner or self-reported change in performance. Individual’s level of insight into their own disease progression has been shown to be low (Logsdon et al., 1999; 2002; Vogel et al., 2004). Whereas impairment is either typically measured against a

comparative population, to determine a level of impairment from an aged matched healthy control, this allows for direct comparisons within an individual earlier than an impairment in a constituent cognitive domain.

Discrepancy Scores Within AD – This Study

The directionality of the discrepancy measure is critical to the interpretation of the outcomes.
Concurrent with prior research the fluid composite measure is subtracted from the crystallised ability score to give the subject specific discrepancy score (McDonough & Popp, 2020; Jacobsen et al., 2002). The essence of this measure is to understand how prior knowledge interacts with acquired knowledge. As such, it is key to the interpretation of the outcomes that the components of the fluid and composite measure have strong measurement validity. As described in McDonough & Popp (2020), a positive discrepancy score, whereby fluid abilities are lower than crystalised abilities, or in other words prior knowledge is not matched by current cognitive ability, represents a lack of adequate retrieval or an absence of fulfilment of maximum ability. The reverse of this, a negative discrepancy score, whereby crystallised abilities are greater than fluid ability, or current ability is below prior knowledge would indicate ability to learn new information and skills would be impaired. In individuals with higher crystallised ability they are expected to have higher fluid ability and vice versa. Therefore, a large much greater crystallised ability would suggest an abnormal decline selectively in fluid ability. Between these two points would be indicative of no discrepancy, with scores around zero are indicative of the absence of abnormal cognitive profile.

Moreover, within the progression of AD fluid abilities are commonly agreed to decline earlier than crystallized abilities (O'Carroll & Gilleard, 1986; Wechsler, 1944). This fundamental differential in timings of decline leads to large discrepancies between the two abilities, the rate of which is indicative of the rate of AD symptom progression (Albert et al., 2011; Bastin & Salmon, 2014; McKhann et al., 2011; Schmid et al., 2013). However, this is usually indexed within a singular domain of memory, primarily episodic, as this is the first domain normally shown to exhibit cognitive impairment within early AD.

Within this cohort a number of questions need to be addressed, primarily the constituent components that will comprise each functional composite and what the scores for this cohort will be compared to in the absence of any normative data. The cohort in question is absent of standardised crystallised measures such as the NART, Boston Naming Test, Category Fluency or Verbal Fluency, but does contain language measures. The NART is often used as a proxy for education, general literacy and overall intellectual ability prior to any disease state (Ryan & Paolo, 1992). However, the measure of

language, consisting of the component part of the MMSE, does not have comparable normative data.

The language domain of the MMSE has long been argued to be stable throughout the initial stages of cognitive impairment due to AD. Research to this extent from Choe and colleagues (2020) investigated the constituent component domains of the MMSE within the ADNI cohort. They found the language domain to be comparable in the MCI progressors and non-progressors with identical means and standard deviations in these cohorts and a non-significant groupwise difference. They also showed that MMSE subscores for orientation and construction, as well as for memory, are useful predictors of conversion from MCI to AD. This is in part due to the nature of the advancement of AD with orientation being one of the differentiators to the functional impairment associated with progression to AD. This finding was also replicated in a natural old age cohort of 500 elderly care home residents. Poor performance on the MMSE orientation domain is associated with faster rate of decline on total MMSE scores over time, although this may be due to the fact the orientation domain consists of a third of the points for the overall global score (Guerrero-Berroa et al., 2010). Language items have been shown to decline much later into the disease with a number of papers showing their decline being indicative of moderate to severe AD (Ashford et al., 1989; Small et al., 1997; Blair et al., 2007). This is in line with the notion of the crystallised composite discussed previously and can be argued to be a highly suitable proxy for a crystallised score within this cohort.

This differential change will be included in part of this analysis by utilising the staging from Jutten & colleagues (2020). This study will assign values to each individual to reflect their disease stage. This will be done in four stages (1;CDR-SB=0.5, 2;CDR-SB=1, 3; CDR-SB=1.5-4, 4:CDR-SB=4.5+). The prior work using this staging showed longitudinal differential declines in a range of measures that varied by stage over one year and three years. These measures spanned semantic memory, episodic memory, global cognition (MMSE), executive function and processing speed. Critically to this analysis, both immediate and delayed memory only showed impairment (-1SD) at stage 4. Executive function, processing speed, attention and working memory were again shown to be stable across the one-year time course for all of the first three stages with very little differentiation in terms of score for each stage also. At stage 4 these measures dropped to around -0.5 SD and as with the memory measures showed a slight decline across the one-year time course. These results show the similarities between these measures at each of these four stages based upon CDR scores.

The relationship between discrepancy scores and both memory and CDR staging is unknown, as is the ability to utilise ubiquitous AD screening measures to comprise a crystallised (language) ability score. It is hypothesised that the discrepancy scores for subjects will be significantly

related to the immediate and delayed memory measures. This will be further analysed by splitting the cohort by CDR stage (as per Jutten et al., 2020) with the hypothesis that at the earliest stages of AD there will be an absence of a relationship, which becomes apparent as the disease progresses.

Methods

Study 1 Details

This analysis will look to validity of being able to compute a discrepancy score from commonly used measures within AD clinical trials. This discrepancy score will comprise composite measures determined from the MMSE for verbal acuity and broader cognitive domain tests from the CBB. These two composites will be based upon indexing crystallised and fluid cognitive abilities of each subject, without being composed of episodic/semantic memory indices used for comparison purposes. The two measures will be compared to find a cognitive discrepancy score for each subject. The study will look to uncover the strength of the relationship between a discrepancy score and memory measures. And understand how this changes throughout the disease course of AD.

As this is the first analysis undertaken for this thesis it is also key to understand the differences in regional variation. Given this dataset comprises 29 different countries the data for each component measure will be assigned a region and analysed to see if any significant differences exist. If there are, this will be taken into account with further analysis.

Assessments & Composite Measures Composition

Crystallised Composite

Within this cohort, the battery of tests primarily consists of memory, functional and executive function measures. There is a dearth of stand-alone vocabulary measures which could be used to comprise crystallised abilities. However, some sub components of the MMSE do index this. The issues with the MMSE have been widely discussed in **Chapter 1**, however it is important here to reiterate the lack of comparative normative data and regional cultural sensitivity. These issues are particularly prevalent for language tasks as the cultural salience will not be always comparable with the direct translations utilised within the data collection. For this proposed comparison another factor to consider is the age matched normative data that will allow a direct comparison to a z-score. However, this is not something available or widely utilised for the MMSE. The measure was directly translated and as such is argued to relate to the normative data pertaining to the original version as no cultural adaptation was undertaken in light of the different languages of administration. Taking this measure for the crystallised composite was done for two reasons. Primarily this analysis was originally

composed to look exclusively at the endpoints utilised within clinical trials, as such, none of the crystallised measures from the McDonough & Popp (2020) paper are normally undertaken within a Phase III trial in early AD. Secondly, from the measures from the screening population these component measures from the MMSE were the only aspects that pertained to language indices. This includes the naming, repetition, comprehension, instruction, reading questions from the MMSE and is scored out of 8. However, as there is no standardised normative data, for the constituent components of the MMSE across all 29 countries, z-scored indices will be computed from the near 10,000 individuals that comprise the screening population. The computed z-score will be restricted to that of the crystallised composite.

These computed z-scores will be validated in part by comparing them to the years of education for each subject. A perfect correlation is not sought, but a weak but significant correlation is hypothesised due to the nature of the construct of the crystallised measure. If the null hypothesis is instead met, a factor analysis will be run to find constituent measures that would comprise a factor related to crystallised abilities. This factor will then be used to contrast to the fluid score to ascertain an individual discrepancy score for each subject.

Fluid Composite

An overall z-score from the Cogstate Brief Battery (CBB) will comprise the fluid composite. The CBB is a widely used and highly researched battery of tests and gives a good broad range of fluid abilities; processing speed, attention, visual learning and working memory. This is a set battery and is z-scored against a normative dataset within the Cogstate system, giving an automatic z-score as the primary outcome measure for each individual test. These 4 individual z-scores can then be averaged to form the final fluid composite measure for each subject.

Discrepancy Measures

As subjects are dichotomised by impairment in the verbal learning task (ISLT) and the overlapping crystallised and fluid nature of the stimuli this will be excluded from comprising the discrepancy measure. The discrepancy score for each subject will be calculated using the following equation: fluid composite – crystallised composite = discrepancy score.

A positive discrepancy score would indicate higher fluid ability and a negative discrepancy score would indicate an impairment in fluid ability compared to expectations based upon the subject’s crystallised ability. Important in the interpretation of discrepancy scores is setting a cut off apriori that seeks to correctly categorise groups who are showing an issue on this measure. Prior studies have set a cut off of 1 standard deviation (+/-1) as a threshold, whereas other subsequent analysis has used a 0.7 criterion for the threshold of impairment. In line with diagnostic criteria pertaining to

cognition (albeit with a focus on memory) the working group thresholds are set at 1 standard deviation or greater to signify impairment in a cognitive domain. This is also in line with wider psychometric philosophy utilised widely across multiple therapeutic areas.

Participants

This study looks to address the issues of prior research comprehensively by analysing a sample size of nearly 10,000 subjects (n=9759). The cohort comprises subjects who have presented to memory centres and consented for testing for the possibility of taking part in a clinical trial in early AD. Each subject has undergone a fully informed consent process consistent with each country/states IRB and EC guidelines. The screening criteria for inclusion into the study mandated the exclusion of medication affecting cognition, known comorbidities prior to brain imaging and depressive illness within the last five years. Full extensive details of this cohort can be found in the methods chapter.

Statistical Analysis

Initial analysis will involve preparing the dataset for analysis and answering some basic questions on the demographics of the cohort itself and how this relates to the cognitive endpoints. Box plots and descriptive statistics will be run on the fluid composite measures (full cogstate brief battery), as well as using correlation analysis to understand the relationships between all measures. Analysis will also be conducted to look at the impact regional differences had on this screening data, with each region being compared across the three key measures (CDR, ISLT, MMSE). Discrepancy scores will be computed using two individual composites (fluid & crystallised). With both composites and the discrepancy score itself will be in the form of z-scored values and computation for the individual composites is described above. The crystallised composite will be correlated with the years of education measure to understand the extent of the relationship between the two.

Subsequently, correlation and linear regression analysis will be performed to understand the relationship between discrepancy score and immediate and delayed memory recall. After these analysis have been run, the cohort will be split by CDR group and the correlation and linear regression analysis will be re-run for each CDR group.

The questions this analysis aims to address is the strength of the relationship between discrepancy scores and those on conventionally assessed verbal list learning measures that index immediate and delayed verbal recall. And secondly, how impairment in a discrepancy score predicts impairment in one of these verbal memory domains at cut off’s akin to MCI criteria.

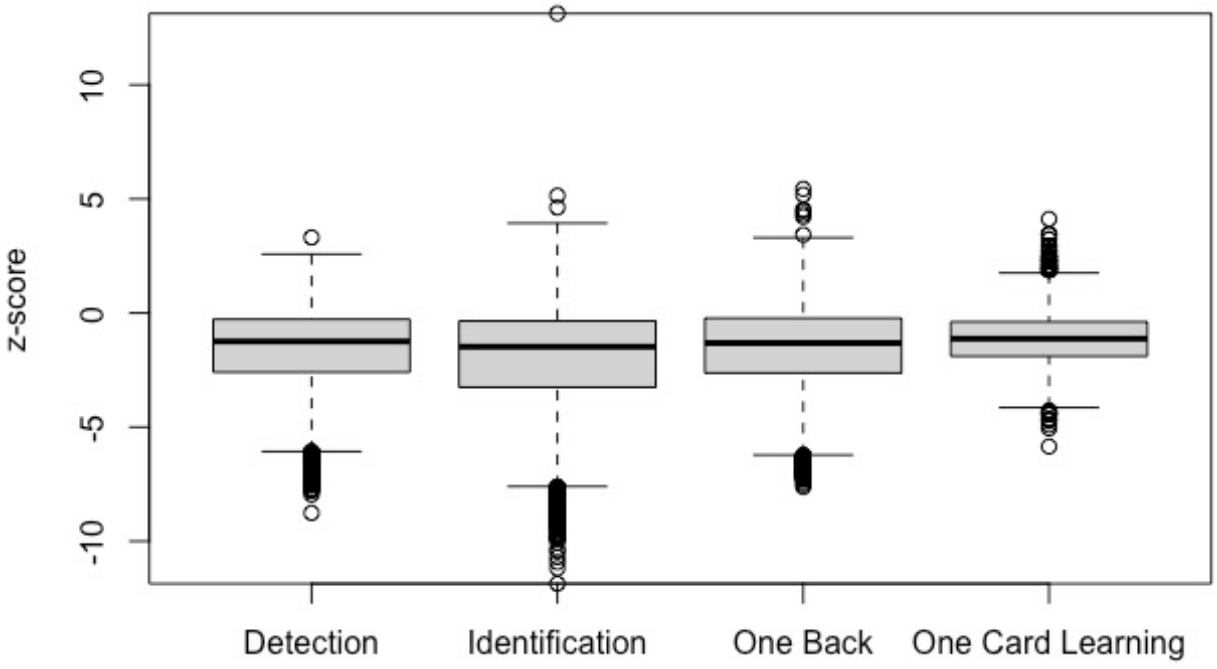
Results

Dataset handling

The dataset initially contained 9,759 screening visits that contained data for the MMSE, ISLT, CBB and CDR. Of these, 879 sessions were rescreened subjects, these were removed from the analysis set to not confound any results due to practice effects or longitudinal variation in these measures. Further to this, basic descriptive statistics were run on the main cognitive variables within the data set to analyse the normality of each variable, this was done on all 4 z-scores of the CBB and the overall MMSE score. ISLT values for both immediate and delayed remained untouched due to the proposed analysis utilising these as independent variables. As with the ISLT, the raw scores for all CBB measures are automatically z-scored by the Cogstate system against their normative database, giving each subject a z-score per measure.

Table 3.1 Correlation Matrix For All Measures (Significant values in bold)

	ISLT Delayed Recall	ISLT Immediate Recall	CBB Detection	CBB Identification	CBB One Back	CBB One Card Learning	CDR-SB	MMSE Total
ISLT Delayed Recall								
ISLT Immediate Recall	0.566							
CBB Detection	-0.016	0.092						
CBB Identification	-0.019	0.112	0.708					
CBB One Back	0.095	0.153	0.438	0.601				
CBB One Card Learning	0.214	0.189	0.014	-0.012	0.102			



CDR-SB	**-0.194**	**-0.185**	**-0.057**	**-0.111**	**-0.123**	**-0.093**		
MMSE Total	0.235	0.268	0.109	0.135	0.190	0.232	-0.258	

Figure 3.1. Box & Whisker Plot For Each Fluid Composite Measure

Table 3.1 shows the covariance of all key measures within the dataset. The CBB has some covariance across the initial two simple measures (identification and detection) and less covariance with the more executive tasks. Memory tasks have some covariance with one another but not with the other measures. Whereas the MMSE has some small level of covariance with all measures and only with the CDR-SB does it have a negative covariance. All four measures were normally distributed but had some larger values in the negative direction (**Figure 3.1**). **Table 3.2** shows the descriptive statistics for each of the fluid composite measures.

Cogstate Brief Battery	Minimum	1st Quartile	Median	Mean	3rd Quartile	Maximum	Bottom Whisker	Top Whisker	Standard Deviation
Reaction Time (DET)	-8.760	-2.595	-1.244	-1.541	-0.274	3.310	-6.077	3.208	1.716
Attention (IDE)	-11.860	-3.252	-1.479	-1.945	-0.349	13.125	-7.607	4.005	2.146
Working Memory (ONB)	-7.623	-2.628	-1.316	-1.509	-0.226	5.438	-6.231	3.377	1.777
Learning (OCL)	-5.851	-1.896	-1.127	-1.083	-0.385	4.120	-4.162	1.880	1.097

Due to the target population for this analysis & original cohort being early AD subjects with an MMSE of 17 or below were also taken away from the dataset to enable better homogenisation of the cohort. Furthermore, subjects with an incomplete set of measures needed for this analysis were also excluded. A total of 1798 datapoints were removed, giving a total of 7082 subjects.

Table 3.2. Descriptive Values for The Cogstate Brief Battery (CBB) – Figure 3.1

Whilst there can be a strong argument to be made for a large number of CBB (fluid) scores significantly deemed to be erroneous due to the level below the norm they are performing at. However, this is contrary to Cogstate administration guidelines (Fredrickson et al., 2010), all of these administrations met the integrity criteria built into the system and were

completed in line with these guidelines and were therefore included in their entirety.

Computing Crystallised & Fluid Composite Scores

The fluid composite was comprised of the entire Cogstate brief battery (CBB) which encompasses similar cognitive domains to that of the fluid composite from McDonough & Popp (2020). Within the CBB are tests of reaction time (Detection), Attention (Identification), Memory (One-Back Memory) and Executive Function (One Card Learning). These 4 measures are automatically z-scored by the Cogstate system and were averaged to give an overall fluid composite measure for each subject.

The crystallised composite was developed by utilising constituent domains of the MMSE. The object naming, repetition, commands and writing aspects of the MMSE were selected to form the crystallised composite, these are also the domains detailed under the language section by the test developers (Folstein et al., 1975). These gave subjects a score out of 8 and each measure was then z-scored against the entire screening population of 7082 who had a complete fluid and crystallised score, who scored above 17 on the MMSE and were not rescreens.

MMSE Concordance & Regional Disparities

In the absence of a stand-alone language paradigm and in keeping with the wide utilisation of the MMSE, the language components of the MMSE were utilised to this end. However, prior to the computation of this composite measure further detailed analysis of the MMSE needed to be undertaken. In order to try to account for some of the variability in such a vast dataset, a regional analysis was undertaken to ascertain the viability of utilising the MMSE within this analysis and for any future analyses that involve this dataset.

As such comparisons were made to ascertain any regional differences in this measure. No overall differences were found within the total MMSE score [$F(1,8930)=1.050$, $p=0.300$] for all subjects and no individual group differences were seen either based on the regional split by 4 regions using Tukey's HSD post hoc test; (1)North America, (2)Europe & South Africa, (3)Asia Pacific & (4)South America (1-2 $p=0.986$, 1-3 $p=0.999$, 1-4 $p=0.106$, 2-3 $p=0.977$, 2-4 $p=0.086$, 3-4 $p=0.141$).

Furthermore, after computing the crystallised composite and z-scoring the measure, this analysis was rerun. Findings showed a significant model effect [$F(1,8939)=26.900$, $p<0.001$] across regions and a significant group differences between APAC and all other regions (1-3 $p=0.00$, 2-3 $p=0.00$, 3-4 $p=0.00$). Subsequent analyses were also run on the key endpoints for each

measure, CDR-Sum of Boxes (CDR-SB), ISLT total (immediate) recall (ISLTTR), ISLT delayed recall (ISLTDR) as well as the fluid composite measure. ANOVAs for the CDR-SB [$F(1,6413)=47.6$, $p<0.001$], ISLTTR [$F(1,8402)=140$, $p<0.001$) and ISLTDR [$F(1,8402)=265$, $p<0.001$] were all significant, however the model for the fluid composite by region was not [$F(1,8294)=2.01$, $p=0.16$]. p-values for each model's post hoc analysis using the Tukey HSD test is displayed below in **Table 3.3**.

Table 3.3. p-values for Tukey HSD post-hoc tests

Region	CDR-SB	ISLT Immediate Recall	ISLT Delayed Recall	Fluid Composite
1-2	0.591	<0.001	<0.001	0.748
1-3	<0.001	<0.001	<0.001	0.120
1-4	0.048	<0.001	<0.001	<0.001
2-3	<0.001	0.488	<0.001	0.628
2-4	0.013	<0.001	0.004	<0.001
3-4	0.757	0.009	0.848	<0.001

Calculating Discrepancy Scores

Removing any subjects without a complete discrepancy score the final number of cases for analysis were 7082. First, the crystallised composite was correlated with years of education within the cohort, there was a weak statistically significant positive correlation ($r(6918)=0.110$, $p=<0.001$) [Figure 3.2]. To note, subjects with scores over 40 for years of education were removed from this analysis, leaving 6918 subjects for this comparison alone.

Discrepancy scores for each subject were calculated by subtracting each crystallised composite score from the fluid composite score, consistent with McDonough & Popp (2020). The crystallised and fluid composites were positively (yet weakly) correlated with one another, ($r(7080)=0.132$, $p=<0.001$).

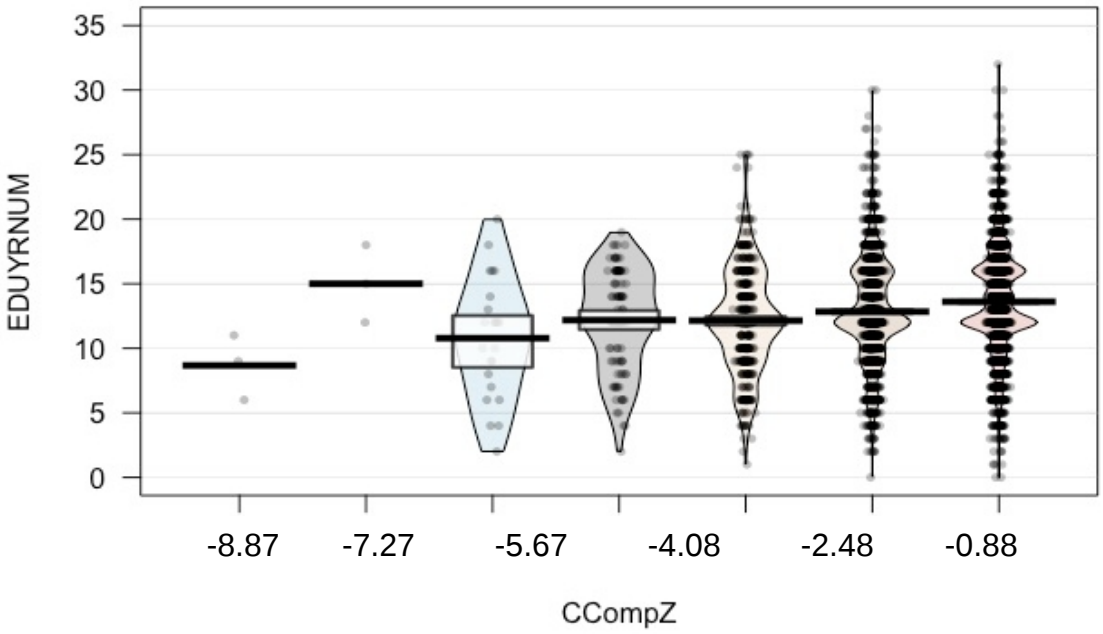


Figure 3.2. Correlation of Years Of Education & Crystallised Composite Z-Score.

Relationship Between Discrepancy Score & Memory Assessment

Part of the underlying hypothesis for this measure is its relationship with impairment on standard memory measures used within early AD clinical trials. To this end correlations were run across the two key endpoints from

the ISLT to ascertain the strength and statistical relationship between these endpoints and the discrepancy measure (DS).

Both relationships were statically significant and positively associated, with immediate recall (**Figure 3.3**) r(7078)=0.193, p<0.001and delayed recall (**Figure 3.4**) r(7078)=0.170, p<0.001, having very similar weak correlations with a subjects discrepancy measure.

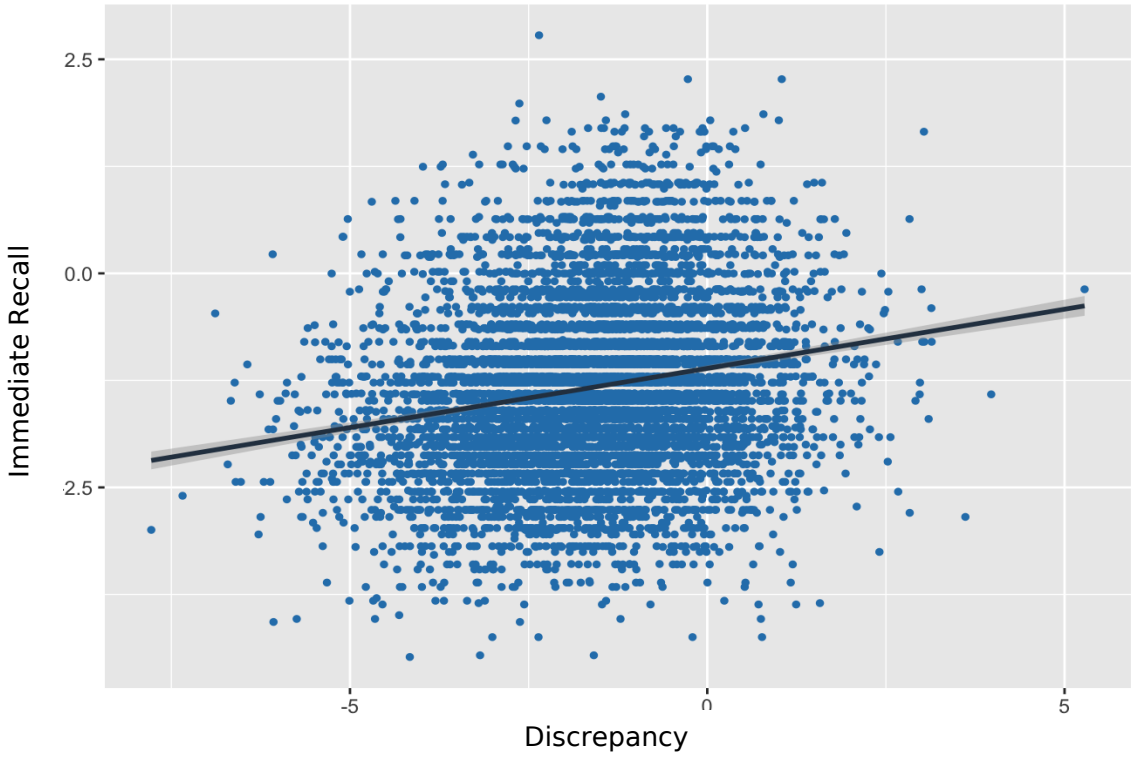


Figure 3.3. Scatter plot with linear regression line and confidence intervals of discrepancy score (DS) and immediate memory recall z-score (ISLTTRZ) for each subject

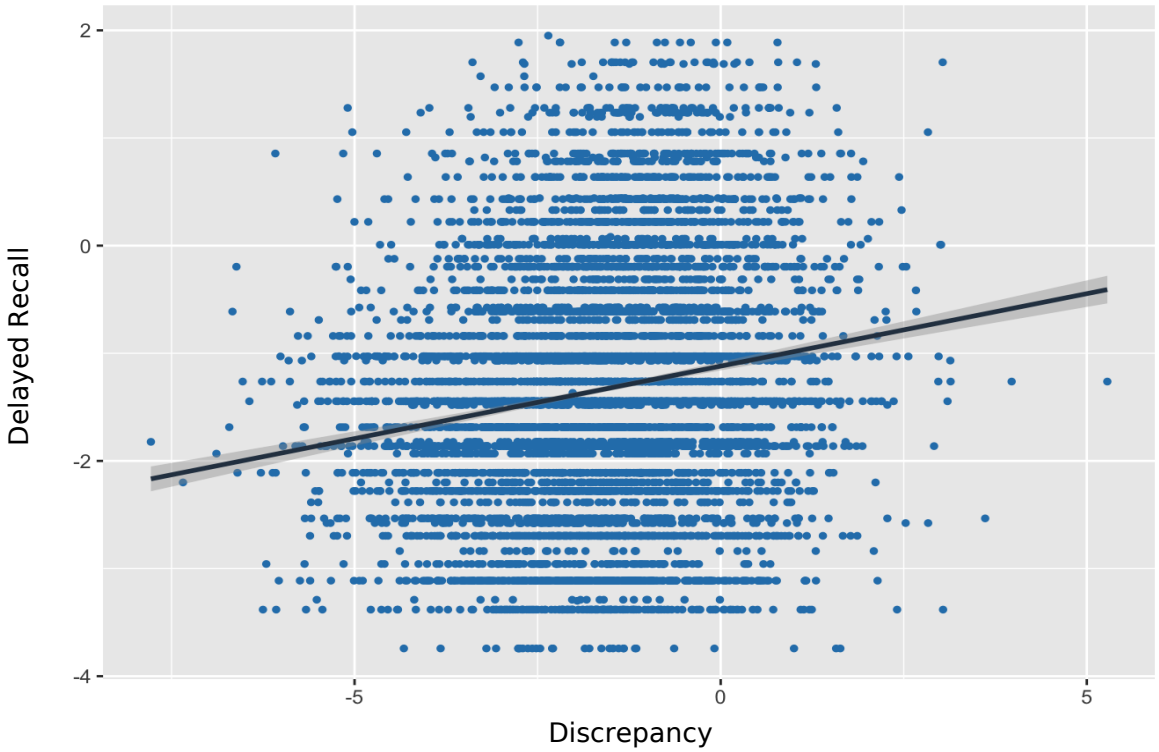


Figure 3.4. Scatter plot with linear regression line and confidence intervals of discrepancy score (DS) and delayed memory recall z-score (ISLTDRZ) for each subject

Assessing This Relationship By CDR Stage

As the population was vast and a broad reflection of subjects presenting to research centres, memory clinics, hospitals and private healthcare sites, the analysis utilised the CDR-SB score for each of the subjects to categorise them into AD stage in line with the research by Jutten and colleagues (2020). In the absence of amyloid status for the full cohort, this measure serves as a proxy for disease progression. As such and in line with NIA-AA (2018) & FDA (2018) criteria the cohort was grouped by CDR-SB score. Descriptives for each group are detailed below in **Table 3.4**. As not all subjects completed the CDR, the dataset was restricted to 5578 subjects.

Table 3.4. N's, Means & (Standard Deviations) for each group on each endpoint

CDR Group (CDRSB)	N	Discrepancy Score	Immediate Recall	Delayed Recall
1 (0-0.5) [Preclincial AD]	245	-1.21 (1.25)	-1.14 (0.89)	-1.17 (1.02)
2 (1) [Verly Early AD]	605	-1.54 (1.50)	-1.41 (0.86)	-1.42 (0.93)
3 (1.5-4) [MCI]	4395	-1.77 (1.45)	-1.66 (0.84)	-1.73 (0.89)

4 (4.5+) [Mild AD]	333	-2.15 (1.58)	-1.86 (0.89)	-1.93 (0.92)

An ANOVA with Tukey HSD post hoc test was also run to ascertain the differences between the discrepancy scores across groups. This was significant [F(1,5383)=105.1, p<0.001], with all individual groups being statistically different from one another too (1-3 p=0.005, 1-4 p<0.001, 2-3 p<0.001, 2-4 p<0.001, 3-4 p=0.01), apart from group 1 and 2 (p=0.11).

As shown in the descriptive statistics in **Table 3.2**, the groupings were unequal with the majority of subjects being classed into group 3 (CDR-SB, 1.5-4). Mean scores also increased with grouping in line with the hypothesised progression that determined the grouping with the CDR-SB score. Pearson's correlation analysis was conducted on relationships between the ISLT and discrepancy scores. Only group 3 (CDR-SB 1.5-4) showed significant associations between discrepancy score and both aspects of the memory measure (immediate & delayed conditions). Further to this, groups 2 (CDR-SB=1) and 4 (CDR-SB=4.5+) had significantly related relationships between only the immediate memory recall and discrepancy score. The r values for all of these relationships were poor with the greatest strength between the immediate memory and discrepancy score was in the most impaired group (4; r=0.180). All r-values (& p-values) are displayed in the **Table 3.5** and **Figure 3.5** below.

Table 3.5. Pearson correlation coefficients for each variable in relation to discrepancy score by CDR Group (r value [p=value] significant results in bold)

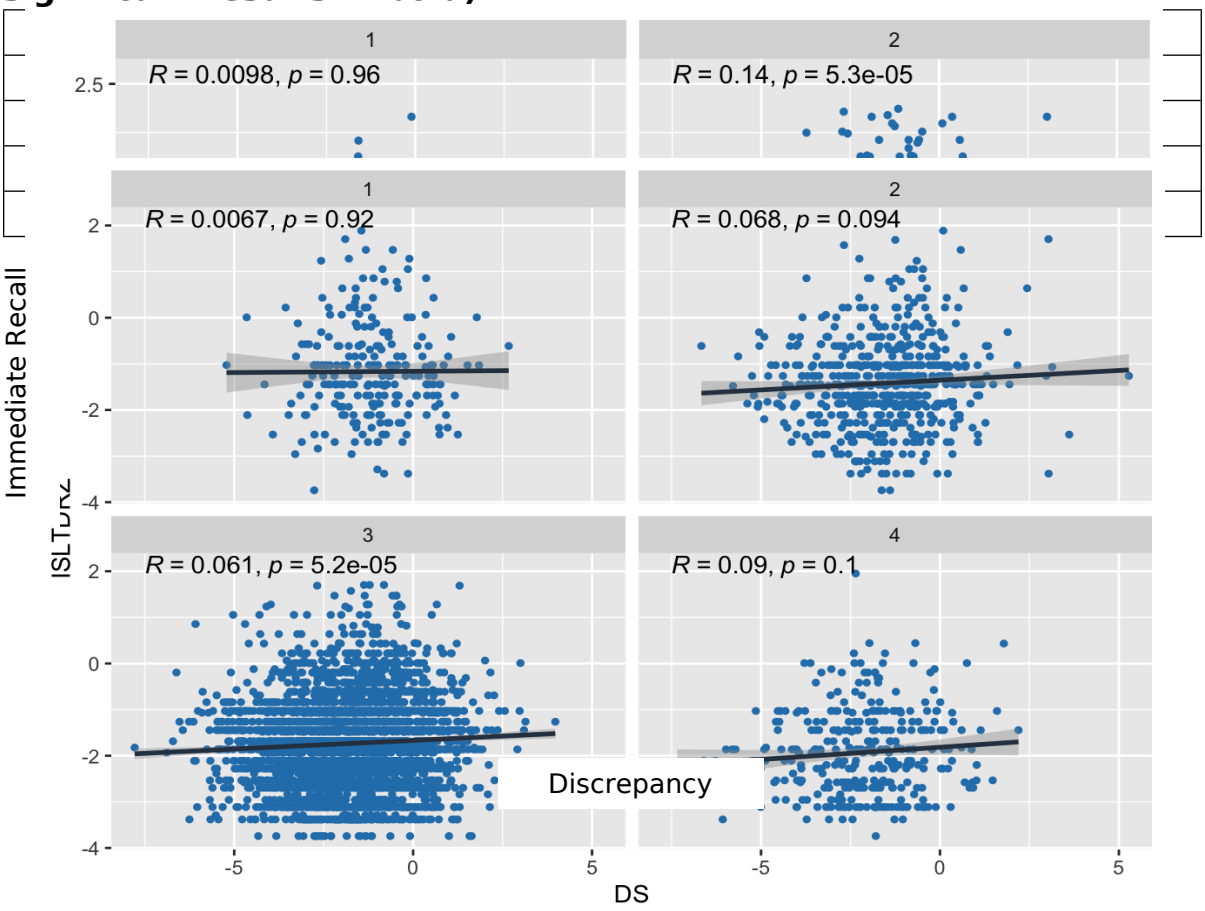


Figure 3.5(a,b,c,d) & 3.6(a,b,c,d). Scatter plot with linear regression line and confidence intervals of discrepancy score (DS) against immediate memory recall z-score (ISLTTRZ) & delayed memory recall z-score (ISLTDRZ) for each subject by CDR Group (Group 1= Preclinical AD, 2=Very Early AD, 3= MCI, 4= Mild AD)

Discussion

The findings from this analysis showed a weak but statistically significant relationship between a subject's discrepancy score and their level of impairment on immediate memory recall in all impaired conditions (2-4). However, this was not replicated in the delayed more cognitively taxing domain of delayed memory recall. The only significant relationship between delayed memory performance and discrepancy score was in the MCI group (3, CDR-SB=1.5-4), however the strength of the association was very close to zero and therefore demonstrates very little relationship between these two measures.

The strength of the relationship between discrepancy score and immediate memory was weak, although fairly uniform across groups 2, 3 & 4 (r=0.16, 0.1, 0.18 respectively). The lack of relationship with the first group shows the relative lack of impairment seen within this subgroup from a functional and proxy cognitive perspective as measured by the CDR. The prior research indicated that a discrepancy score may show up early for those who have amyloid pathology present in the progression of AD. However, as no amyloid markers were utilised in this analysis, and although a number of these subjects were subsequently diagnosed with AD based on biomarkers criteria, this population were presenting to memory clinics and may have unknown multiple pathologies and/or co-morbidities. This analysis shows that within a screening population for a clinical trial discrepancy measures, built utilising common measures for AD, index something different to that of traditional memory measures. The lack of a relationship with the primary index of memory, delayed recall, suggests that across the stages of AD, a discrepancy measure is not related to that of the primary initial domain of impairment within an amnestic presentation of AD. This can be argued to be due to the composition of the fluid composite score comprising half of the discrepancy measure and this includes memory indices but not a delayed component of recall. This therefore suggest that a discrepancy score and memory measures used for diagnosis do not have convergent validity and should be explored further to assess their relationship within screening and diagnosis for AD. Further to this, it is important to note that the relationship mirrors the expected impairments across AD whereby a discrepancy score becomes significantly related to that of memory the more progressed an individual's AD is deemed to be (as measured by the CDR). This demonstrates the criterion validity for a discrepancy measure such as this within AD.

Consistent with prior work (e.g., Cattell, 1965; Kaufman et al., 1996), the crystallised and fluid composites were positively (yet weakly) correlated with one another. This may go some way to explaining the dearth of relationships between these measures. The lack of stand-alone language measures and the exploration as to the suitability of common AD screening measures for this component of the discrepancy score limited the range of performance with the majority of the cohort performing at ceiling on the MMSE language domains.

There are also several other limitations of this analysis that are important to consider in any generalisation. There was a weak relationship between crystallised composite and years of education which demonstrates the lack of suitability for utilising the language domains of the MMSE as a proxy for crystallised ability. The significant relationship demonstrates the validity of exploring this as an option for overall language ability but the strength of the relationship and large number of subjects at ceiling restricts the applicability for future research.

When comparing this analysis to prior work and findings, it is important to differentiate the make-up of the composite measures. Whilst the crystallised composite is theoretically similar to that of the prior work of McDonough & Popp (2020), the fluid composite in this analysis is divergent. As the objectives of this analysis were different to that of this prior research, the crystallised composite computed here within, related to uncovering the serviceability of common AD screening measures and their relationship to common severity indicators. The fluid composite by nature of the analysis had to be composed of other cognitive domains in order to compare this score to their relationship to common severity indicators (delayed memory recall). This engendered differences that may explain the lack of early-stage relationships between discrepancy scores and both immediate & delayed recall. The fluid composite was composed of indices of processing speed, working memory, short-term memory & attention (Fredrickson et al., 2010), which is in contrast to that of the prior composite of short-term and long-term memory, as well as processing speed & attention (McDonough & Popp, 2020). Within the McDonough & Popp paper, the differences between fluid and crystallised ability were manifested through memory and executive function, whereas this analysis showed these differences between fluid ability to be weaker when this paradigm is altered. This underlines the inability of the MMSE language domain to comprehensively index crystallised ability within a screening cohort, but still shows the potential for this measure to uncover an alternate index to that of typical cognitive domain measures. Overall, this absence of fully concordant results with prior research shows that a) discrepancy scores index different domains to that of delayed memory recall and b) within a screening population this is not a useful indicator of impairment when utilising the MMSE language domains for crystallised ability.

Due to the high rate of ceiling effects on the crystallised composite, the fluid measures (CBB) comprised the majority of the variations in discrepancy score. It is important to note here whilst the CBB was completed under supervision and administration guidelines, as well as having effort tests and thresholds for incompletion built into the measure, the CBB was done as a distractor task in between the immediate and delayed conditions of the verbal memory measure and as such carries with it the confound of such an distractor task, in that, subjects may not engage with it in the same way if they have been given advice to the contrary of the administration guidelines, due to the CBB not being part of the inclusion criteria for the trial from which the dataset was taken from.

Conversely, the regional analysis is something that uncovered broad variation across these three common screening measures for clinical trials within AD (MMSE, CDR, ISLT). The differential pattern of divergence showed that the screening population for this trial was significantly different in the Asia-Pacific region, with subjects being objectively impaired on the ISLT to a greater degree but had lower CDR values compared to other regions. This is suggestive of caregiver bias commonly seen in this region based upon cultural practices when it comes to elderly care and not wanting to seem as though an individual's family does not take care of their relatives well by scoring highly on the CDR interview. The opposite was true for North America whereby subjects on the whole showed much less impairment than that of all other regions on both indices of the memory measure (ISLT). However, this calls into question the cross-cultural validity of these measures, they can be used in these regions and completed, however they produce divergent results. Looking forward to future analyses, the region should be included as a covariable for any future analysis on this dataset.

The overall strength of these relationships between the measures analysed here precludes any strong conclusions however, there are some key takeaway points and future considerations for any further analysis utilising this cohort. The small relationship shows the absence of similarities but indicates a different index of cognitive performance, further research is needed with variables linking proteinopathy and biomarkers, to these clinical presentations as shown in prior work research into this area (McDonough & Popp, 2020). Nevertheless, this does not remove the ability of this measure from having the potential for finding earlier impairment, however this disease specific biomarker information is crucial to this endeavour. The criterion validity of this as a measure within AD can be supported by this study as it tracks as expected with memory indicating progression akin to that seen in AD. Furthermore, this study also calls into question the cross-cultural validity of these measures as the clear significant regional differences shows the differences in screening populations across the globe.

Chapter IV – Comparisons of Verbal Memory Measures

Chapter Outline

Previously we have shown that you can derive discrepancy scores from global cognitive measures. However, its validity as a measure of verbal acuity is limited. This discrepancy score measures something different to that of amnestic memory at screening, with that relationship changing across the disease stage of AD, demonstrating its criterion validity. To further understand the variances within specific measures of amnestic memory, this chapter will examine this aspect of psychometric validity within two different cohorts. One with a confirmed AD diagnosis, both pathologically and clinically. The other smaller cohort without unconfirmed diagnosis and unknown comorbidities. These studies aim to understand cognitive variability of memory measures within key AD-related populations by exploring their concurrent, convergent, construct and content validity.

Introduction

Commonly, many measures of memory are frequently interchanged with one another, on the basis they produce concordant results. This chapter will explore the validity of this across four memory measures commonly used to find impairment within AD. This is key to understanding sources of heterogeneity in cognitive performance of individuals across the spectrum of AD.

Variations in cognitive performance are something that researchers try their best to minimise. Within AD there are many reasons for variations within a single “homogenous” cohort. Within this thesis I explore how much of this is down to clinical trial design. How much is down to the fundamental diagnosis and pathology of individuals. And finally, how much is down to the psychometric validity (or lack thereof), of these endpoints. This chapter will explore the latter two aspects of this, looking at between-measure comparisons.

Finding out if an individual is exhibiting cognitive impairment can be done in three ways. Firstly, and most commonly within screening paradigms, a comparison is made to that of a normative data set. A second way is comparing within individual performance across cognitive domains. This is done by measuring expected performance using a stable measure such as IQ, NART or vocabulary ability, against a single or multiple domains of interest, to give a composite score or discrepancy score. Thirdly, the gold standard, is to measure a subject longitudinally on a single domain to minimise other factors that may induce variability. This is especially

important within neurodegenerative disease where individuals decline from prior levels of ability and function. However, longitudinal measurement is rarely available and often a costly endeavour.

Across all three paradigms, tests that are thought to measure the same domain are often used interchangeably. For example, there are many different versions of list learning measures, all of which are often seen as measuring the same things, both within research and clinical trials. However, there is limited knowledge about the comparisons between these measures themselves and some have very divergent constructs. It can be argued that this is an oversight and is something that induces cognitive variance in results and contextual interpretations of study results.

Across the AD continuum cognitive deficits present in differential domains at different stages of the disease. This can also vary greatly from person to person based upon a wide variety of factors, but heterogeneity within AD is argued to comprise three distinct areas; risk factors, protective factors and concomitant AD pathology (Ferreira et al., 2020). From a cognitive perspective, the trajectory of impairments is fairly well characterised. Preclinical AD is characterised by normal cognition, but has shown signs of absence of practice effects in some cohorts (Hassenstab et al., 2015). This is also something to consider when looking at any variances between measures within a single cohort. The inception point of symptomatology for AD in >80% cases is amnestic impairment (Albert et al., 2011). Recent research has also shown a relationship between genetic risk factors and visuospatial, attention-based measures of memory (Lu et al., 2021). All of which is important to consider when indexing impairment. The final stage of cognitive interaction is the relationship between AD biomarkers and cognitive impairment longitudinally. This primary interaction is related with tau and to a lesser extent amyloid pathology (Hanseeuw et al., 2019; Mortamais et al., 2017).

All of these disease specific differences further elucidate the intricacies of cognitive domains across the course of AD. This is compounded by the lack of congruency between the common measures employed to index the same hypothesised cognitive function. A greater psychometric understanding is required to better understand the relationship between these commonly employed measures of amnestic memory in early AD. As currently the assumptions are these different measures index the same impairments and result in homogenous populations, when often the reality is hypothesised to be highly cognitively heterogeneous cohorts.

Practice effects related to AD

In order to properly examine any differences due to measure variability, it is important to first understand the interplay between AD disease pathology

and cognitive performance. Recent evidence, enabled by more frequent cognitive testing paradigms, has shown some indications that subjects early in the disease course show a lack of improvement, alternatively termed practice effects, when repeated testing is undertaken (Jutten et al., 2020). Repeated testing within subjects, who are amyloid positive but cognitively unimpaired, show an absence of improvement on common neuropsychological testing paradigms (Hassenstab et al., 2015). One explanation for these deficits is that this may be the result of disruption to Medial Temporal Lobe in preclinical AD that manifest as deficits in learning rather than as a progressive decline in memory recall, thus manifesting in subtle impairment rather than scores well below normal performance. This is something that is only picked up by more frequent testing than the typical 3-6 months period employed in research and clinical trial settings. Lim and colleagues (2020) used a test and paradigm of repeated measurements in learning ability over 6 days, whereby subjects had to learn new language characters, specifically Chinese symbols. The results showed cognitively normal participants who were amyloid positive had significantly worse learning trajectories than cognitively normal participants who were not amyloid positive.

Looking at the wider literature within this area, a recent review has shown the consistency of impaired practice effects being useful markers of early cognitive decline (Jutten et al., 2020). As shown within this literature review there has been a consistent, whilst small in size, diminished practice effects associated with either current diagnosis and/or indicative of future cognitive decline. Primarily the focus of these re-testing papers focused on testing across 1 week. When restricting these studies to those looking specifically at early AD, this absence of practice effects have been associated with known AD risk factors, such as APOE (Oltra-Cucarella et al., 2018) and overall amyloid deposition (Duff et al., 2014; 2017; Ihara et al., 2018; Galvin et al., 2005), although these studies have been limited by small numbers of participants and findings have not always been consistent in relation to APOE status (Duff et al., 2014; Hassenstab et al., 2015; Wilson et al., 2018) and amyloid deposition (Wilson et al., 2018). Recent progressions in the staging of AD, have not been investigated in relation to practice effects. Further research is warranted to inform how other biomarkers such as tau and neurodegeneration fits into the overall interplay of the absence of practice effects being a hypothesised early indicator of AD. To note these studies have primarily used verbal list learning measures as indicators of memory performance in both control participants and as well as though with a diagnosis of MCI or AD.

Interestingly some studies have found that whilst the absence of practice effects is widely associated with amnestic recall and learning ability, it is also prevalent across other cognitive domains (Duff et al., 2010). However, again results in these domains are often limited by sample sizes and incongruency

of domains. This may be down to a poor understanding of the constituent cognitive measures employed in these studies and lack of broader biomarker testing, resulting in heterogenous cognitive and pathological status' within cohorts.

APOE allele influence on cognition in AD

On top of ATN biomarkers discussed in detail previously, many environmental and genetic factors affect risk of developing sporadic AD. The strongest genetic risk factor being APOE ε4 (Roses, 1996). This gene being primarily responsible for lipid metabolism with carriers of the ε4 allele being at increased risk of AD. With evidence showing a quicker disease progression and earlier onset (Haan et al., 1999; van der Flier et al., 2011). However recent research has suggested a potential benefit of this allele, in that carriers of the same abrogating gene have a superior ability in a task of visual working memory (Lu et al., 2021). They presented strong evidence that within amyloid carriers who were cognitively normal this improvement remained (although to a smaller degree). Thus, indicating a subtle interplay of divergent cognitive functions early in the disease. It is clear however, this is only specific to short term memory, as carriers of this allele are impaired in long term memory in healthy elderly (Zokaei et al., 2019). This juxtaposition indicates nuance to the development of impairments in cognitive functions as there are bad and good parts to APOEε4 carrier status. This improvement in specific tasks indicates a possibility for APOE to impart resilience to certain systems within the cortex responsible for encoding short term visual memories. Understanding these mechanisms within AD and interaction with risk factors could preclude insight into the wider suggested phenotypic variability within AD as a whole. This emphasises the importance of understanding the cognitive measures used at a given stage of AD. The heterogeneity of cognitive measures is something that should be better understood and is an unabridged source of variance, which can and should be, mitigated against. Ultimately being vehemently avoided through thorough research and understanding of each constituent measure within the studied population.

Trial inclusion/diagnosis of MCI/AD

Conversely, it is currently agreed that the first and easiest deficit exhibited by individuals with probable AD is one of amnestic, usually episodic, memory impairment (Karr et al., 2018; Mortamais et al., 2017). In line with research (Jack et al., 2018) and clinical criteria, both for MCI due to AD (Albert et al., 2011) and early AD (McKhann et al., 2011), impairment below the mean of 1 to 1.5 standard deviations is at the very essence of the disease. However, many different assessments have been utilised to measure such impairment. Measures of episodic memory share a predominant construct validity but, their fundamental paradigms are unique to each.

However, there is no standardisation in scale selection across trials, resulting in a siloed approach by company and compound. With each compound under development reliant upon a different assessment to the detriment of consistency across trials and compounds, further burdening the field as well as clinical trial sites. In the case where an individual site is running multiple trials in one area of AD, sites potentially have to decide which trial to propose to a subject based upon the concordance between the inclusion scale and the subjects' current level of impairment. This quandary also gives rise to higher screen fail rates, leading to longer development timelines, greater patient burden and increased cost.

The gold standard way to uncover MCI on memory measures is by comparing an individual's performance to the normal performance for their age, education and gender. It is also highly preferable to have multiple assessments of individuals cognitive performance to accurately index any cognitive decline from a baseline level. Tracking over time can also give greater levels of reassurance to an individual by confirming that the impairment they are exhibiting isn't transient and is likely not to have an alternative aetiology. Larger and highly successful clinical trial sites commonly employ this continuous measurement of subjects in their database. This leads to quick and efficient screening of their trial database and getting potential treatments to individuals quicker. Which episodic memory measure a site uses widely varies and is driven by site's experience, level of qualification of staff, preference of paradigm and cost/availability.

Verbal List Learning Measures

There are a wide variety of measures that sites can employ. Verbal list learning measures such as the California Verbal Learning Test (CVLT; Delis et al., 1987), the Rey Auditory Verbal Learning Test (RAVLT; Rey, 1941; Taylor, 1959), Hopkins Verbal Learning Test (HVLT; Brandt, 1991), the International Shopping List Test (ISLT; Thompson et al., 2011), Verbal Recognition Memory (VRM; Robbins et al., 1994) and other word list measures share a concurrent paradigm. But measures such as the Wechsler Memory Scale Logical Memory measure (WMS-LM/LM; Wechsler, 2009) and the Free & Cued Selective Reminding Test (FCSRT; Grober et al., 1988) have very different constructs and can be argued to index slightly different cognitive processes. These measures are often used very interchangeably but have differences in underlying constructs.

The stability and validity of these measures is also fundamental to comparisons of performance in AD. The ISLT has been shown to highly stable across an 18-month repeated testing paradigm (measurements every 3 months) (Lim et al., 2021). In amyloid negative cognitively normal subjects scores on both the delayed and immediate recall on average remained

within a 1-point range across the time period. However, in amyloid positive MCI subjects, as well as amyloid positive AD subjects, an almost linear decline occurred in both groups across both conditions, with those with MCI showing a slightly more precipitous decline. This study also showed minimal declines over a 3-month period in line with other measures of memory within AD populations, showing gradual decline over many years.

The concordance between these memory measures has been an ongoing point of contention without resolution. This has been studied in prior decades with different diagnosis paradigms for AD. Studies such as Rabin & colleagues (2009) has looked at this, showing CVLT to best distinguish MCI subjects from healthy controls more accurately than the LM. However, they did not have amyloid confirmation within their subject groups. This lack of confirmation has been shown to lead to a 20-30% amyloid negativity (aka non-AD subjects) within cohorts such as these (Egan et al., 2019). Further to this, the ADNI memory composite (Crane et al., 2012) utilised a number of episodic memory components. ADNI's sample was selected to best resemble a clinical trial population (Petersen et al., 2010), thus making this a highly suitable sample to look for memory congruencies within. Within the ADNI composite, indices of the LM, ADAS-Cog and RAVLT are combined. The ADNI memory composite was slightly better at detecting change than total RAVLT recall scores, which performed best within the sample out of all the memory indices. Further to this, a confirmatory factor analysis (Park et al., 2012) of the full ADNI neuropsychological measures, showed an excellent absolute fit of the memory composite. However, model modification indices revealed the largest source of misfit arose from constraining memory factor loadings, particularly for RAVLT short and long delay recall, to be equal across groups. This study didn't include the LM measure within the factor analysis and the levels of unexplained variability in memory items were greater in the less functionally impaired group than in the more functionally impaired group. This indicates suitability of some measures in certain subgroups more so than in other subgroups. Overall, further study on a direct comparison with ATN biomarkers and expanded episodic memory measures, would be highly beneficial for the field.

Logical Memory & Alternate Forms

Story recall is often used interchangeably with verbal list learning, but has a fundamentally different construct. Story recall is argued to be clustered into chunks of information therefore, when compared to other memory measures, giving the recall of the story a slightly elementary slant. This is thought to allow subjects with greater executive deficits, whose ability to "cluster" is impaired, improved recall on this paradigm. In contrast to this are verbal list learning measures, where the often-random nature of the word lists, requires subjects to utilise their ability to cluster independently of the test paradigm. As executive deficits manifest slightly later in AD compared to amnestic

ones, this could lead to cognitive variability confounding comparisons across studies using these two measurement constructs.

The WMS-LM is used widely within clinical trials due to its initial utilisation within one of the first academic longitudinal cohorts, ADNI (Mueller et al., 2005). This is the primary measure used when indexing story recall and has been well validated as part of the wider WMS battery (Schnabel, 2012; Sullivan. 2005; Taler et al., 2020). However, the fundamental story has not changed since its inception decades ago, leading to familiarity with repeated testing as there are no official alternative versions. For clinical trials this problem is often compound by the use of direct translations. In order to have culturally valid stories this would require complex and timely validity studies, something which is often not in line with development programs. These direct translations can often carry biases and not accurately reflect regional discrepancies in phrases or cultural nuances. Direct translations can often mean something stands out more to a native speaker and as such becomes more memorable resulting in a better score than would have normally occurred in culturally sensitive cognitive testing. This results in subjects who are actually cognitively more impaired than they are testing at due to poor culturally insensitive testing paradigms leading to within country cognitive variations (Kave et al., 2021). Which when translated longitudinally, often means divergent biomarker profiles and disease trajectories. However, the alternative is to have unvalidated measures that are divergent to that of the original measure. This would mean the literature validation from its development would be lost. Nevertheless, cultural nuance is often missed and as such, cognitively heterogeneous populations are often recruited within screening for clinical trials.

Alternate forms of these memory scales are critical to enable wider understanding of both the measures and how they perform within disease. These measures allow for repeated testing longitudinally whilst still allowing for comparisons against normative data, the best of both worlds. Prior work has shown the validity of an alternate form of the WMS-LM, often termed the Morris paragraphs (Morris et al., 2014; 1997). The alternate Morris paragraphs were developed in concordance with the phonetic and psychometric principles applied to the original WMS story recall. Originally these were developed for the third version of the WMS (Morris et al., 1997), then updated to reflect the updated version four of the WMS based upon the same founding principles (Morris et al., 2014). Each composition is lexically and linguistically comparable to the original WMS, based on specifications used to originally develop the re-test stories equivalent to those in WMS-III/IV. Importantly direct translations are not used but translations are slightly adapted to fit not only the language in question but also the country targeted for use.

Looking outside the Morris versions, several other alternate story measures have been developed for use in place of traditional WMS-LM paragraphs for the purposes of repeat testing (Trifilio et al., 2021; Taler et al., 2021; Bolognani et al., 2015; Schnabel, 2012; Sullivan, 2005; Newcomer et al., 1994). However, not all have been matched to the scoring or administration guidelines of the WMS-LM subtest. This can lead to divergent concordance between these measures. Further to this, and key to utilising this measure within an elderly cohort, age matched normative data and age specific test variants are fundamental to wider usability of the scale. This is something which is not possible without the linguistic validity of alternate versions of any scale. As such the Morris paragraphs represent the best alternative to the WMS from a psychometric standpoint based upon the prior validity studies undertaken by their authors (Morris et al., 2014; 1997). This allows the scoring of the Morris paragraphs to mirror the cut offs for the WMS-LM and as such be deemed as an unofficial alternate form, based upon the prior research studies. The premise of Study 2 is to understand if an alternate form of the Morris stories mirrors the performance of the original WMS-LM in a screening cohort across countries.

Study Outlines

Study 2

Key to understanding the interplay of AD pathology on participants memory ability and memory performance, is to look for congruencies and differences when observing memory indices within these populations. The concurrent and convergent validity of these measures will be assessed using statistics for measuring agreement between measures. The second study of this thesis looks at ascertaining the concordance of two story recall measures, taken from a memory clinic cohort looking to be screened for a clinical trial. This was a multicentre study carried out across three countries in two languages (English & Spanish).

Study 3

The third study of the thesis will look at a different measure construct for assessing in a confirmed amyloid positive population. This will be studied across two, word list memory measures, with differing levels of ecological validity. The concurrent and convergent validity of these measures will again be assessed using the same statistics for measuring agreement between measures. By comparing word list recall which has been tailored to each location and language and a test using direct translations, testing paradigms often expected to yield identical test can be better understood. Also, within participants who have the prevalence of amyloid pathology, these measures can also be explored for the absence of practice effects across the disease, which is important given the absence of improvement is often found as a hallmark of early AD. This will further help explore the criterion validity of measuring memory with these tests within AD.

Overarching Objective

By assessing the concordance of episodic memory measures based around two divergent constructs in two populations, this can help tease out differences between measures widely utilised and often deemed to be similar. Overall, these findings will help explore the construct and content validity of these memory measures allowing greater understanding of their inherent measurement of cognition.

STUDY 2

Methods

Study details

This study was designed to understand the relationship between two measures of episodic memory impairment commonly used for indexing MCI deficits thought to be related to AD. Participants in this study were pre-screened at sites conducting the Eisai Clarity AD phase III clinical trial. They first underwent testing on the Alternate Story (AS; Morris Alternate Paragraphs) paradigm before being screened in full for Clarity AD. At the investigator's discretion, but primarily based upon scores on the AS measure, subjects were then screened for the Clarity AD study whereby they underwent cognitive testing on the WMS-LM. Clarity AD was an 18-month treatment, multicentre, double-blind, placebo-controlled, parallel-group study with open-label extension in subjects with early AD. To be eligible for the study, subjects must have objective impairment in episodic memory as indicated by at least 1 standard deviation below age-adjusted mean in the WMS-LM. Site raters were trained to undertake both measures and were given stringent guidelines on how to administer both measures and importantly, the time between administrations. For the WMS-LM, all administrations were recorded and reviewed for accuracy of administration and scoring errors by a clinical psychologist. For the AS measure no review took place and was undertaken at the direction of the sites principal investigator (PI).

Assessments

For this study the two measures of story recall were the WMS-LM & AS. Each measure had three different stories, one short form and two long form stories. For individuals that were aged under 65 both long form stories were used. This began with a trained rater reading out the first long form story in full, the individual then had to recall as much of the story as they could. This process was then repeated for the second long form story. Between 20 and 30 minutes then elapsed, without any intermediary cognitive testing, and individuals were asked to recall as much of the first story as they could remember. If no details could be recalled a set prompt was given and an opportunity was given for recall of that story. This process was then repeated in full for the second story. Scores were taken from a sum of both immediate recalls and both delayed recalls to give representative scores for each condition for each individual. For those over 65, the first short form story was replaced by a short form story totalling half the number of sentences (3) of the long form story (6). This short form story was repeated twice during the immediate recall condition. The rater read the story, asked for as much information as could be recalled, before repeating the same story again, then again followed by an immediate recall of the same story. The same 20-

30 minute delay occurred before a single recall of the short form story (however, no prompt was allowed for this short form story), followed by a single recall of the long form story (this was the same set prompt as given in the younger age group condition).

This paradigm is identical for both the WMS-LM and AS. Both measures were employed to measure memory ability in the same cohort. Individuals underwent testing on the AS first then undertook the WMS-LM. The time between assessments was recommended to be between 3-8 weeks (21-60 days). Initial analysis comprised all individuals, however if this testing paradigm was diverged from then sessions outside of this range were put into group 2, with those within the suggested time range in group 1, in a subsequent sensitivity analysis. The AS forms were given to sites in the appropriate for country and primary language, who were participating in the Clarity AD phase 3 clinical trial. Sites staff were trained on the AS by the author. No prior experience of scales was mandated for those administering this measure. Subsequently, at the investigator's discretion, but partly based upon scores on this measure, subjects were then consented, and formally screened for the Clarity AD study, whereby they underwent cognitive testing on the WMS-LM. To note the WMS-LM were direct translations of the original text with only names changed. This is in direct contrast to the linguistic translations for the AS which were subtly changed to reflect culturally and linguistic properties pertaining to the language in question. The WMS-LM was part of a battery of tests for inclusion into the clinical trial. This battery included the MMSE, WMS-LM and the CDR.

Participants

In total 196 individuals underwent testing on the AS and then subsequently the WMS-LM. Each individual had undergone a fully informed consent process consistent with each countries IRB and EC guidelines. Those who were screened for this study underwent a full consent process both prior to the AS being undertaken then further full study consent prior to the administration of the WMS-LM. The screening criteria for inclusion into the study mandated the exclusion of medication effecting cognition, known comorbidities prior to brain imaging and depressive illness within the last 5 years. These exclusionary factors were investigated prior to the first measure being undertaken within this study. Individuals were seeking consultation for suspected memory impairment with the view to screening for a clinical trial for AD. To note none had confirmed amyloid pathology or progressed signs of AD at the time of entry into this study. The cohort comprised of 161 US English speakers, 17 US Spanish, 4 Canadian English and 14 British English. Each individual had language appropriate forms of each measure.

Statistical Analysis

The primary analysis looks to ascertain the concordance of two story recall measures commonly used within MCI and early AD populations. The AS measure was developed on the basis of being a direct analogue of the WMS-LM measure. Psychometric analysis was undertaken for the comparison of these two measures, this began with simple t-tests in order to look at group differences. Correlation analysis was then run to look at the strength of the relationship and finally Bland & Altman (1968) methodology for looking for agreement between two measurements methods was run (*a full discussion of this methodology and why it was selected can be found in* ***Chapter 2****).* By looking for group differences and the strength of the relationship between the two measures on an individual level, this can help better uncover if participants display a concordant level of performance across two measures that have identical underlying psychometric constructs and properties. Given the large range in durations between assessments and the potential for this to confound any findings, participants were subsequently split into those who underwent WMS within 20-60 days of AS testing, which was per guidance and those outside of this range of dates, both less than 20 days between administrations and 60 days or greater.

Results

Group data is displayed in **Table 4.1.** Descriptive statistics for the whole group and the subsequent split groups was run. To note 80 subjects had no date for pre-screening so were excluded from the second analysis. Group level analysis showed significant differences between the delayed recall from the alternate Morris paragraph and from the logical memory stories ($t(1,195) = -4.59$, $p=0.0001$). Correlation analysis revealed a significant, yet moderate, positive relationship between the two measures of delayed recall ($r=0.64$). Individuals scored nearly 2 points higher on the logical memory as compared to the similar AS measure.

Table 4.1. Descriptive & psychometric properties of all groups

	Whole Group	Within Pre-screening Guidelines (Group 1)	Outside of Pre-screening Guidelines (Group 2)
N	196	49	67
Alternate Story Mean	6.41	6.31	6.15
Logical Memory Mean	8.14	9.00	7.72
Alternate Story SD	4.65	4.22	4.16

Logical Memory SD	6.87	7.10	6.11
Bland & Altman Bias	-1.74	-2.69	-1.57
Bland & Altman Limit Of Agreement Range	8.62 – -12.10	10.47 – -15.86	8.07 – -11.20
r	0.64	0.60	0.39

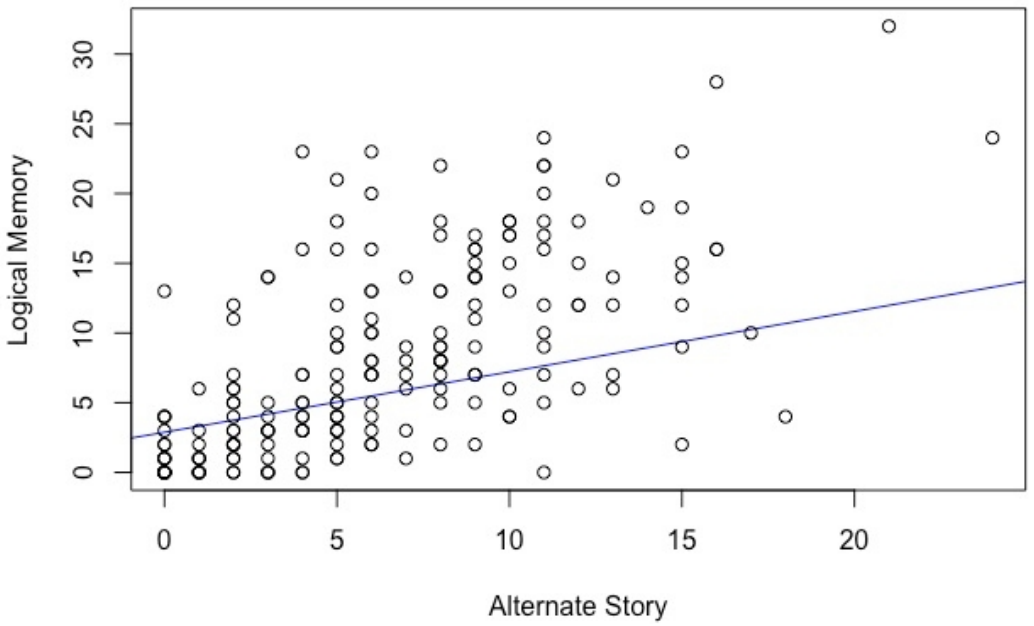


Figure 4.1. Scatter plot of individual scores on Logical Memory & Alternate Story Recall. r=0.64

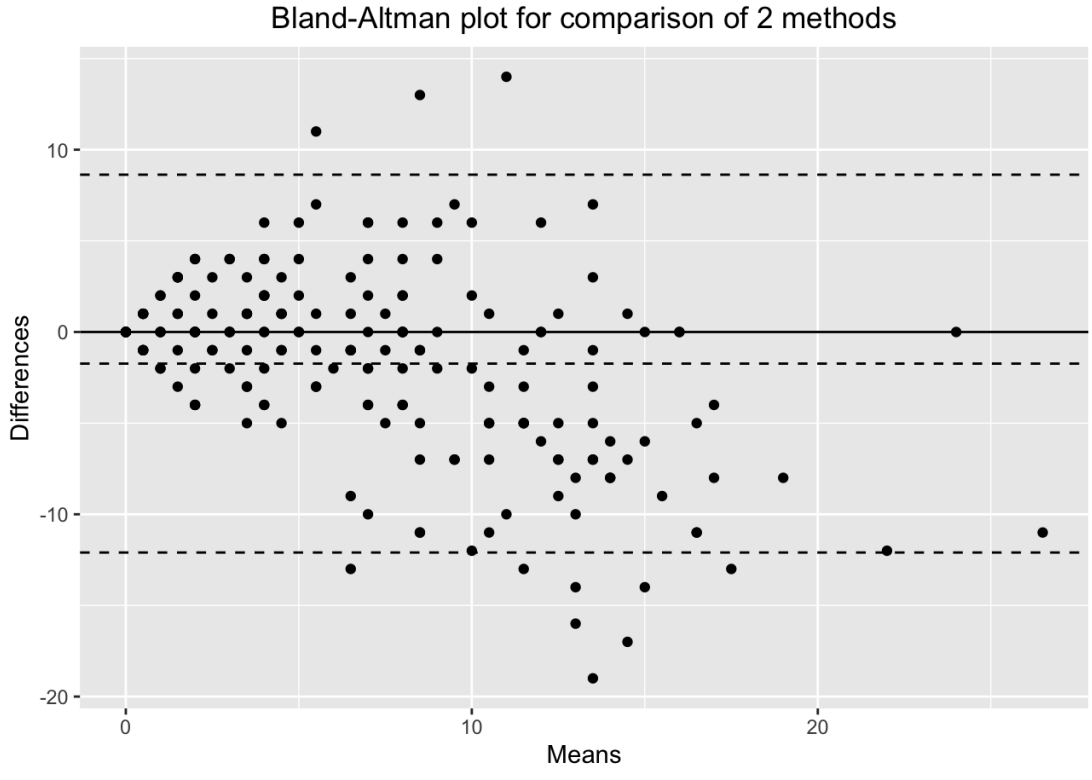


Figure 4.2. Plot of individual mean scores against individual differences between 2 measures

Further analysis utilising Bland & Altman (1968) methods of agreement showed reasonable agreement between these two measurements, with over 95% of subjects falling within bounds. The overall group had a mean difference of -1.73 points, however the limits of agreement were wide. An upper limit of agreement of 8.63 and a lower limit of agreement of -12.1.

Looking into these findings further, individuals were grouped into categories based upon duration between administrations. Individuals who fell between 20-60 days of an interlude between administrations were grouped as 1(within guidelines) and those outside of this timeframe grouped as 2(outside guidelines). 67 subjects were in group 1 with a range of interlude of 21-60 days. In group 2 there were 49 subjects who had an interlude range of 0-20 & 61-232 days between assessments. 80 subjects had no date for their administration of the initial assessment and were excluded from this analysis.

Dichotomising the cohort by this grouping, all analysis were re-run. Again significance differences were found between measures across both groups [1($t(1,48)=6.05$, $p<0.00$); 2($t(1,66)=-2.61$, $p=0.01$)]. Pearson correlation coefficient values were again significant for both groups, however, both values were below the overall group value [1 ($r=0.387$, $p=0.01$); 2 ($r=0.600$, $p<0.00$)]. These two independent group level correlations were compared and were not significantly different from each other ($p=0.139$).

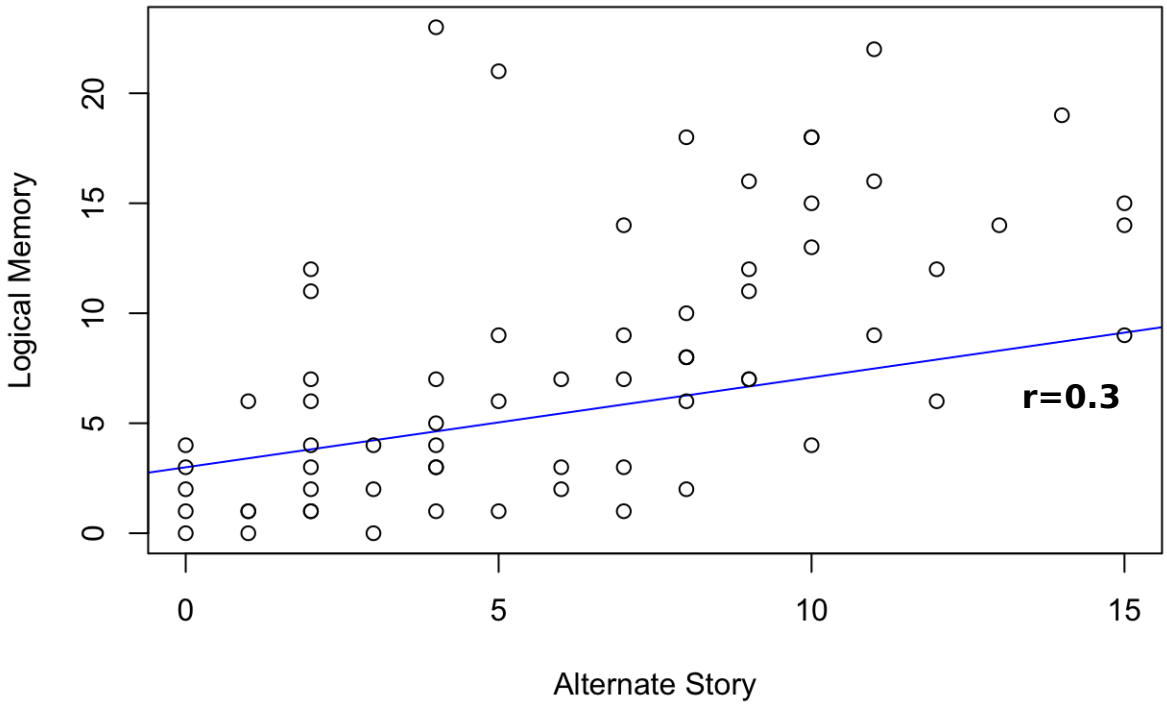


Figure 4.3. Group 1 Scatter plot of individual scores on Logical Memory & Alternate Story Recall.

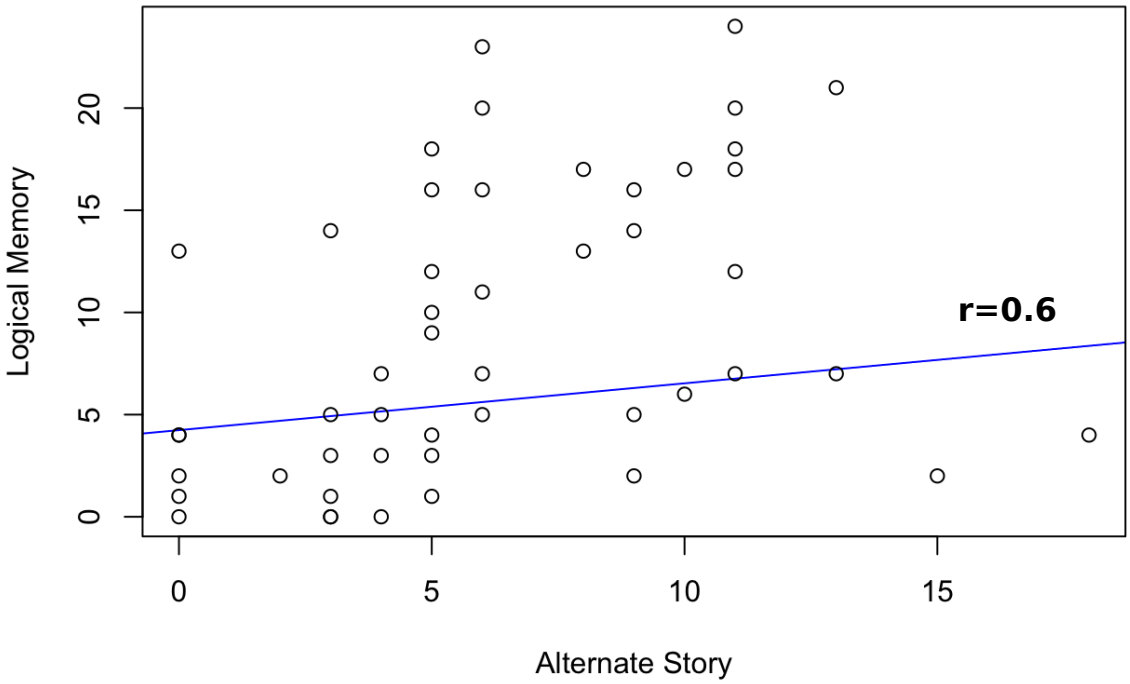


Figure 4.4. Group 2 Scatter plot of individual scores on Logical Memory & Alternate Story Recall.

Agreement between these two measures by group utilising the Bland & Altman methods showed poor agreement between these two measures

within both groups with much larger upper and lower limits of agreement [1(MD=-2.69, ULOA=10.46, LLOA=-15.86); 2(MD=-1.57, ULOA=8.07, LLOA=-11.20)]. However over 95% of subjects fell within the limits of agreement in both groups.

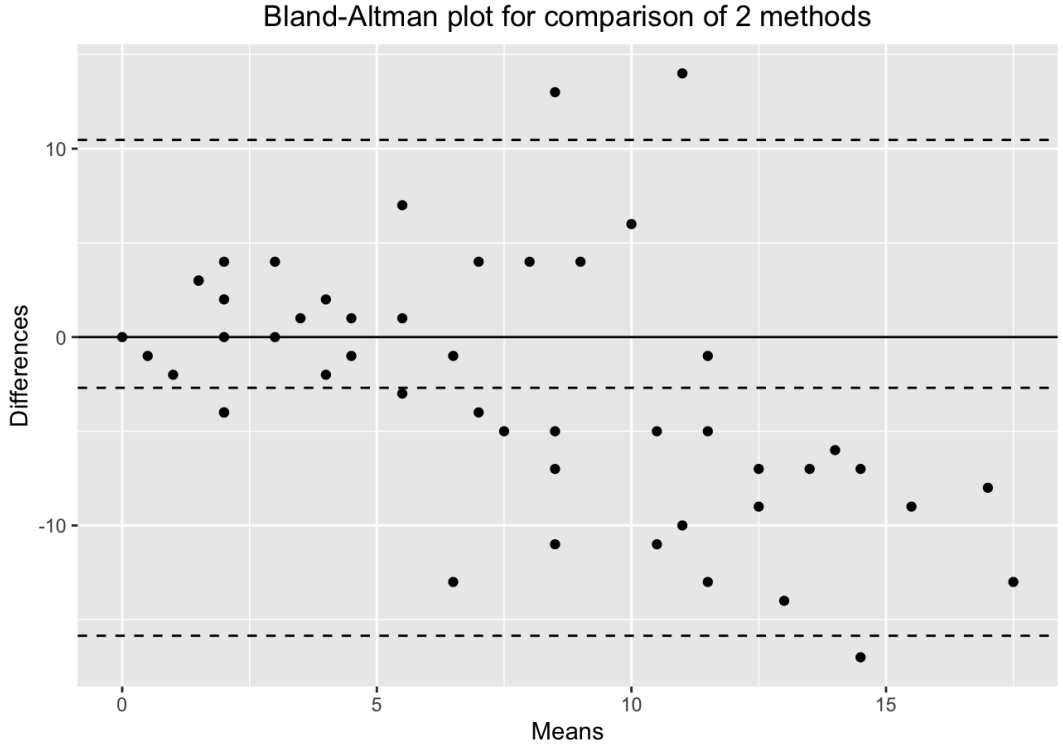


Figure 4.5. Group 1 Plot of individual mean scores against individual differences between 2 measures

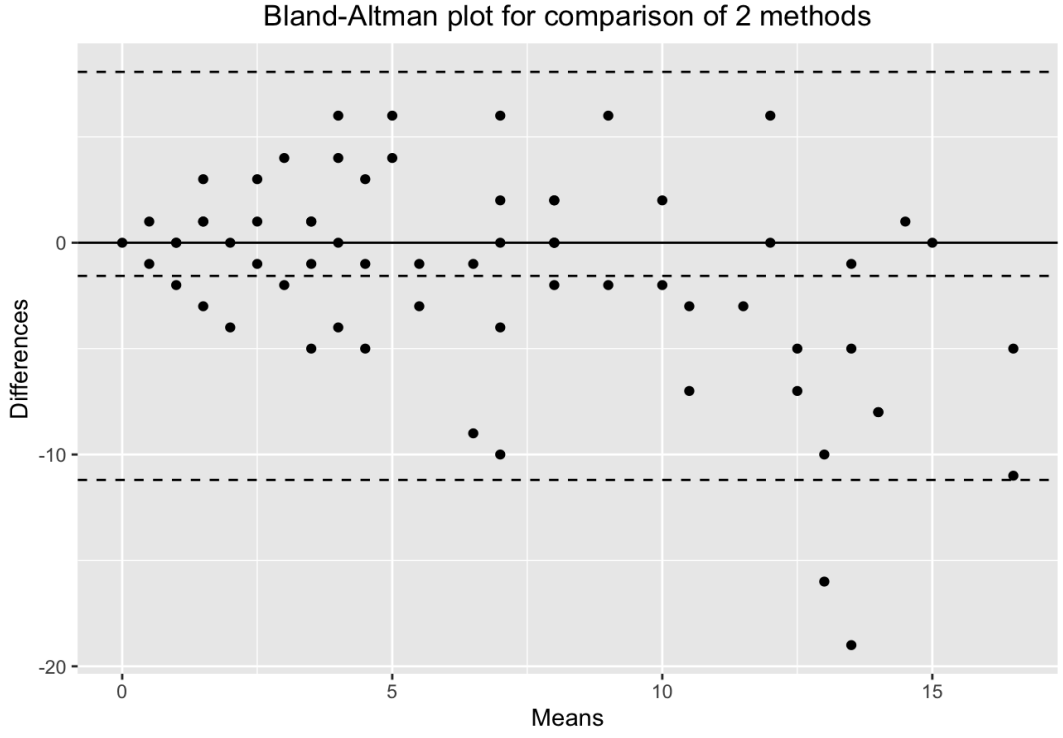


Figure 4.6. Group 2 Plot of individual mean scores against individual differences between 2 measures

Study 2 Conclusions

This study demonstrates the concurrent fundamental underlying constructs between the AS and WMS-LM. Overall, the agreement between the tests was good. However, this was negatively influenced by the lack of adherence to the administration timelines. This measure shows strong concordant scores to that of the logical memory paradigm from the WMS. However, it should not be described as an alternate form, based on these results. The applicability for wider use of this story recall is well founded but requires further research within the bounds of administration guidelines. This was the first use of this measure in a number of new cultural and language settings, within which, these findings support future wider use within similar settings in the future. The results show clear evidence that repeated testing within a certain timeframe is a pre-requisite for homogeneity between these two measures.

One factor not captured within this cohort was a level of education. There may be some who have been highly educated and therefore performed better than someone else at the same stage of AD, or have more progressed pathology and therefore greater variation of performance from visit to visit. However, without these data the effect biomarkers for AD may have is unknown. Given that these subjects are not fully worked up to the point of a diagnosis of AD or MCI, it is difficult to draw strong conclusions for performance between the two measures as there is likely high clinical variability within the cohort. Factors such as undiagnosed depression and poor cardiovascular health are common within the elderly and often have similar cognitive difficulties to that of early AD/MCI (Ferreira et al., 2020). There are also the probable divergent AD biomarker profiles within this cohort to consider. As discussed earlier, those at the preclinical stage of AD often exhibit impairment in practice effects which might explain some of the variance between the two measures, however without these additional data points for individuals it is difficult to suggest the level of influence this may have on the results.

There are often many common factors that can affect cognitive testing that happen to everyone, not just those with MCI/AD. These include duration of sleep, dysphoria, caffeine intake, stress levels and diet, all of which influence test performance. Intra-individual variation is expected to a small degree due to these factors. However, this is exacerbated due to the population in question displaying significant memory impairments. For example, with the widely utilised MMSE measure, 3/4-point variations are commonly seen across yearly visits in cognitively normal subjects (Hensel et al., 2007; Clark et al., 1999). Repeated testing can be influenced to a greater degree to these environmental confounds from one visit to the next, on the basis of

chance alone, whereas, those of greater duration tend to have greater preparation for them and as such are less susceptible to chance and can be rescheduled if required. Another factor outside the control of the study was rater change between the two measures. All raters were trained on both measures by neuropsychologists and whilst some variation could be expected between administrations, the non-subjective nature of the scoring guides meant variance due to rater change should have been minimal.

The fixed design of this study meant the administration order of the two measures was fixed as counterbalancing was not possible. This was due to the scenario of screening for the clinical trial in question which utilised the WMS-LM. As such, the AS measure had to be utilised first as a prescreener as not to engender bias or variance in the screening procedures for the clinical trial. This study design issue may explain the increased score on the WMS-LM as it was always administered second. Although other factors such as differences in translation method may also have factored into this. The WMS-LM was a direct translation, whereas the AS was subtly changed to take into account cultural and language specific phrasings. This may have resulted in improved recall for the WMS-LM in comparison to that of the AS as words or sentences that may not be akin to everyday speech could be easier to recall. However, due to the predominance of US English in the cohort this may have only be a small factor in the study's findings.

This study shows that whilst these aforementioned factors are taken into account, particularly duration between administrations, the performance and concordance on these measures are not exact. This indicates the need for constrained administration procedures when repeatedly testing participants and a concordance between the two translation processes. Enforcing these constraints would no doubt lead to an improvement in concordance between these two measures as shown with the group sensitivity analysis. However, even with these constraints, there still wouldn't be an exact correlation between scores across visits. Therefore, the AS should be seen as a close analogue of the WMS-LM, but not an alternate form in these different languages. Though, even without counterbalancing the administration order, there was a consistent improvement on the WMS-LM from the AS on all analysis. This shows there are clearly some underlying commonalities between the two measures. Further research is needed with the confounding factors described being accounted for and minimised.

This first study shows that even within identical test paradigms of story recall, these measures have far from the level of concurrent validity needed to be used interchangeably within this population. Whilst there is a level of convergent validity between these two measures, the way they are used suggests the performance should mirror one another more closely, the results indicate this is far from being the case within a memory clinic population for screening for signs of early AD.

STUDY 3

The second study will aim to look test variance in a cohort of individuals with concordant amyloid profiles, who were part of a fully diagnosed clinical trial population. This comparison will look at two similar measures of word list learning, the ISLT and the word list recall from the ADAS-Cog. Both verbal list learning measures require participants to learn new words with only the salience of the word lists differing between the two measures. The ISLT looks at ecologically valid words akin to a shopping list specific to the country and language of administration, whereas the ADAS-Cog word list is set and contain random words which is directly translated for each language it is administered in.

Method

Study Details

This study is designed to assess the fundamental relationship between two, word recall measures within a diagnosed, amyloid positive cohort of AD individuals. Site raters were trained to undertake both measures and were given stringent guidelines on how to administer both measures. The time between administrations was determined by the speed that a site managed to undertake the full complement of screening procedures for the clinical trial in question (detailed in full in **Chapter 2**). This was an average of 52 days from the initial administration of the ISLT to that of the ADAS-Cog_{14} (26-103 day range). All subjects were from the randomised population of the MissionAD program. The MissionAD program in Early AD was a Phase III program conducted across more than 500 sites in 29 countries and was designed to assess the efficacy of elenbecestat. Diagnosis of MCI due to AD or Mild AD was required, an MMSE score ≥24 and a CDR Global score of 0.5 and a CDR Memory Box score of ≥0.5. Cognitive impairment of at least 1 SD from age-adjusted norms was also required and the objective test of episodic memory that was used in these studies was the ISLT. Once these assessments were passed, confirmation of brain amyloid pathology by either amyloid PET or CSF assessment or both was also required as well as blood, ECG and medical examinations which had to be passed (Roberts et al., 2020).

Assessments

For this study two measures of verbal list learning were used to compare performance within the same group of individuals. Both measures require the list of words to be read out loud to an individual. Upon the completion of this, the individual is then asked to recall as many words from that list as they can. Correct answers, as well as incorrect words recalled, are counted. This is then repeated for three trials. A total for immediate recall is then summed for all of the trials giving an overall score. A delay or intermediary

period then takes places which lasts between 20 and 30 minutes. During this time other cognitive testing that is non-verbal and non-memory reliant is undertaken. After this delay individuals are then asked again to recall as many words as they can from the word list they have been told prior to the delay. This is a single trial with the total from this comprising a delayed recall score. This paradigm is the same for both of these measures with only the word list length (10 for the ADAS-Cog, 12 for the ISLT), word salience and distractor tasks, differing between the two.

The first measure is the ISLT, which was undertaken as part of the screening procedures of a Mission AD clinical trial. The ISLT is an ecologically based verbal memory measure designed to mirror a real-life shopping list by having culturally and location relevant lists (Lim et al., 2014). 12 words are shown initially with three trials of immediate recall for subjects to retain the full set of words. Distractor tasks are then undertaken with a set 20-30 minute delay before the delayed recall condition is administered (Thompson et al., 2012). These distractor tasks begin with a basic detection task to orientate an individual with a card paradigm. This is followed by a simple reaction time task. The final two measure incorporate aspects of working memory in a one back task, followed by a one item learning task. All of these four distractor tasks are collectively known as the Cogstate Brief battery (CBB). This battery has been widely studied both in healthy adults and those with MCI/AD. Importantly these four tasks are non-verbal, with the stimuli being playing cards. Therefore, they do not interfere with the verbal recall of ISLT word list. The ISLT word list shown to individuals never varies in length but always varies in context. The words shown will be randomised from a bank of 50 for each language but also location. For example, there is a Korean word-list for those in America. This is how the test has greater ecological validity than the ADAS-Cog.

Approximately five weeks after the ISLT was administered, the ADAS-Cog word list was then administered to the same participants as part of the wider ADAS-Cog battery. Each participant is given three trials to learn a list of ten high-frequency, high imagery nouns. A delay of approximately 20 minutes occurs between the immediate and delayed recall conditions, this is based upon the length of two distractor tasks, commands and constructional praxis (full details of these tasks can be found in **Chapter 2**). These two measures are fundamentally divergent measures to that of word recall and are do not engage any aspect of verbal or episodic memory. The order of administration of the ADAS-Cog is set and has been widely studied within healthy controls and MCI/AD individuals ever since its inception (Baker et al., 2017).

Participants

There were a total of 2176 individuals that undertook all procedures within this study. Each individual had undergone a fully informed consent process consistent with each country/states IRB and EC guidelines. They had been

screened for other MRI, blood, ECG or medical abnormalities, had prior confirmed cognitive impairment, a confirmed threshold (or greater) presence of amyloid beta (assessed through CSF collection via lumbar puncture or PET imaging) and the absence of any other co-morbidities and taking any medication, known to influence cognition (again full details of this cohort are contained within **Chapter 2**). Individuals within this cohort were recruited at over 500 trial sites in 29 different countries making this a highly diverse population. As shown in the **Chapter 3**, the individuals screened for this trial had diverse cognitive profiles however, the individuals within this cohort are only those who passed all screening measures and as such all have concordant clinical and biological presentation. This means that whilst there is known regional variation in screening populations such as this, the cohort here under investigation can be analysed in full as the variations seen at screening are not present for this cohort.

Statistical Analysis

Given the nature of the cohort, the analysis will look to expand upon and follow similar statistical methods to **Study 2** in this chapter. The primary analysis looks to ascertain the concordance of two verbal recall measures one specifically developed to be an ecologically valid measure and the other a stalwart of every AD study undertaken both in academia and clinical trials. Understanding the relationship between the more ecologically valid ISLT and the ADAS-Cog wordlist within a clinically diagnosed AD clinical trial population who are amyloid positive, is key to furthering the understanding of these verbal memory measures. Psychometric analysis was undertaken for the comparison of these two measures, initially this will be correlation analysis to look at the strength of the relationship and then Bland & Altman (1968) methodology for looking for agreement between two measurements methods will be run. Finally given the additional data (including APOE information on individuals) within this dataset, MANOVA analysis will be performed controlling for a number of confounding variables known to influence cognitive performance within AD.

Results

Table 4.2. below shows the breakdown of the demography of the group under investigation in Study 3. Group level comparisons are not made here as all subjects underwent the same testing battery and did not do so under different testing conditions or order.

Table 4.2. Descriptive Statistics

Raw Whole Group Statistics	Value/Mean (SD) [Range]
N	2091
Age	72.0 (7.1) [50-85]

Sex (Female %)	51.3%
Years of Education	13.5 (4.1) [0-28]
APOEe4 +	63.3%
ADAS-Cog Immediate Recall* (/10)	5.3 (1.4) [0-9.33]
ADAS-Cog Delayed Recall* (/10)	7.1 (2.3) [0-10]
ISLT Immediate Recall (/36)	14.3 (3.7) [3-29]
ISLT Delayed Recall (/12)	2.8 (1.9) [0-9]

*ADAS-Cog is reversed scored so values represent the number incorrectly recalled.

In order to allow for direct comparisons of the two measures the ADAS-Cog scores were reversed so that the score for each individual represents how many correct answers they recalled, with graph axis's mirroring one another. Immediate scores were also multiplied by three as the scoring convention takes a mean of the three immediate recall trials. To note scores for each scale had different maximum scores the ISLT contained a possible twelve words to recall for each trial whereas the ADAS-Cog has ten words across both recall conditions.

Correlation analysis was run to ascertain the strength of the relationship between the two measures by recall condition. This showed statistically significant moderately strong positive correlations between both measures in both conditions, with delayed recall having a stronger Pearsons r value of r=0.55 (p<0.001), compared to r=0.46 (p<0.001) for the immediate recall condition. The correlations for each condition (immediate and delayed) were significantly different from each other (p<0.001).

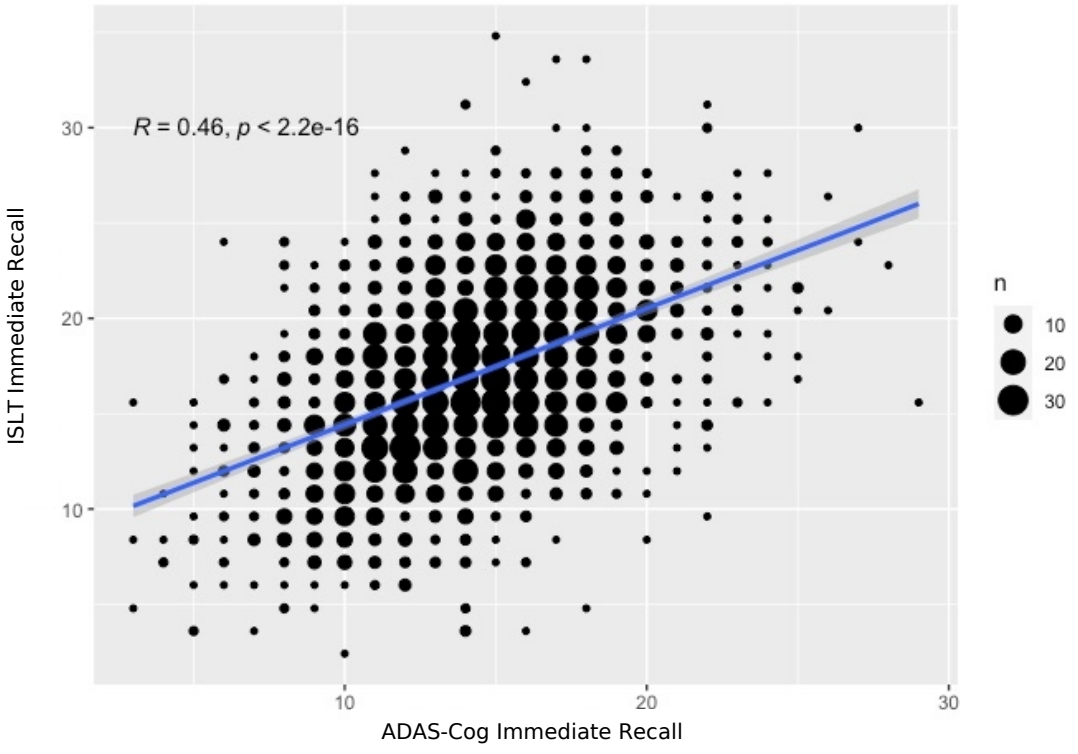


Figure 4.7. Scatter plot of individual scores on immediate recall from ISLT & ADAS-Cog r=0.46

Bland and Altman analysis was then run which showed uniform levels of agreement between the two scales on the immediate recall condition (**Figure 4.9**). Less than 95% of individuals fell within the bounds of agreement indicating poor agreement between these measures. This was also run for the delayed condition which again showed uniform agreement between the two scales (**Figure 4.10**). The upper and lower limits of agreement were much smaller for delayed recall in comparison to the immediate condition [Immediate Recall (MD=2.75, ULOA=11.64, LLOA=-6.15); Delayed Recall (MD=0.67, ULOA=5.20, LLOA=-3.86)]. However, again less than 95% of individuals fell within the bounds of agreement.

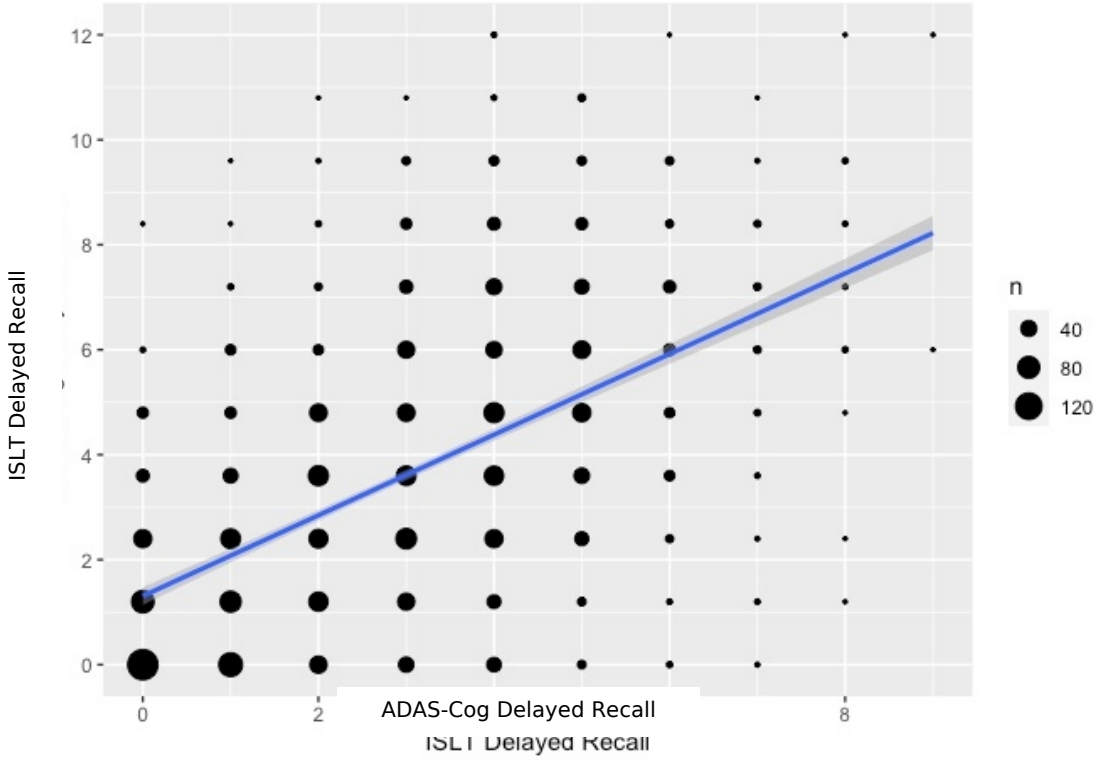


Figure 4.8. Scatter plot of individual scores on delayed recall from ISLT & ADAS-Cog. r=0.55

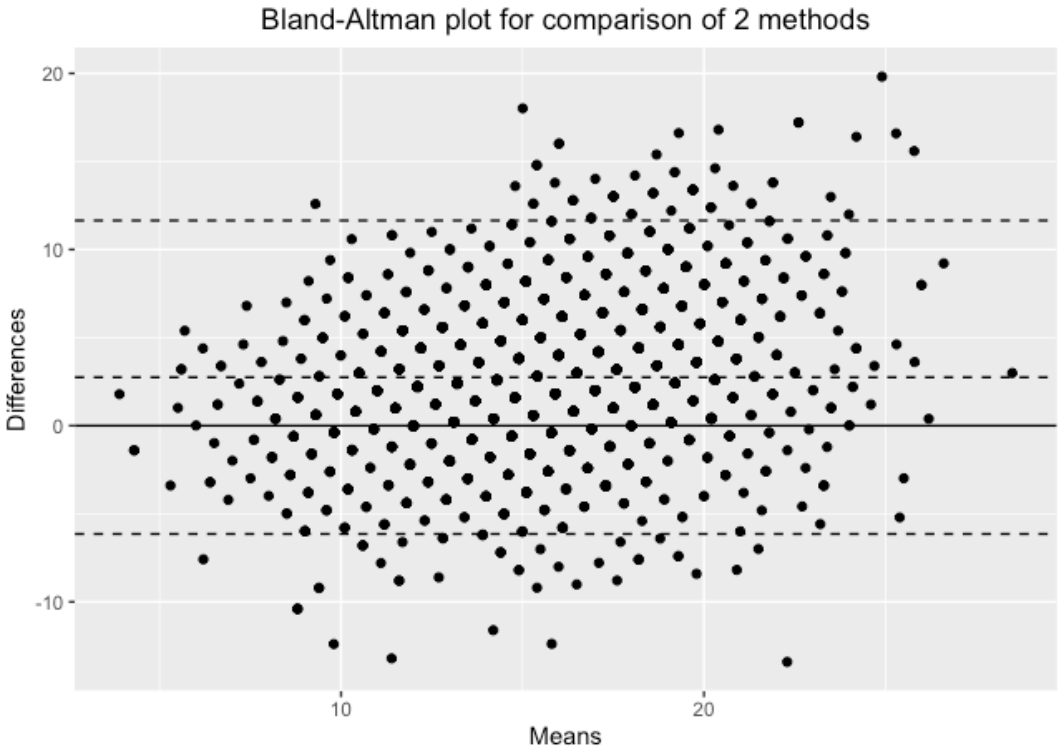


Figure 4.9. Plot of individual mean scores against individual differences between 2 measures of immediate recall

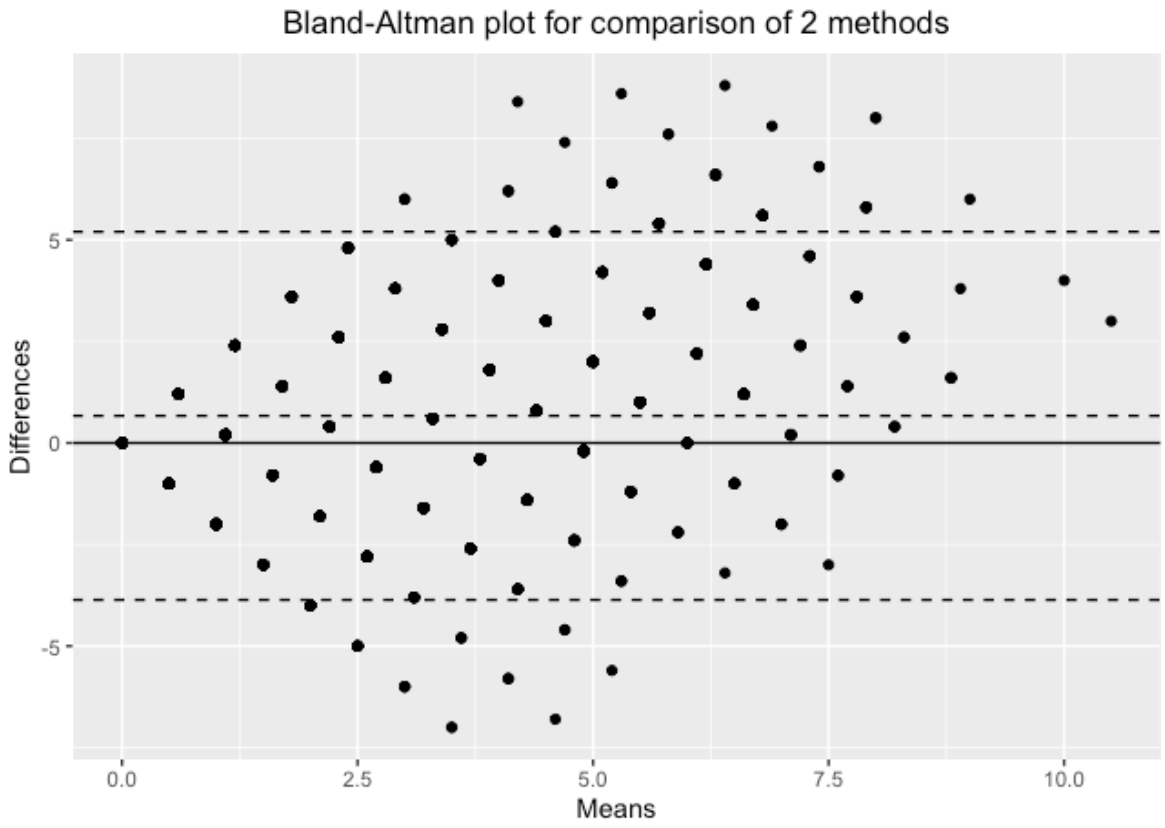


Figure 4.10. Plot of individual mean scores against individual differences between 2 measures of delayed recall

Finally, in order to understand the influence of APOE status (e4 carriers versus non-carriers) had on the performance, both on delayed and immediate recall indices from both measures, a MANOVA was run for each condition (immediate & delayed) across both measures. Within the MANOVA the independent variables were APOE status (carrier vs non-carrier) the dependant variables were the memory recall measure scores, which were either both immediate recall conditions or both delayed recall conditions from the WMS-LM and the AS.

The multivariate result was significant for APOE status for delayed recall, Pillai's Trace = .03, F = 13.99, df = (2,2063), $p<0.001$, indicating a difference in the performance on both delayed recall measures by APOEe4 carriers and non-carriers. However, the result was not significant for immediate recall, Pillai's Trace = 0.00, F = 0.25, df (2,2063), p=0.91, indicating an absence of significant differences between APOEe4 status groupings and immediate recall performance across both measures.

Study 3 Conclusions

This study demonstrates the intricacies of the relationship between two fundamentally similar tests of immediate and delayed memory recall. Given the well characterised cohort, being amyloid positive, with confirmed AD, (with level of cognitive and functional impairment fundamentally similar) this allows for wider generalisation of the findings of the analysis. Overall, there was poor agreement between the two measures, with moderate yet statistically significant correlations between the two measures, with a slightly stronger relationship between the delayed recall measures. The significant effect of APOEe4 status has on the delayed recall on both measures was also demonstrated. This further emphasises the importance of factoring this key AD risk allele into analysis and investigations within any AD population. This also suggests further subtlety of this risk allele in the interplay of cognition across the disease course. And confirms that at this stage of AD, APOE carrier status has an effect on longer term memory recall.

There are some drawbacks in the analysis, with common factors associated with changes in cognitive performance such as level/years of education, age, gender and geographical location, have not been taken into account. These factors should be considered when looking at subsequent analysis and should be considered when assessing the applicability of the results from this study. They were not incorporated into this analysis due to the statistical method employed.

The broader common issues associated with variability in cognitive performance are important to acknowledge. This includes duration of sleep, dysphoria, caffeine intake, stress levels and diet. These factors can be influenced to a greater degree to these environmental confounds from one visit to the next, on the basis of chance alone, whereas, those of greater duration tend to have greater preparation for them and as such less susceptible to chance and can be rescheduled if required. The size of the effect of each of these issues is small however, they are factors that can affect intra-individual variation on the relationship between two measures. This is something that is also exacerbated within the population in question, as all subjects have significant memory impairments as well as confirmed AD pathology and diagnosis.

As discussed previously the absence of practice effects within repeated testing is often seen at the early stages of the disease (preclinical AD), however, this is clearly not seen here as there is clear group level improvements between measures. This absence of practice effects at this stage of the disease confirms prior findings in this regard. Although this group level improvement can be argued to be, at least in part, due to an

issue with the study design was the absence of counterbalancing the order of assessments. It was not possible to swap the administration order of the two measures in a 2x2 factorial design, due to the scenario of screening for the clinical trial in question. The trial utilised the ISLT as the initial stages of inclusion and the ADAS-Cog was undertaken at a subsequent clinic visit. The study design was the set order of administration for this study. This resulted in higher scores on the ADAS-Cog verbal recall measures (immediate & delayed) compared to the ISLT, which can clearly be argued to influence the strength of the relationship between the two measures.
Fitting this into a broader cognitive picture, it is often found that the largest practice effects seen across repeated administrations is seen between the first and second administrations, which is the same as these two administrations analysed here. In future it would be of benefit to look for subsequent longitudinal assessment to mitigate against any influence practice effects may have had on this analysis, as well as counterbalancing the assessments.

Another factor outside the control of the study was rater change between the two measures. However, all raters were trained on both measures, which was conducted by trained professionals and whilst some variation could be expected between administrations, the non-subjective nature of the scoring guides necessitated a constrained approach to measurement of performance of these subjects. The ISLT rater certification was less stringent than that of the ADAS-Cog, with no prior level of qualification needed to administer the ISLT due to its computerised automated administration, there is a small likelihood there was a different rater for the two assessments. The ADAS-Cog required a much greater level of prior experience as well as training on the measure. However, given the strict requirements for the ADAS-Cog rating and automated administration of the ISLT, rater change here would not be enough to engender large discrepancies that would significantly influence the relationship between the two measures.

Fundamentally these two measures are designed as such that they index the same cognitive domains and have concordant paradigms. The findings show there is still variation between the two measures even within a highly homogenised population. The subtleties in assessment composition may go some way to account for this. The ISLT is designed to be highly analogous to everyday memory tasks, with strong ecological validity for the words recall in each country and dialect spoken (Lim et al., 2014). This is not the case for the ADAS-Cog, where words are translated from the original word list and have no ecological validity to normal day to day tasks or indeed shopping. These subtle design differences may go some way to elucidating the statistical differences between the measures and is important to consider when interpreting clinical findings within this population. As discussed above often, similar measures such as these are habitually interchanged and interpreted to show exactly the same impairments, when the findings from

this study would suggest the opposite to be the case. There is clear variation within a highly homogenous population. As such this indicates these subtle differences in psychometrics within AD are worth considering when interpreting any cognitive or clinical endpoints and should not be used interchangeably in the future.

Further to this, there is a clear influence of APOEe4 carrier status has on delayed memory recall. As shown widely in prior research, APOEe4 status infers greater levels of cognitive impairment, faster cognitive decline and divergent disease progression to those who are not APOEe4 allele carriers. Recent research suggesting a different disease classification for APOEe4 carriers to that of non-carriers (Frisoni et al., 2022) would also be supported by this finding as, within this homogenised population, there are clear cognitive differences purporting to divergent disease trajectories (however this would need further longitudinal research to confirm this). Given that not all amyloid positive subjects go on to develop dementia, clear differences exist within our current understanding of AD diagnosis and pathology which should be taken into account in further studies and clinical trials alike. Overall, this study demonstrates a moderate and imperfect concordance between two measures commonly used to index memory recall within AD populations. The divergence in this relationship is argued to be down to three main drivers of variance. Firstly, the fundamental constructs of the two measures have different levels of ecological validity which would enable improved recall on the ISLT which has greater ecological validity due to the familiarity of the stimuli utilised. This however was not what was found within this analysis, the opposite was in fact the case. This is argued to be down the influence of the other two drivers of variance. Firstly, the lack of counterbalancing of the two measures led to an improvement across the two measure with inflated scores found in the ADAS-Cog compared to that of the ISLT. Secondly, the influence of APOEe4 carrier status was statistically significant across the delayed components of both measures. Indicating this risk allele to factor into the homogeneity of the studied population and thus inducing variance on cognitive performance across measures. Nevertheless, this study demonstrates clear evidence of moderate concordance between these measures and highlights the importance of understanding the fundamental constructs of any measure being employed within cohort studies, for the evaluation of findings and results within an AD population such as this. It also emphasises that these measures should not be treated as analogues of one another.

Chapter Conclusions: Studies 2 & 3

This chapter aimed to uncover the concordance and relationships between measures that are widely used interchangeably to uncover memory impairment across the spectrum of AD. Overall, the measures show fair to moderate correlations with moderate to poor levels of agreement. The

statistical relationships between the two pair of measures all are imperfect and far from concordant enough to allow for direct comparisons of “amnestic impairment” within populations across the spectrum of AD.

Nevertheless, taking into account the issues with counterbalancing, that were unavoidable across both studies, the analysis the AS and LM measures show to be good, but imperfect analogues of each other when undertaken within the administration guidelines. However, given the divergent ecological validity of the ISLT and ADAS-Cog it could be argued these to be fundamentally different measures, however the results of this study would suggest similar levels of concordance to that of the AS and LM which share indistinguishable constructs. However, the cohort for the AS & LM is highly heterogenous in comparison to that of the ISLT and ADAS-Cog, which is highly homogenous and argued to comprise subjects at a limited stage of AD. Whereas the AS & LM cohort included subjects without diagnosis or that were worked up and as such where likely not all subjects with AD. These differences in cohort composition, are important when understanding the implications of the findings from the two studies.

Looking at the Bland & Altman analysis across the two studies, clear similarities exist across both sets of measures. However, the large limits of agreement of the AS & LM cohort are suggested to be in part due to a fundamental lack of adherence to the administration guidelines and heterogeneity of disease pathology within the cohort itself. The limits of agreement within the ISLT & ADAS-Cog cohort again show the imperfect concordance of the measures that fall within this amnestic paradigm. This further emphasises that these measures share some convergent validity but differ enough to have an absence of concurrent validity. The homogeneity of the cohort can be strongly argued to influence the concordance of the relationship between cognitive measures, as this second analysis was highly homogenised both cognitive and clinically. The Bland & Altman analysis also demonstrated the influence that the lack of counterbalancing had on the concordance between the two measures in both studies, with both sets of analyses showing improvements in the scores for the measure administered second. This was shown by the Mean Difference (MD) being negative for Study 2, as the second measure was the Logical Memory and by the MD being positive for Study 3, whereby the second measure administered was the ADAS-Cog. Where possible, this study design implication should be incorporated into future studies in order to remove a clear source of variance and remove the impact of repeating cognitive measures on the results for improved interpretations.

The wider implications of these analyses are that future studies need to place greater importance on the construct of the measure selected to index “amnestic impairment/immediate recall/delayed recall” within the spectrum of AD. Whilst there is a clear relationship between fundamentally similarly

constructed measures, within a homogenised population that relationship is nowhere near strong enough to allow for direct analogous comparisons to be draw. This is fundamentally critical when looking at clinical trial results where these measures vary across nearly every clinical trial and compound in development. It is also an important point to consider when looking at patients in the clinic, with the ecological validity of a measure important to assess when trying to ascertain level of impairment. Clearly a comparison to a normative cohort is key within this scenario but measure selection is also critical to a broad understanding *apriori* of the outcome of the cognitive assessment in question.

Overall, the aim of this chapter was to look at the relationships between measures of memory, commonly utilised to index this domain within AD cohorts. These studies have shown that across cohorts and across similar measures, variability exists and even when looking at subjects with confirmed biomarkers amnestic impairment, cognitive performance on both immediate recall and delayed recall can vary from measure to measure. The convergent validity of memory measures is something that exists but concurrent validity does not. The interchangeable nature of the way measures are used, is at odds with the findings of Study 2 & 3. This has broad implications for the use of cognitive measures within AD as these measures are used interchangeably throughout AD research.

Chapter V – Biomarker Stratification Modelling

Chapter Outline

The first three studies of this thesis have revealed the variances in cognitive measures widely used within AD diagnostics and clinical trials. These studies have also looked at using a novel discrepancy score within AD, and explored how this measure performs in an AD clinical trial population. It has long been a fundamental goal of trialists to predict AD pathology from the clinical presentation, and to bridge the gap between clinical presentations and biological classifications. This is particularly relevant when it comes to patient selection for clinical trials, which has been an area of broad failure over the last two decades. The current chapter will build upon the prior studies by furthering the understanding of the interpretability and criterion validity of these measures, by trying to marry this with the biological phenotype of AD. This is being sought by using additional discrepancy measures and machine learning methods to build sophisticated models for classification of amyloid positivity, the biological cornerstone of AD.

Introduction

While many factors are thought to contribute to the high attrition rate in AD clinical trials, a major contributor is currently argued to be cohort heterogeneity, both in terms of the cognitive and clinical presentation, as well as divergent ATN biomarker profiles. The challenges of assessing treatments of neurodegenerative disorders in clinical trials includes heterogeneity in an individual's drug response, the underlying disease pathology, the length of disease course and the trajectory of the decline, amongst other factors. Only by bringing in biomarker endpoints, specifically amyloid, the cornerstone of an AD biological diagnosis, can we begin to bridge the gap. This chapter aims to achieve this by using more complex classification methodologies, specifically through machine learning, a technique growing quickly in popularity and widely used in other areas of AD research. Being able to predict AD pathology without invasive and expensive techniques has long been a big goal for the field. If this is achieved, this will enable quicker and broader diagnosis of AD, saving costs and also allowing quicker access to any potential disease modifying AD treatments, through quicker clinical trials.

ATN & Cognition

As outlined in detail within **Chapter 1**, ATN is recognised as defining AD as a biological construct, in order to select more homogenous patient populations. The difficult step is marrying the cognitive staging and clinical

presentation into this model. Cognition is fairly well characterised at the latter stages of AD and has been fundamentally defined across mild, moderate and severe AD. However, it is less so at the very earliest stages of the disease. As shown in the prior studies, cognition is one thing that is highly variable across this population, and there are a number of confounding variables that influence an individuals' ability and performance at any given timepoint. Thus, in order to accurately incorporate cognition into ATN, a more accurate staging approach of the clinical and cognitive presentation is needed – one that takes into consideration these prior findings and variables.

Within AD, research has robustly shown that decline first occurs approximately 15-20 years *after* the start of the formation of amyloid plaques in the brain (Jack et al., 2018). This cognitive impairment is argued to be the first clinical manifestation of AD. Cognitively impaired, in the context of AD, is widely agreed to be an impairment on a cognitive measure that is equivalent to or greater than 1-1.5 standard deviations below the appropriate normed population scores (McKhann et al., 2011). However, there is still no agreement on which measure or particular memory domain this should be. As shown with the ATN categorisation method, individuals who fall into some of these categories can in fact have the absence of AD or even concomitant pathologies, this is the same with cognition (see **Chapter 1** for full discussion). Those with a recognised cognitive impairment but without specific amnestic decline, are very much questionable in terms of a diagnosis of "frank AD", "MCI due to AD" or "possible AD dementia". Cohort studies have estimated that non-amnestic/atypical AD occurs in around 5-10% of diagnosed AD cases (Graff-Radford et al., 2021). These cases are often argued to have comorbid pathologies, however currently many of these pathologies are only able to be diagnosed post mortem as in-vivo techniques are currently not as abundant as the pathologies.

Marrying biological classification to the clinical phenotype

Moving into new diagnostic territory with the ATN classification system brings a new set of challenges. For many of these modalities, moving from a clinical classification to a biomarker one tends to involve having a single cut point in order to define a positive or negative grouping. This is the norm for clinical trials in AD. However, having such stringent criteria does help clearly define AD classification but also hinders a more nuanced approach that has a degree of flexibility in clinic. This ATN diagnostic criteria has subsequently led to nuanced model comprising of a number of different combinations of pathology, some related to AD and some which are not. For a biological diagnosis of AD it is a widely held view that amyloid biomarkers represent the earliest evidence of AD neuropathology in vivo. As such, in combination with phosphorylated paired helical filament tau (p-tau) they are seen as categorical determinants for the definition of AD (Montine, et al., 2012;

Hyman et al., 2012, Frisoni et al., 2022). Furthermore, abnormal amyloid being the earliest pathologic change, can be argued to be the defining signature of AD (Jack et al., 2018). This positive amyloid biomarker is critical for the categorisation of AD as without it, any positive biomarker is indicative of non-AD pathologic change and without a positive amyloid AND tau biomarker, a positive (N) biomarker is suggestive of concomitant non-AD pathology. These definitions allow for further subtyping of the biological presentation to allow for a homogeneous look at the complexity of AD pathology.

Nevertheless, this biological construct is somewhat at odds to the clinical phenotype. Approximately 30% of those diagnosed with clinical AD do not have amyloid pathology. On top of which, of those who are amyloid positive not all will decline to AD or progress from MCI to AD dementia. A longitudinal study of the Mayo Clinic database showed that of those who were amyloid positive at baseline, only 30% of them declined after 5 years (Jack et al., 2019). Their population data also showed biologically defined AD is also three times more prevalent at 85 years of age than the clinically defined phenotype. This is argued to suggest amyloid alone is not enough to infer future decline. This is only when tau positivity is found on top of amyloid positivity. However, looking at the longitudinal modelling done on this cohort, at age 70, 30% are amyloid positive with 10% exhibiting MCI or worse, at 80, 55% are amyloid positive with 30% exhibiting MCI or worse. This suggests that whilst amyloid isn't immediately a prerequisite for decline or AD dementia, it does infer a much greater risk of future cognitive decline. This is in line with our current understanding of AD as a disease, with a 20-30% prevalence of disease and with a number of risk and protective factors that significantly interact with an individual's disease trajectory. As with all neurodegenerative diseases, AD is highly complex both clinically and biologically and its aetiology is not currently fully understood.

Further to this, disease and pathological comorbidities are common within the general elderly population (Ferreira, Nordberg and Westman, 2020). This is also common even within clinical trial cohorts aimed at a singular disease or process homogenising the trial population. Nevertheless, trials for diseases such as AD, have clinical populations that still tend to display divergent protein and other biomarker values, arguably indicating that they are at differing stages of the disease. This provides fundamental support of heterogeneity being common even within complex homogenised clinical trial populations. Disentangling the heterogeneity of a neurodegenerative disorder such as AD is critical in order to get the right disease-modifying drugs to the right patients, at the right time. This biological evidence has shown indications of AD heterogeneity on top of the ATN framework referred to as the ATX(N) continuum where additional X represents additional candidate biomarkers such as neuroimmune dysregulation, synaptic dysfunction and blood-brain barrier alterations (Jack et al., 2018; Hampel et

al., 2021). These recent suggestions propose distinct subtypes of AD (Frisoni et al., 2022) with additional evidence of further subtypes yielded from machine learning methodologies on RNA datapoints (Tijms et al., 2020; Neff et al., 2021). Studies of multiple cohorts with inter-disciplinary data (Ferreira et al., 2020) also strongly argue for three distinct drivers of heterogeneity which include risk factors, protective factors and concomitant non-AD pathology, all of which need to be addressed within clinical trial populations to better address the underlying disease process. Further to this, prior phase trials can yield important information that informs a go or no-go decision. As outlined in Wessels et al, prior cognitive data as well as biomarker outputs can fundamentally drive future development (Wessels et al., 2021). However, these two strands of research need amalgamation before any real progress is seen.

Positive steps are being taken though as clinical trial datasets are incredibly rich and highly characterised. And with the constant developments in research, they have the potential to inform future trial designs on these complex classification issues. Stratifying patients with a disease such as AD into subgroups based on these multiple factors will ensure a greater accuracy in predicting a patient's response to a drug in a trial that targets a specific phenotype of the disease. However, as discussed above the clinical presentation doesn't always overlap with the biological one. AD causes dementia but is influenced by many variations in an individual's genetic profile, which can be argued to precipitate disease trajectory and result in subtle pathological variations. This is the fundamental crux of the probabilistic theory of AD based upon PSEN & APP mutations in familial/dominantly inherited AD and APOE genotype in sporadic AD (Frisoni et al., 2022). Those who are not carriers of those specific mutation then infer a broad plethora of risk factors and minor risk alleles and develop AD. It is suggested that these genetics are fundamental factors in driving these divergent disease trajectories. Within the cohort being studied the APOE gene are the only applicable factors and are incorporated into the models being built.

All of this current work leads to the argument that patient selection and stratification, both within clinical trials and broader research, should take these considerations into account to find the patients at the right pathological stage/trajectory of AD. Despite the attraction of simplicity, phenotypic stratification systems are currently not reliable predictors of disease progression or subtype and more complex methods such stratification by multiomic scores are most practical for the generation of homogeneous groups, but are not particularly feasible within clinical trial paradigms or to measure drug response and baseline underlying disease status. This is something that machine learning methods are able to be employed for, as they are able to deal with multiple disease stages,

endpoints for classification purposes, or when looking longitudinally, trajectories of decline.

Machine Learning/Models For Prediction Applications In Clinical Trials & AD

When contemplating trial design and patient selection, patients' stage of disease along with anticipated future decline, cognitive performance and biomarkers endpoints are key factors for success. With trials only lasting 18 months, in diseases that typically span decades, matching the right individuals to the right compound at the right time is fundamental to clinical trial outcomes. These factors can be used to characterise the majority of past, repeated, clinical trial failures. By refining patient selection and homogenising patient's disease pathologies, there is an increased likelihood of an efficacious outcome and subsequent regulatory approval, all of which is achieved through reduction in inaccuracies in patient level responses to the given therapy.

Measures within clinical trials are often under-utilised with specific hypotheses for each of those said measures being poorly designed. Thus, leading to a discordance in findings across each clinical phase of development. Machine Learning (ML) can play a key role in improving these decisions, trial design and boosting the power of trials meaning less individuals need to be recruited. This can be achieved by assisting trail enrichment with appropriate individuals through imaging, biomarkers and cognitive performance (when measured over time) to allow for subject stratifications that classify individuals prior to enrolment (Kohannim et al., 2010; Ithapu et al., 2015; Mathotaarachchi et al., 2017; Ahmad et al., 2018; Calvin et al., 2020; Thall, 2021). Enriching cohorts with individuals who are at the correct stage of AD for each individual compound, can be done by fundamentally categorising based upon biomarkers that relate to homogenous staging and prospective endpoint declines. Many markers now exist for accurate classification of individuals within the ATN criteria for AD, nearly all of which are captured within clinical trials already. However, instead of incorporating long lead in times, as well as additional endpoints and analyses, modelling progression based upon disease specific biomarkers and subsetting the analysis a priori can alleviate these operational constraints. Further to this, as biofluid markers become more accessible and cost effective in larger trials, algorithms that can predict biomarker status become particularly useful in increasing screening success rates and could lead to quicker trials and speedier decisions on compound efficacy. The start of this must begin with the prediction of amyloid, the cornerstone of the biological definition of AD.

Many groups have produced work on ML paradigms within AD cohorts. The EPAD cohort for preclinical AD is a large longitudinal cohort assessing

individuals at risk of developing dementia and as such has a wide range of groups across the ATN staging criteria. This cohort contains individuals at a much earlier disease state than the one being studied within this thesis but these findings are nevertheless important to discuss. There have been encouraging results from studying this cohort and in particular in predicting ATN classification. This was done using Gaussian mixed models and multinomial regression to estimate the likelihood of these ATN groupings. This was achieved through looking at a number of lifestyle, risk factors, cognition and imaging endpoints. Their ANOVA (Chi-squared where applicable) based models showed AUCs of up to 0.89 in distinguishing ATN groupings (Calvin et al., 2020). They approach building models on the basis of basic significant group differences between A+, T+ and N+ groups on a plethora of clinical, demographic and cognitive measures. Once these likely risk factors had been established, they simplified the groupings to a binary distinction of AD pathologic change (A+ ATN groupings) and non-AD pathologic change (All A- groupings). Building their models in an additive way allowed for the strongest possible classifiers to be found and included in the final prediction model. This meant an improvement of AUCs for the initial model variables of 0.82 improving up to 0.89 based upon the key risk factors shown in prior analysis. This additive approach is clearly beneficial for selecting the relevant variables to include in any model. The alternative approach relies on unsupervised learning which is hypothesis generating, which is in direct contrast to the approaches taken within the literature, as well as this chapter, which are hypothesis driven.

Further work from Ingala and colleagues (2020) also used Gaussian mixed models to determine CSF cut points for ATN. They split their cohort by ATN grouping and excluded individuals with non-AD pathology. To note P-tau and t-tau strongly correlated (R^2=0.98) so hippocampal volume was used for N. In cognitively normal individuals p-tau181 drives cognitive dysfunction. However, given recent evidence showing this is intrinsically linked to amyloid and subsequent amyloid removal (Budd-Haeberlain et al., 2021), rather than tangle accumulation, this posits that the pathological process of cognitive decline is still poorly understood. This analysis from Ingala also showed that whilst being cognitively normal, individuals who had greater amyloid burden still had memory and language domains impacted as the first domains of decline within preclinical AD, further supporting the prior findings of these being the key domains of interest in early AD.

MRI imaging is another key piece of information that can be explored prior to more invasive procedures, such as CSF lumbar puncture or PET scans. This imaging modality allows for rich datasets to provide additional anatomical and structural information of the brain both cross sectionally and longitudinally. Given the vast number of endpoints within a single scan, ML methods have been commonly employed to analysis these deeply rich datasets. One common method is to use Support Vector Machines (SVM) to

provide a hyperplane which characterises a large number of variables through a dichotomy. This has been shown to be successfully applied to numerous analysis paradigms across multiple diseases, both for cross-sectional analysis as well as transition predictions and treatment prognosis (Orru et al., 2012). This method of categorisation provides optimal grouping characteristics across multiple, divergent endpoints. An example of this is in practice within AD comes from the EMIF-AD cohort. This is an analysis of individuals across the spectrum of AD as well as normal healthy elderly (ten Kate et al 2018). They looked to predict clinical diagnosis of MCI or AD using clinical, cognitive, demographic variables plus volumetric MRI volumes. AD was confirmed using PET or CSF amyloid positivity assessment within the cohort. Their feature selection was different to that of Calvin et al (2020) whereby apriori they defined the tree-based feature selection strategy using a random forest method (hypothesis generating). With the Gini index was used to measure the relevance of each feature (Cutler et al., 2012). The analysis also then used an additive approach to find the best predictive model, with demographic and cognitive measures model providing an AUC value of 0.64. The best model included demographic, cognitive, APOE status and MRI to give an AUC value of 0.85 for the model for distinguishing amyloid positive and negative individuals.

The papers from ten Kate et al., (2018) and Calvin et al., (2020) will form the basis for the proposed analysis, however in a much larger sample size (prior work; n=863 & n=1010 respectively). The variables of interest will be based on the key cognitive domains of interest and include discrepancy scores. The aim will be to build model of prediction of amyloid status using a hypothesis driven approach. This will be founded on the basis of the cognitive, demographic and clinical variables, with the addition of discrepancy scores based upon work in prior chapters and built in an additive manner to discern the best feature selection possible within this cohort.

Methods

Study Details

This study aims to use a machine learning approach (SVM) to build a predictive model of biomarker grouping (amyloid positive or amyloid negative) based upon scores on clinical and cognitive measures as well as demographic factors. The model will also incorporate findings from prior chapters on discrepancy/change scores to help predict pathological status. The proposed analysis will build three SVM models of prediction, in an additive manner to find the best predictors of amyloid status. For each model the dataset under investigation will be split into two in a 70:30 split as common with all ML work within AD. The first 70% of the cohort will be used to build the model, with the final 30% used to test the classification properties of that model on brand new data. This split will be done from

scratch for each model being built. This allows for the models to be built with the largest number of individuals possible, whilst maintaining a set of unseen data by the model sufficient to test the performance and generalisability of the model on. This will be repeated twice with two sets of cognitive measures alone, with the best performing model then taken forward to a third model whereby it will be added to the full batch of clinical and demographic variables within the cohort. All three models will also use 10-fold cross validation within the training data set to minimise the bias of the classifiers and maximise the generalisability of each model. The first model will comprise of total scores from ISLT (Immediate & Delayed recall), MMSE, CDR and all four tasks from the CBB. These are the commonly used normal cognitive indices which will be tested compared to the second model. The second model will look at using discrepancy score and a memory difference score, on top of total scores from the MMSE and CDR. The MMSE and CDR are included across both models as they are fundamental to the measurement of AD. The third and final model will use the top performing model (1 or 2) in combination with demographic factors (age, gender and years of education) with ApoEe4 carrier status.

Participants

All individuals used for all analyses within this chapter are from the main cohort of screening and baseline assessments, from the clinical trial program of elenbecestat. The cohort comprises individuals who have presented to memory centres and consented for testing for the possibility of taking part in a clinical trial in early AD. Each subject has undergone a fully informed consent process consistent with each country/states IRB and EC guidelines. The screening criteria for inclusion into the study mandated the exclusion of medication affecting cognition, known comorbidities prior to brain imaging and depressive illness within the last five years. Full extensive details of this cohort can be found in **Chapter 2**. Pertinent to these analyses only individuals who underwent amyloid positivity assessments are included here. This is a subset of the full screening cohort as individuals had to meet inclusion and exclusion criteria in a tier (step)wise manner. Thus, those who underwent amyloid positivity assessment, are all cognitively impaired (>-1 standard deviation below the norm), have no comorbidities known to effect cognitive performance, have normal blood tests, ECGs and neurological and physical examinations, as well as having no abnormalities on a screening MRI. Excluding those with any missing data, this resulted in an analytic cohort of 3675 individuals.

Assessments

The measures used in these analyses will be the same as previously discussed in prior chapters. The CBB, ISLT, ADAS-Cog, CDR and MMSE are all discussed in full in **Chapter 2**. These measures will be used in a component-based manner, based upon their corresponding domain features. Further to this, the same discrepancy measure from **Chapter 3** will be computed to be

used within the models, with an additional memory difference score computed too. This memory difference score will be computed by subtracting the delayed recall score from the immediate recall score from the ISLT. This is included on the basis of a within domain comparison of amnestic performance, something that was not possible within the analysis in **Chapter 3**. Overall, for all of these individual screening assessments, there were approximately 59 days (range 12-89) between the initial cognitive and clinical measures being conducted and the final amyloid assessment. Within that time period the other screening procedures were done, many of which were exclusionary (again full details are contained within **Chapter 2**). This timeframe is in line with a minimisation of any practice effects due to repeated testing.

Discrepancy Scores

As broadly discussed in **Chapter 3** discrepancy score adds another dimension to the cognitive profile of individuals. As such these scores will be computed again using the same methodology described in the prior chapter. The fluid composite will be comprised of the 4 tests from the CBB and the crystallised composite from the five language domains/questions from the MMSE. The discrepancy score for each subject will be calculated using the following equation: fluid composite – crystallised composite = discrepancy score. A positive discrepancy score would indicate higher fluid ability and a negative discrepancy score would indicate an impairment in fluid ability compared to expectations based upon the individuals crystallised ability.

Amyloid Classification

Fundamental to this analysis is the classification of individuals for confirmatory amyloid pathology. The binary classification within this analysis is split by levels of amyloid beta present on a PET scan or low levels of the protein in CSF (Roberts et al., 2021). Individuals were given a positive or negative determination on Amyloid PET by an expert radiologist, a team of whom performed a visual read of all images centrally. The radiologist always blinded to cognitive status of the individual. The thresholds for positivity were dependent upon which radiotracer was used for the PET scan. With the label of each tracer defining the criteria for the number of positive regions to claim an individual being amyloid positive on a visual read. All scans were analysed by readers trained on the guidelines established by the manufacturer for each constituent tracer. For Florbetapir (Amyvid™) the regions are; Frontal cortex (excluding midline), medial frontal cortex (including Anterior cingulate), parietal cortex (excluding midline), medial parietal cortex (precuneus and/or posterior cingulate), temporal cortex, and occipital cortex. For Florbetaben (Neuraceq™) the regions are; Lateral temporal, frontal lobes, posterior cingulate/precuneus, parietal lobes. And for Flutemetamol (Vizamyl™) the regions are; Frontal lobes (axial & sagittal views), posterior cingulate and precuneus (sagittal & coronal views), temporal lobes – lateral regions (axial views – coronal views as supportive),

parietal lobes – lateral regions (coronal views – axial views as supportive), striatum (axial views – sagittal views as supportive). For CSF samples taken for amyloid positivity assessment, these were all analysed at a central laboratory with a set cut off level for amyloid beta 42. Two assay platforms were used across the duration of the study for analysis of samples for logistical purposes. These were assays of Aβ(1-42) with a cut off for positivity of <250 pg/mL from Alzbio3 which was run at the ADNI core lab and the Lumipulse™
platform from Fujirebio, using a total tau:Aβ(1-42) ratio greater than 0.37 to indicate positive amyloid status. Both cut offs have been validated, analysed for concordance and broadly used within many cohorts to date (Kaplow et al., 2020).

Machine Learning Approach - Feature Selection & Performance Evaluation

In order to properly validate the model, the full dataset of 3675 individuals will be split into two equal groups. These groups will be selected from the larger sample at random for each model, with ten-fold cross validation being used (in line with the prior work from ten Kate et al., 2018 and Calvin et al., 2020) to prevent overfitting of any of the models. Demographics and cognitive characteristics will be sought on the apriori defined key outcome variables in a hypothesis driven approach. Models will be built in an additive manner. Model 1 will contain z-scores for Immediate & Delayed memory recall, attention, reaction time, working memory & executive function, as well as total scores from the MMSE and CDR. Model 2 will contain variables of discrepancy score, memory difference score as well as the total scores from the MMSE and CDR. And finally Model 3 will contain all demographic measures (Age, Gender, Years of Education) as well as APOEe4 carrier status along with the variables from the best performing model (1 or 2).

The SVM models will each consist of 10-fold cross validation as using an optimum cost function for each individual model. This will be done within the first 70% (training set) of the dataset whereby after splitting in two, that part will be split into 10 separate subsets to train the hyperparameters of the SVM model (cost parameters and kernel function). In order to accurately compare and assess the performance of each model, the receiver operating characteristic (ROC), area under the curve (AUC), specificity, sensitivity, Youden index and accuracy will be computed each of the testing datasets. The SVM modelling anlaysis was run in R using the e1071 package, which allows for the tuning of the model parameters on the training data prior to the overall model being built. The kernel function was always radial, with the cost parameters tuned across six decimal places and the gamma function between minus eight and positive one.

Results

Demographic & Cognitive Comparisons

The features selected for use within the classification models for the whole sample are presented in **Table 5.1.** Across amyloid positive and negative groupings, APOE status, Age, Gender and Education were all significantly different across amyloid status groups. Effect sizes for each variable across groups were also calculated using Cohen's d formula. All measures outside of delayed recall showed small or minimal (d<0.2) differences between groups. The effect size on the delayed recall measure across amyloid groups was the greatest in size.

Table 5.1 Demographic Group Level Characteristics

Characteristic	Whole Group	Amyloid Positive	Amyloid Negative	Group Differences	Cohen's d
N	3675	2056	1619	-	-
Age (years)	70.7 (7.8)	72.1 (7.1)	68.9 (8.2)	*	-
Gender (% Male)	48.5%	48.8%	48.1%	*	-
Education (Years)	13.4 (4.4)	13.7 (4.2)	13.1 (4.6)	*	0.14
APOE e4 Status (% Carrier)	45.9%	64.0%	22.2%	*	-
MMSE	26.6 (1.8)	26.4 (1.8)	26.9 (1.7)	*	0.29
CDR-SB	2.35 (1.0)	2.63 (1.2)	2.24 (1.0)	*	0.35
Immediate Memory (ISLT)	-1.69 (0.8)	-1.81 (0.8)	-1.52 (0.7)	*	0.39
Delayed Memory (ISLT)	-1.78 (0.8)	-1.97 (0.8)	-1.55 (0.7)	*	0.59
Reaction Time (DET)	-1.45 (1.6)	-1.26 (1.5)	-1.69 (1.6)	*	0.28
Attention (IDE)	-1.85 (2.0)	-1.60 (1.9)	-2.17 (2.1)	*	0.28
Working Memory (OBM)	-1.52 (1.7)	-1.57 (1.7)	-1.46 (1.7)	NS	0.07
Executive Function (OCL)	-1.21 (1.0)	-1.36 (1.0)	-1.02 (1.1)	*	0.32
Memory Difference Score	0.10 (0.8)	0.15 (0.9)	0.03 (0.8)	*	0.14
Fluid Composite	-1.51(1.2)	-1.58 (1.2)	-1.45 (1.1)	*	0.11

Crystallised Composite	0.02(1.0)	-0.03 (1.0)	0.05 (0.9)	NS	0.08
Discrepancy Score	-1.52(1.4)	-1.56 (1.5)	-1.50 (1.4)	NS	0.04

*Means and (standard deviations) are displayed for each group. *denotes a significant group level differences using simple t-tests with the exception of Gender & APOE Status for which Pearson's $\chi 2$ test was used ($p<0.05$). Correction for multiple comparisons was also undertaken using the Bonferroni correction method. Values represent raw scores for the MMSE, CDR-SB & demographics. All other measures are z-scores.*

Looking at the groupwise differences between amyloid positive and negative individuals, all cognitive measures apart from discrepancy scores show significant differences between the two groups. Noticeably, amyloid positive individuals had better scores on the simpler tasks (reaction time and attention), whereas for tasks with higher cognitive load (memory, executive function), amyloid positive individuals performed worse than amyloid negative individuals. For the composite scores there were small differences between groups on the fluid and crystallised composites, with a slight numerical difference favouring better scores in the amyloid negative group on the overall discrepancy score (although this difference was not statistically significant) and the memory difference score (which was statistically significant).

Multi-variable Classifier Results

SVM Models were run in an additive, stepwise manner with the model being built on a test data set that comprised of 70% of the cohort, with the model then tested on the remaining 30% of the data. This split was done in a random manner and was repeated separately for each model built. Each model was built first using the training dataset, using 10-fold cross validation to tune the model parameters in order to optimise its performance. The final model was then run on the test data set (the remaining 30% of the cohort in each case) with classification performance measures calculated for each individual model in turn.

The first model used the total z scores from both components of the ISLT (immediate and delayed memory recall), all 4 z-scored components of the CBB (reaction time, attention, working memory and executive function) plus total scores from the MMSE and CDR. This performed fairly poorly in classifying amyloid status within the testing cohort (AUC=0.681). The second model was comprised of the discrepancy score calculated for each subject as well as a memory difference score. These were combined with total scores from the MMSE and CDR. The second model performed even more poorly than the first (AUC=0.627), leading to the selection of the variables from Model 1 to be taken forward to Model 3. This third model combined the

variables from Model 1 with key demographic details for each individual. The ISLT (immediate and delayed memory recall), all 4 z-scored components of the CBB (reaction time, attention, working memory and executive function) plus total scores from the MMSE and CDR, were combined with age, gender and years of education, plus APOEe4 carrier status to try to predict amyloid positivity classification. Model 3 performed moderately in classifying amyloid status in the testing dataset (AUC=0.734).

Table 5.2 gives the full classification properties for each of the three support vector machine models. To note the performance of model 1 showed moderate sensitivity at picking out the true positives as compared to the false negative from the cohort. The model showed poor specificity, giving the model overall poor classification performance for distinguishing amyloid positive individuals from amyloid negative individuals. Model 2 improved slight on these classification properties, this was driven by its ability to show good sensitivity but very poor specificity, showing very limited ability of the model at selecting the true negatives in the cohort, as compared to the false positives. Model 3 improved in all performance metrics from those of Model 2. Out of the 3 models, Model 3 was able to correctly classify the most individuals as amyloid positive and amyloid negative (73.4%) within the testing cohort.

Table 5.2 Optimised Performance Characteristics of Each SVM Model/Classifier

Model	Sensitivity	Specificity	Youden Index (J)	PPV	NPV	AUC Value	Accuracy
Model 1	75.4	49.3	0.247	65.5	61.1	0.681	0.224
Model 2	83.3	30.9	0.142	60.6	59.3	0.627	0.728
Model 3	89.0	33.0	0.220	64.2	69.1	0.734	0.684

Model averaged across 10-fold cross-validation. Youden's J statistic employed for the Youden Index. NPV=negative predictive value, PPV=positive predictive value, AUC=area under the curve.

Performance of SVM Classifiers for Amyloid Positivity Assessment

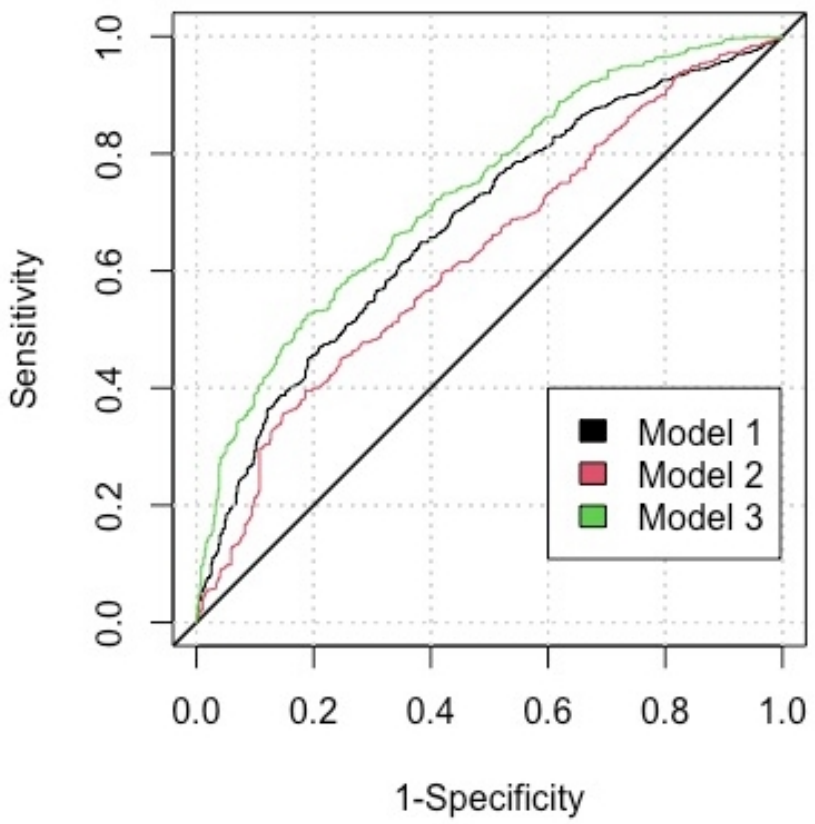


Figure 5.1
Receiver operating characteristic (ROC) curves of support vector machine classifiers to predict amyloid pathology. Model characteristics are displayed in Table 3.2. *Model 1 (AUC=0.681): Z-scores for Immediate & Delayed memory recall attention, reaction time, working memory & executive function, as well as total scores from the MMSE and CDR. Model 2 (AUC=0.627): Discrepancy score Memory difference score as well as the total scores from the MMSE and CDR. Model 3 (AUC=0.734): all demographic measures (Age, Gender, Years of Education), APOEe4 carrier status + Model 1.*

Discussion

This study aimed to build upon the prior studies by assessing whether it is possible to predict AD pathology using a new discrepancy measure and/or commonly used measures of cognition. This analysis has shown that within this population total scores from cognitive measures show the strongest classification properties compared to that of discrepancy scores. Overall, all models showed fairly moderate to poor AUC values. This demonstrates the incongruency between cognition and amyloid dichotomy within AD. However, despite this poor performance all three models showed good sensitivity in classifying true positive cases (sensitivity), but in nearly all three cases, the models performed poorly at selecting the true negatives (specificity). To note, AUC values were used as the primary performance measures as the

accuracy measures can be argued to be unduly influenced by a 40:60 proportion of amyloid negative to positive cases.

The poor classification performance for these models could be in part due to the unknown levels of each of the ATN biomarkers. As previously seen the three ATN biomarkers each have a highly variable relationship with measures of cognition. Although this analysis dichotomised cohort labels for A, using set thresholds for amyloid positivity, individuals were able to be at any level (high or low) either side of the boundary. In such a large sample, it is highly likely there is a high variability in amyloid load across the cohort. Although the model build (in each case) was done in a randomised fashion, as the continuous data on amyloid load was not available, it is plausible that this may have given rise to broad cognitive variability within amyloid positive and negative groups. On top of this, recent research (as discussed above and in prior chapters), has shown a stronger association between levels of tau (T) and cognition. When the data were collected, no tau imaging or plasma biomarkers were validated or available for levels of phospho-tau. Therefore, this data was not available within this cohort, but it is highly plausible this will have a stronger association with cognition than amyloid. Tau plasma data is currently being sought for analysis within this cohort and is a key future area for research. The same issues also apply to markers of neurodegeneration (N), as these levels will likely vary broadly within this cohort too, and as such measures of this carry nearly as much importance as markers for tau as a future avenue for research.

Looking at the prior work of ten Kate et al., (2018) & Calvin et al., (2020), this work shows that without markers of MRI (ten Kate) or other biomarkers for AD (Calvin), the classification utility of cognitive and demographic variables is somewhat limited, even when used with these cognitive and clinical variables being indexed in a novel manner (discrepancy scores). Nevertheless, there is still merit in approaching biomarker classification using these machine learning approaches. As has been shown in the prior work these analyses were founded upon, the poor performance here was not due to the ML methodologies employed but in the utility of the cognitive measures in predicting the absence of presence of biology. It is clear there are limitations in the interpretability of cognitive performance on these measures, as from these models, it is clear there is a weak relationship between amyloid classification and cognitive performance.

There is clearly still some way to go for a better confluence between the clinical and biomarker classification perspectives. This analysis goes to show that the clinical and cognitive measures do not aid diagnosis when looking for the cornerstone of biomarker classification for AD, the presence of amyloid. The clinical presentation of AD is still broadly varied in primary care, with doctors starting by looking for impairments in cognition, primarily memory and diagnosis should be approached in a different manner. Primary

care does not yet have biomarkers available, plus the clinical presentation of AD often occurs even without a subsequent finding of amyloid. This cohort is a primary example of this, with ~40% of the studied group being amyloid negative yet presenting with a clinical diagnosis of AD or mild cognitive impairment due to AD. Amalgamating biomarker classifications which are fundamental to clinical trials, to that of clinical presentations of AD which are highly varied, still needs much work and further study. This could primarily be done by assessing screening cohorts, such as the full complement of individuals within this one (n=9758), in greater detail regardless of their cognitive performance, by assessing them in full (including biomarker and imaging workup). However, this would be incredibly costly and operationally complex. Current initiatives in this vein have begun studying preclinical AD individuals and biobank cohort studies, which will hope to yield some further findings to help bridge these classifications. However, as with any cohort studies implementing new biomarkers into existing programs often lags behind that of clinical trials. As such, findings from cohort studies take longer to implement into both clinical practice and clinical trial design further hampering clinical diagnosis.

This research shows for trial designs, given the dearth of measures used, it is not possible to accurately capture a measure of crystallised ability/intelligence or accurate language ability from the MMSE, CDR or other basic cognitive measures within this population. For further research to progress understanding of discrepancy scores, a broader range of measures is required, with one/s that accurately index verbal acuity key to this endeavour. However, given the reluctance of the field and regulators to shift away from the CDR, MMSE and ADAS-Cog, this appears unlikely for now.

As well as the SVM models, initial analysis looked at group differences on the cognitive measures between amyloid positive and amyloid negative individuals. This showed a number of interesting points worthy of discussion. The 0.4 raw score difference (d=0.35) between groups on the CDR-SB is notable (as this is the primary outcome measure used in clinical trials) and also indicates the likely greater AD progression of a number of the amyloid positive group, potentially indicating higher tau load (T) and/or neurodegeneration (N). This poorer performance by individuals who are amyloid positive is also mirrored on the key memory measures, with a slight bigger difference on the delayed memory recall, which is in line with current understanding of AD related impairment. This worse performance from the individuals who are amyloid positive is replicated over all cognitive measures apart from those of reaction time and attention, which amyloid positive individuals perform better than amyloid negative individuals on. The simplistic nature of these two tasks may allow for any deficits to be masked in task performance, however, a comparative improvement compared to amyloid negative individuals is intriguing and would be aided by the addition of vMRI measures in the hope of better explaining this finding. This could

shed light on the nature of the pathology behind the cognitive impairment being exhibited by those individuals who do not have significant amyloid burden.

Furthermore, the lack of difference between groups on the discrepancy score is indicative of the lack of variability in the measure. This is fundamentally due to it being derived for this population using the MMSE and having very little range or variability. This goes back to the flawed nature of the MMSE in indexing verbal acuity and language performance. This was also seen in greater detail within the analysis in **Chapter 3**. However, this prior analysis was in the screening population the much larger n (9758:3675) which allowed for a broader range of scores. Conversely with this analysis the cohort had much greater clinical and cognitive homogeneity and as such restricted this measure of crystallised intelligence and discrepancy score even further.

The effect size calculations also point toward the largest difference being seen within delayed memory recall, this is in line with the understanding of the prior literature reviews (Mortamais et al., 2017; Baker et al., 2017). However, even with the moderate (d=0.59) group level difference between amyloid positive and amyloid negative individuals on delayed memory recall classification based on this alone is insufficient to be implemented in any meaningful way. This was a rationale for the alternative approach taken in this analysis, of using discrepancy scores to approach cognition in a different way. The results show this alternative approach is not beneficial within this setting.

Overall, there are clear statistically significant differences between amyloid positive and negative individuals when looking at indices of cognition. However, these group differences do not translate to a dichotomisation of groups based upon biomarker classifications, even when implementing sophisticated machine learning models such as SVM. This suggests indexing cognition alone, even across multiple domains, is not sufficient to predict amyloid pathology. Furthermore, a discrepancy score comprised of standard clinical trial cognitive measures, is not able to cross-sectionally indicate this AD related pathology above that of the standard measures themselves. Future research needs to broaden the scope of the characterisation of ATN within the cohort studied and widen the scope of measures used in clinical trials to better capture verbal acuity. Nevertheless, the SVM classification method used here should be utilised in future by looking at this these paradigms of cognitive impairment and biomarker groupings, longitudinally, across the ATN spectrum, with the hope to see improvements in classification performances.

Chapter VI – Conclusions

Chapter Outline

The overarching aim of this thesis was to explore the differences between cognitive measures used within AD clinical trials and for clinical diagnosis of early AD. As well as this, it also explored these measures in a novel way by utilising a novel cognitive discrepancy score. This was done with the aim of improving clinical trial design, which would result in saving costs, time and patient burden as well as improve the chances for a successful trial outcome. This was also undertaken to bridge the underlying pathology of AD with the clinical presentation through a better understanding of the cognitive measures used to define this clinical presentation. This aim of this chapter is to conclude and discuss the merits and drawbacks of the research conducted. This chapter will draw overall conclusions, how this research adds new information to the field, how these analyses were restricted and potential future directions for this avenue of cognitive research.

Research in context

Data is always gathered in a highly controlled way within clinical trials. The funding levels for clinical trials allow for studies to reach more individuals than in conventional research projects. This gives broader context and scope to any findings and is normally cross-continental. Half the sample is normally designated to receive a new molecular entity and the other half a form of placebo. Whilst the placebo group is a prime candidate for broader study, data is always held in house within the very pharmaceutical companies who have heavily invested in this data. Even so, initiatives such a CPAD (Critical Path for Alzheimer's Disease), pool placebo groups from these multiple clinical trials and are accessible to researchers. However, projects require company sign off and current cohorts often take a minimum of 5-10 years before they become accessible to the wider research community. This results in a lag between any research exercises and current theoretical understanding, which often becomes obsolete in this timeframe. The main cohort studied within this thesis was taken from the two largest clinical trials undertaken to date within AD. This broad applicability is one key reason why this cohort study is important and gives important findings to the field.

Whilst this analysis benefited from broad data collection, it is also important to highlight that this is also a large potential source of heterogeneity. By having 644 clinical trial sites this inherently induces variability in the data. There was however, a great effort to minimise this to every extent possible, with standardised cognitive, clinical and imaging protocols. Study staff (including the author) travelled and gave regional specifics talks on suitable individual profiles to screen for this study. The rigorous nature of the clinical

trials also necessitated robust monitoring of any deviations from the set administration and scoring criteria.

Nevertheless, when dealing with multiple languages, dialects and cultures, variance is an unavoidable part of research. In this regard the cognitive tests across both cohorts warrant discussion. The translations for the WMS, ADAS-Cog and MMSE stimuli are all direct translations. This meant words were translated verbatim and no cultural sensitivity was applied to the stimuli. This was not the case with the alternate story recall, which inherently induces variability in the relationship between the two measure (full discussion can be found in **Chapter 4**). This lack of veracity in the translations for the cognitive stimuli can be argued to be one cause of the suboptimal variance seen in the relationship between the cognitive measures. This is something that is also applicable in a broader clinical context when conducting studies in multiple countries. Future studies should also look at accounting for this and exploring the invariance in greater detail. As shown in **Chapter 3** there is broad regional variation across the battery of cognitive measures used for screening.

On top of this, whilst clinical trials themselves typically provide homogenous broad data, there has been a near complete failure of AD clinical trials over the past 2 decades. This has meant a complete absence of any disease modifying treatment, which leaves AD as the only disease in the top 10 of mortality rates worldwide, without a treatment to prevent or slow its onset. All of which means, AD remains one of the most costly and prevalent diseases afflicting the world today. Importantly, the failure of clinical trials has been predominantly down to the absence of efficacy rather than safety concerns. As such, the endpoints used need further scrutiny to ensure they are measuring the progression of the disease accurately, as they have not changed for nearly three decades. These failures also point towards high variability of the endpoints and cohort heterogeneity. Alleviating these measurement issues combined with a better theoretical understanding of the lack of confluence between biomarker and clinical phenotypes of AD, will go a long way to further our understanding of AD. And can only help give new compounds a better chance for success meaning treatments to the right individuals at the right time at a greater pace.

This thesis goes some way to explaining some of this variability. For instance, there are different nuances between cognitive measures used for inclusion into AD clinical trials and diagnosis of AD. Overall, the findings from two distinct cohorts, comparing two distinct memory paradigms (story recall and verbal list learning) showed moderate correlations and with moderate levels of agreement between these two individually studied pairs of measures. Furthermore, the strength of the relationships between these measures all are imperfect and far from concordant enough, to allow for direct comparisons of “amnestic impairment” within populations across the

spectrum of AD. The implication of this finding is that standardisation across companies and research groups is badly needed. By having a broad range of screening scales this adds to already variable scores of cognition, when measuring cross-sectionally, which is nearly exclusively the case when testing for early AD.

On top of this with amyloid being the cornerstone of a biological diagnosis of AD, no matter which stage, predicting this is key to the utility of aligning endpoints. Alas, even with sophisticated classification approaches such as support vector machines, no individual measure can accurately predict amyloid status or load. This has been seen in prior work across screening and clinical trial cohorts. The work of Calvin and colleagues (2020) as well as ten Kate et al (2018) had a broader range of variables but were still only able to achieve AUCs of 0.84. Using only one invasive biomarker endpoint this research managed to reach AUCs of 0.73. Even when looking at novel index (discrepancy) measures of cognition this was not able to be improved upon. This shows that even in a highly homogenised cohort cognitive performance is highly varied within AD and struggles to marry the clinical presentation with the biological one within AD. As such, this poses the question are these endpoints really fit for purpose, as the underlying presence of amyloid within this classification of the disease, conclusively does not fit with the clinical presentation.

The future directions from this piece of work would be to have agreement from overarching clinical and regulatory bodies on new endpoints that mirror disease progression more closely and make these stalwart measures for research into AD. The current ones are outdated, highly variable in some cases and do nothing to measure underlying cognitive domains of key interest, specifically within early AD. These new measures should replace the MMSE, CDR and ADAS-Cog giving more in depth, psychometrically valid information on cognitive performance in those with AD. These measures should also be freely available to trialists and researchers alike. Furthermore, to improve our understanding of the disease we need to have broader cohort studies that follow individuals who present at memory clinics that are not impaired enough for trials or studies, only then will we have a broader understanding of cognition in AD to allow us to find earlier signs of the disease.

Overall findings from this research

Not all measures are created equal

Undertaking this research, a clear message came through from the prior literature that many common neuropsychological tests are used interchangeably in AD research. However, whilst most share underlying construct validity, measurement of cognition is messy and subtle variations

in test paradigms, result in greater variation in individual performance. Whilst much effort and research is undertaken on the initial validation studies for most (if not all) measures, understanding how they perform within current understanding of a single disease (such as AD) is broadly overlooked. As a result of this, the variability in cognitive endpoints have the possibility to mask or increase differences within study results. Results from this thesis speak to this across two divergent cohorts. Firstly, within a screening paradigm alternate forms of the same measure, studied in the same individuals presenting to a memory clinic, showed only moderate correlations (r=0.64) and poor agreement on Bland and Altman (1984) comparison statistics. This finding was broadly repeated in a second much larger cohort of amyloid positive individuals diagnosed with AD. Two different measures of verbal list learning (the international shopping list and ADAS-Cog word recall) were compared within this highly homogenised population. The same analysis methods were employed and similar results were found. Agreement between the two measures was moderate and correlations again moderate, if not slightly poorer (r=0.55). Whilst the verbal list learning measures have slightly different ecological validity, the overall findings remain concordant. Further to this, from a statistical perspective measuring the same domain twice can improve reliability of the measurement, however if the same tool is used to measure the same construct this could potentially introduce problems with multicollinearity in the analysis. However, on this occasion different measures were used to provide indices of the same underlying domain. Psychometrics such as these found in Study 2 & 3, are widely deemed to be poor for repeated measurement and it is clear from this they should not be used interchangeably, but currently are. A harmonisation approach is badly needed. This is especially pertinent in a disease such as AD, whereby early detection is key and the first notable sign of the disease is amnestic impairment.

Harmonisation of measures would help drug development & patient selection

As discussed above one way to help reduce this variability in cognitive endpoints would be to have agreement on the gold standard measures used within AD research. Harmonisation currently exists to a degree, in that the MMSE, CDR and ADAS-Cog have to be used for clinical trials in AD. This was also the founding principle of the ADCS initiative in mild to severe AD (Rosen et al., 1984). So, this idea is not something that doesn’t have precedent. And as much as the ADCS initiative followed the best methodology to find the best outcome, the field has moved on since then. The argument here based upon the research undertaken in Chapter 4, is that agreement should be sought over the use of a measure to use for detection of amnestic impairment due to AD.

As shown in Chapter 4, even within fundamentally similar paradigms both verbal list learning and paragraph story recall, have broad variability of performance within the same subjects. This is the case even when controlling for as many environmental factors as possible within clinical trial settings (ie. time of day, rater change, testing surroundings). The counter point for this variation is that this in itself may actually reflect the disease course of AD. However, research into preclinical AD shows that in frequent testing paradigms (ie. multiple short assessments within a number of weeks), it is actually the absence of improvement (practice effects) that is indicative of AD pathology (Jutten et al., 2020; Duff et al., 2014; 2017; Ihara et al., 2018; Galvin et al., 2005; Hassenstab et al., 2015; Oltra-Cucarella et al., 2018). This broad variability within both the heterogenous screening cohort and the homogenous trial cohort shows that no matter what the setting these measures should not be used and interpreted in an interchangeable manner. There should be a far-reaching harmonisation effort undertaken, whereby new or existing measures are selected for use and are made broadly available to researchers and clinicians alike. Only this, from a credible research or regulatory body will allow for this cognitive variability to be minimised through measure alignment.

This measure would be suggested to be short and avoid the need for long batches of cognitive testing, focus on amnestic memory recall (both delayed and immediate components), capture real world function, premorbid ability better (ie NART, W-TAR) and mirror disease progression of AD over time. It should also have broad psychometric validity across all the key validity domains detailed in **Chapter 1.** This measure should also have a broad range of scoring, as not to suffer from the same ceiling and floor effects within early AD. This consensus needs to be drawn from an evidence-based roundtable of leading experts putting together the latest biomarker and cognitive work from the field, in order to find and/or develop from scratch the best cognitive measure/s for AD.

Current measures do not reflect amyloid pathology very well

Amalgamating biomarker classifications, which are fundamental to clinical trials, to that of clinical presentations of AD, which are highly varied, still needs much work and further study. One way this could be achieved is by assessing screening cohorts in greater detail regardless of their cognitive performance. The drawback to this is that it would be incredibly costly and operationally challenging. Current initiatives have begun by studying preclinical AD individuals and biobank cohort studies, regardless of cognitive performance, which will hope to yield some further understanding of this relationship, to help bringing these classifications together. However, as with any cohort studies implementing new biomarkers into existing programs often lags behind that of clinical trials. As such, findings from cohort studies take longer to implement into both clinical practice and clinical trial design.

Another critical aspect of any harmonisation effort, would be to look at the specific relationship between cognitive endpoints and amyloid pathology. As wide-ranging reviews from Mortamais and colleagues (2017) and Baker et al., (2017) both showed, amyloid deposition has very little relation to cognitive performance or decline (Cohen's d ranging from 0.24-0.3). However, Study 4 showed an effect size of d=0.59 for delayed recall in comparing the amyloid positive and amyloid negative groups. Whilst this is a strong finding, caution should be urged as this is not directly comparable to the prior work given the homogeneity of the cohort studied within Study 4. The prior lack of concordance between pathology and clinical manifestation is argued to be primarily due to amyloid deposition building up over the 15-20 years prior to the onset of symptoms and cognitive impairment. Nevertheless, some studies within those prior reviews and since, have shown that in relation to amyloid load and cognition, cross sectionally cognitive impairment is more widespread than a single domain. This is likely to be due to heterogenous biomarker profiles. However, the irrefutable evidence longitudinally is for significant early decline in episodic memory in relation to amyloid load. Something Study 4 also strongly reiterates. What has also been shown is the standard cognitive measures such as the MMSE and ADAS-Cog have little relationship with amyloid load and tracking the disease course longitudinally (Baker et al., 2017). This goes on to further support the case for new measures more akin to measuring the subtilties of amyloid deposition. Something this thesis tried to achieve with the novel discrepancy score using existing measures but was unable to do.

Here it is also important to further discuss the lack of confluence between the biological phenotype of AD and the clinical presentation. This is further borne out in the results from Study 4 showing the lack of utility of the clinical and cognitive measures to predict amyloid positivity. This showed commonly used measures have little to no bearing on the amyloid status of individuals. Despite the attraction of simplicity, phenotypic stratification systems such as this are not reliable predictors of disease progression or subtype. What is needed is to design trials aimed at those expected to respond can only be beneficial for all engaged in this area. Something that phenotypic stratification can help with but further understanding needed. Specifically, across the AD continuum the progressive pathology and differential rates of decline in ATN biomarkers (Dubois et al., 2021) argues for the necessity of a more nuanced approach to clinical trial design across all stages of drug development. To further this confluence, wider adoption of the probabilistic model of AD is suggested (Frisoni et al., 2022). With sporadic AD split by APOE genotype, non APOE-E4 sporadic AD is driven by varied genetic and environmental factors. As such this subgroup is thought to be the source of a lot of variance in trial and research data sets. As shown within the analysis within Study 3 & 4, APOE infers clear and concrete influence on cognition and whilst all studies factor this into any analyses, broader and more

overarching diagnostic grouping, using more biomarkers from the ATN classification system, would aid the marriage of biology and clinical presentation.

Predicting pathology from cognition requires greater insight into ATN, something now possible with the advent of blood-based biomarkers

Clearly the very nature of AD pathology is highly complex and varied. This is highlighted by the fact the field is still unsure of the underlying aetiology of the disease and its initial inception point. There is therefore a clear need for broader biomarkers of pathology that influence the disease course and by proxy cognitive and functional deficits. Whilst this is something that was unavailable for either cohort under investigation within this thesis, current efforts are beginning to yield reliable blood-based biomarkers. These are still having broad validation efforts undertaken; however, it is clear these new assays will provide clear accessible biomarkers for future research efforts across all avenues of AD research. This is a cheap and easy was to measure AD pathologies, including those across the ATN spectrum and is at odds with the current invasive and costly CSF & PET procedures, which have been available to the field for the past decade. Research has already shown that different analytes of phosphorylated tau to be strong indicators of both future decline and amyloid pathology (Mila-Aloma et al., 2022), with more findings likely to be brought to the fore in the coming years due to this proliferation.

Discrepancy scores are (likely) not useful at indicating presence or absence of amyloid but do show something different at screening

At the start of this project one core aim was to better understand a novel measure of discrepancy between cognitive domains and how this fits into our current understanding of AD. Prior work from McDonough & Popp (2020) and Jacobson et al (2002), had previously shown comparing cognitive domains to compute a discrepancy score was an early indicator of cognitive impairment in early AD. However, what was missing from both research efforts was a link to the pathology of AD. This was something that was addressed within this thesis. Looking at the screening cohort of nearly 10,000 individuals, the results showed that a discrepancy measure clearly measures a different aspect of cognition to that of immediate and delayed memory (the common indicators of early AD). The larger analysis from Study 1 also showed the relationship was weak but changed over the course of the progression of the disease when measured cross-sectionally, as staged by CDR score. This change with disease progression mirrors that of amnestic impairment and is cause for further research in other cohorts with more biomarker measurements available for analysis.

This work went on to look at the potential predictive value in a discrepancy score in relation to amyloid positivity. The results here were a near failure of the model to predict groupings in over 3,000 individuals. This conclusively shows a discrepancy score (and other stalwart cognitive measures) cannot accurately predict amyloid positivity, even when grouped with a combination of measures and known demographic risk factors, such as APOE. At the inception of this thesis the holy grail was being able to predict amyloid positivity from cognition, alas, both from this research and others, as a field, we are still no closer to this. The poor classification performance using a discrepancy score within this research can be argued to be down to how it was comprised. The hypothesis was to test if it was possible to find a discrepancy measure accurate enough within existing cognitive measures that are used within every clinical trial. As such, the MMSE was used, as this was the measure with components indexing language and verbal ability.

The MMSE does not accurately index verbal acuity in early AD

As part of computing a discrepancy score within this cohort a measure of crystallised intelligence comprised a fundamental part of this new measure. This was built from the 5 verbal measures that are part of the MMSE (Folstein et al., 1975); naming, repetition, comprehension, instruction, reading and had a total score of 8. As there were no other measures of verbal performance within the studied cohort, no comparisons could be drawn or alternative computed. Given that this sub-score from the MMSE produced a max score of 8 this severely limited the breadth of scoring on this index and as such hampered the comparisons inherent within the discrepancy score itself. For comparison the NART (Blair & Spreen, 1989) measure has a total item score of 61. The basic nature of these 5 component scores also allows for learning of these tasks if repeated more than once, as the MMSE has no alternate form. Whilst the MMSE was not designed with this use in mind, it was designed to measure language performance, something that it does in a far too rudimentary way. This shows that subjects, even those who are badly impaired, perform at ceiling in the majority of cases. As shown in the results from Study 1, those who are at the earlier stage of the disease have no trouble with the basic commands and it is therefore highly unsuitable to measure verbal acuity in those who do not have mild to severe dementia.

These ceiling effects are also broadly seen within the MMSE, as well as within the language domains. The psychometric limitations of the MMSE such as learning effects and large ceiling and floor scoring have been shown to vastly limit their diagnostic accuracy in AD (Tombaugh et al., 1992; Sperling et al., 2012; Mitchell, 2009; Spencer et al., 2013). Both prior and current work argues against continued use of the MMSE, albeit brief and inexpensive, for use in AD trials. As this is argued to be a contributing factor to the inaccuracy of efficacy measurement within these trials. Furthermore, on top of this other negative studies may have occurred not because of a lack of

efficacy of the compounds, but due inappropriate subject populations selected based upon MMSE scores.

The Overall Psychometric Validity Findings

As discussed at the start of this work, it is important to look at the different psychometric validity aspects of the cognitive measures used within these studies from this thesis. **Chapter 3** began by examining the cross-cultural validity of the ISLT, MMSE, ADAS-Cog and CDR. This showed different cognitive profiles in the screening populations across regions on nearly all measures. Those individuals in western countries (North America, Europe/South Africa) tended to exhibit less objective memory impairment on the ISLT. Whereas the CDR and MMSE showed significantly less impairment in eastern regions of the world (Asia Pacific and Japan) but the ISLT showed greater impairment. This is suggested to be due to cultural differences in the sharing of subject impairments to clinicians and raters for the CDR which is more objective than the ISLT.

Looking at the validity of the novel discrepancy score, Study 1 looked at the criterion and convergent validity of this as a measure in AD. Its convergent validity (as compared to amnestic measures) was good when looking across the course of AD and it was clear this reflects good criterion validity as a measure within AD. Where this measure falls down is on its construct validity. Here it was designed to measure language/crystallised ability, however the way this novel measure was constructed it is restricted in this regard due to the MMSE. Future work should look at this paradigm with a different measure of verbal acuity as discussed above.

Chapter 4 sought to explore the concurrent validity of amnestic memory measures. Study 2 & 3 showed clearly these measures lack concurrent validity and whilst they have convergent validity, it is moderate to poor based upon the Bland and Altman analysis. The scores from these two pairs of measures do at least have strong construct validity, however greater nuance in early AD could be added to them, through broader error scores to capture deficits in learning.

Finally, Study 4 looked to find the interpretability of these measures through aligning them with individuals diagnosed with biological AD. The interpretability of these measures still has some way to go before they accurately mirror the everyday function performance of someone with biological AD. Nevertheless, the criterion validity of the ISLT was demonstrated by its confluence with the amyloid positive group (d=0.59) showing that this measure is the most relevant to biological AD. This in line with prior work on amnestic impairment being fundamental to AD related cognitive impairment (Mortamais et al., 2017; Baker et al., 2017).

Restrictions on these analyses

After discussing the interpretations of the findings from this thesis it is important to also discuss in detail some of the drawbacks of the analyses undertaken. These analyses were undertaken on one large clinical trial cohort, of which two smaller subgroups were also studied and a smaller memory clinic sample. The restrictions of these analyses are primarily founded in the data collection process and cohort compositions.

Whilst a large part of these analyses focused on implementing findings for clinical trials, this resulted in an extremely homogenised cohort. This isn't necessarily a bad thing for extrapolating findings to AD clinical trials. However, the cohort analysed for Study 4 contained subjects who had been through five stages of screening procedures and were unlike a normal research population. This is because many comorbidities had been excluded over and above a normal research setting (**Chapter 2** contains full details of all screening procedures) and as such can limit the generalisation of any results outside of a clinical trial setting. The main crux of this issue was that all individuals were already impaired with less comparative cognitive variability post screening compared to a general population, or one that would present at a memory clinic. However, this was why a second cohort was sought and studied, from a memory clinic to broader the applicability of these results. Nevertheless, given the results from this large homogenous cohort were aimed at aiding clinical trial design, this does not provide rationale for the negative findings in Study 4. However, one factor that did prevent wider generalisation of these findings was the lack of biomarker data available within either cohort. The larger screening cohort only contained data points on amyloid. Ideally, volumetric MRI, blood or CSF markers of p-tau, t-tau and NFL (and other across the ATN spectrum) would have aided broader analysis of the cognitive endpoints and their confluence between clinical biological phenotypes of AD.

Further to this, the other smaller screening cohort was not followed further than initial investigations of memory. Results of memory screening can only be more broadly interpreted if a diagnosis and further investigations are undertaken. This should be more broadly done to better understand any cognitive impairment an individual is exhibiting. Due to this fact, those that present but don't show significant impairment will not be followed up. This therefore can negatively bias a sample, although this was not the case for this smaller cross-sectional cohort analysis. However, in order to properly understand cognitive performance, both those with and without cognitive impairment should be studied in greater detail.

Future directions

Looking towards new avenues to further our understanding of these concepts, future work should keenly focus on the relationship between these cognitive measures and a broader spectrum of AD biomarkers across the ATN framework. As discussed, predicting amyloid dichotomy from cognitive measures carries little predictive accuracy. However, this was on the basis of a single cut point for amyloid. As current research shows measures of tau and neurodegeneration more closely mirror the declines seen in cognition (Jack et al., 2018; Dubios et al., 2021; Frisoni et al., 2022). As such, further research should explore these predictive models (using SVM) within the context of the full range of ATN groupings, not just A+ and A-.

The analyses within this thesis were in the main, cross-sectional and as such lack any longitudinal measurement akin to tracking AD progression, something which occurs over years and decades. With this in mind, further research and data collection should focus on the longitudinal and repeated measurement of discrepancy scores. This measure should comprise a better language measure as discussed, but should focus on broad fluid cognitive domains akin to this research. How this novel measure behaves over time will help uncover its utility as a screening measure, for which as no current longitudinal study has been undertaken, it is still relatively unknown. Including a quick and robust measure such as the NART in any future study or trial would be highly beneficial in this endeavour.

As discussed within the restrictions of these analyses using the smaller screening cohort, tracking all individuals regardless of cognitive impairment, is something that would further aid our understanding of these common cognitive measures. With this in mind an avenue for further analysis would be to look at a similar discrepancy score at screening with a full cohort of individuals (those with and without cognitive impairment) who are followed for over time along with biomarker assessments. This will provide greater details of the longitudinal performance of a discrepancy score in a screening population and how this relates to AD pathophysiology.

What is clear from both prior work and the studies within this thesis, is that AD is highly heterogeneous. Commonly studies are only run for a couple of years at most. This is problematic as this disease lasting decades and from onset of symptoms to fatality can be over 10 years. As such, the factors that precede a precipitous decline in cognition are still poorly understood. One further question to answer following on from these analyses is that, can a discrepancy score be indicative of decline on CDR at 1 year, 18 months or 2 years. It is difficult to find markers of cognition that can differentiate decline across these short time period, so using a novel measure such as this, with AD biomarker stratification, has the potential to help answer this question.

Looking to the future it is important to speak to the shifting of the broader landscape of AD research. Blood biomarkers were not available or validated

when thesis started, during data collection or upon its conclusion. However, they have been broadly validated in research and clinical trial settings over the last 2 years. There is now momentum for their uptake into primary care. This will be a seismic shift in how AD is diagnosed and detected. This calls into question the utility of cognition in screening and in particular how useful using cognition to predict pathology. These blood tests will be quick, cheap and widely available to all relatively soon. However, the clinical presentation of AD is intrinsically linked to cognition in a way the biology is yet to be. Therefore, there is still much benefit to cognitive measures being used at the point of care. And critically they allow a fundamental understanding of clinical benefit due to being inherently linked to the everyday function of an individual.

Final Summary

At the start of this thesis the goal was to better understand the nuances between cognitive measures used for inclusion into AD clinical trials and diagnosis of MCI. The aim was to also to utilise these existing measures in a new way, by using a discrepancy score between cognitive domains. This was sought in order to save cost, time and patient burden, with the ultimate aim to better clinical trial design and predict biomarker status of the underlying pathology of AD.

At the top of this thesis, the studies conducted also aimed to provide important new insights into some fundamental questions – all of which have been answered from the results of the studies;

- Why do cognitive measures vary so much within clinical trials and cohort studies?
    - Study 2 & 3 suggest this is at least in part due to the construct validity of these measures. Study 1 also shows that variables such as APOE also inducing broader variation in these measures within AD. Leading to the argument that the classification AD needs further refinement in line with the probabilistic hypothesis of AD (Frisoni et al., 2022).
- Why have all trials to date failed to meet their cognitive endpoints?

    - The lack of confluence between cognitive measures and the biological classifications of AD, as shown in Study 4, can be argued to significantly contribute to this failure.
- How can you accurately diagnose Alzheimer's Disease with cognitive tests?
    - It is clear that the clinical presentation of typical AD must be diagnosed with memory measures, but as discuss above is something that needs harmonisation.
- Can you use existing (and commonly used) cognitive measures in a new way to better understand cognitive performance over the course of the disease
    - Yes, it is clear discrepancy scores for individuals capture something different to that of amnestic memory, which warrants further study. However, it is also clear they are not able to improve the marriage between pathology and the clinical presentation of AD whilst using the language components from the MMSE.

At the end of this thesis the research undertaken has shown plausible reasons for the variability in cognitive measures, that those measures used within cognitive trials have significant drawbacks and even when using them in combination and providing discrepancy scores, cognition alone (and with demographic factors), is not a useful predictor of amyloid status. Discrepancy scores have potential merit for screening purposes within early AD settings and in order to move this research area further forward, longitudinal measurement and analysis with broader AD biomarkers is a necessity for future study.

Appendices

Appendix I – Current Clinical Trial Summary

The current landscape of drug development is detailed within this appendix. This review looks at the key areas for compound development for AD as of year-end 2021. The current areas of focus for potential disease modifying compounds are predominantly split into two main areas; monoclonal antibodies for amyloid beta and tau specific antibodies (Cummings et al., 2021). There are also a large number of compounds outside to this that pertain to other aspects of the disease. The predominance for non ATN targeting compounds is within phase I & II of development. However, the scope of targets has greatly increased over the past 5 years, with now the majority of compounds not directly targeting amyloid.

Given the complete halt of all trials of compounds modulating BACE inhibition, the following section may seem redundant, however, as the cohort under investigation was from two large phase III trials of a BACE inhibitor (elenbecestat), the discussion of these compounds warrants discussion.

BACE Inhibitors

With the amyloid hypothesis (discussed in **Chapter 1**) driving both research and clinical development alike, the first class of drugs stemming from this hypothesis was the beta-site amyloid precursor protein cleaving enzyme inhibitors (BACEi). As shown in **Figure A1.1**, the depiction of the formation of Aβ plaques by $A\beta_{42}$ deposition. Inhibiting the beta secretase cleavage at point two in **Figure A1.1** is hypothesised to rapidly decrease the amount of toxic $A\beta_{42}$ deposited in the cortices.

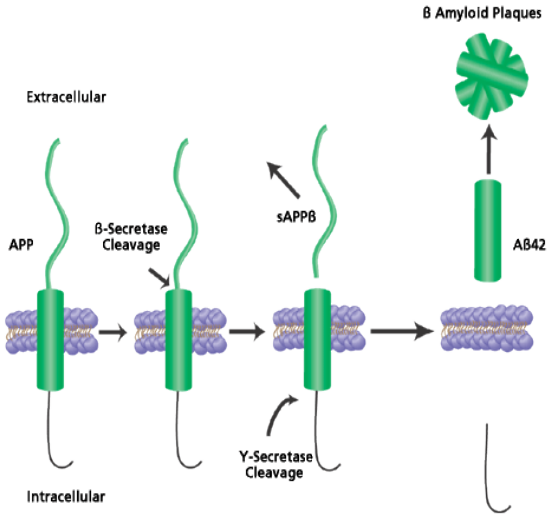


Figure A1.1. Process for the formation of Amyloid Beta 42

The proteolytic processing of APP (amyloid precursor protein) begins in the extracellular space with β-secretase cleaving APP to form sAPP β and CTF99/CTF89 (Chow et al., 2010). sAPPβ is then cleaved again at the plasma membrane by γ-secretase which produces the Aβ fragment of variable length and another amino acid complex known as AICD50. Beta-site cleaving enzyme 1 (BACE1) is the major β-secretase in the cortex (Vassar et al., 1999) and has been a major therapeutic target for new disease modifying

compounds (Cummings et al., 2018). BACE inhibition has been shown to be beneficial at many aspects of impairment in AD mouse models [memory measures, cellular level impairments & LTP] (Keskin et al., 2017), however there are conflicting findings within human studies of these compounds, most notably widespread failures, from either side effects or a lack of efficacy, in larger human clinical trials (Mullard, 2018).

Thus, the hypothesis around the mechanism of action for BACEi's is that the inhibition of this cleaving enzyme will allow the normal processing to occur with APP, thus nullifying the toxic cascade of $A\beta_{42}$. However, there have been a number of recent setbacks that challenge the validity of this hypothesis, with all of the BACEi compounds in development having their major phase III trials halted. All were stopped due to a significant adverse side effect profile/futility (Panza et al., 2018). To the surprise of most in the field the results from Verubecestat showed a significant impairment on treatment when compared to placebo on ADAS-Cog in the dosing cohort (Egan et al., 2019). This compound worsened cognition to such a degree that it showed up at the three-month time point and was maintained throughout the dosing duration. Eigen and colleagues hypothesised this could be down to a number of reasons, but the key differences between all of the BACE compounds are the selectivity of BACE1 over BACE2 and amount that the primary dose lowered amyloid by (see **Table A1.1**).

Table A1.1 Breakdown of the BACEi Compounds

Compound	Company	AD Population	Phase	CSF Aβ	Status	Cognitive Outcome
Verubecestat[1]	Merck	Mild to Moderate & Prodromal	III	-60% to -80%	Stopped for safety	Trend for/ cognitive worsening
Lanabecestat[2]	Lilly & AZ	Prodromal - Mild	III	-55% to -75%	Stopped for safety	Trend for/ cognitive worsening
Atabecestat[3]	J&J	Preclinical	III	-50% to -82%	Stopped for safety	Trend for/ cognitive worsening

LY3202626[4]	Lilly	Mild	II	-70% to -90%	Stopped for safety	Trend for cognitive worsening
Elenbecestat[5]	Eisai & Biogen	Prodromal-Mild	III	-57%	Stopped for futility	No data available
Umibecestat[6]	Novartis & Amgen	Preclinical	II/III	-20% to -90%	Stopped for futility	No data available

1. Egan et al., 2019. 2. Wessels et al., 2020. 3. Henley et al., 2019. 4. Lo et al., 2021. 5. Roberts et al., 2021. 6. Tariot et al., 2020.

One question that remains elusive to answer is the level of interaction of BACE2 versus BACE1 within these compounds. All of the aforementioned BACEi compounds have a degree of interaction with BACE2 as well as its primary target of BACE1. BACE2's function is comparatively under researched, in the rodent brain this receptor appears in oligodendrocytes but also in some other astrocytes and neurons (Voytyuk et al., 2018). It has been shown to be fairly rudimentary in the brain but then ramps up its activity due to inflammation implicating some potential role in AD. Nevertheless, BACE2 is widely considered to be irrelevant to amyloid pathogenesis or cleavage, due to it being expressed at significantly lower levels than BACE1 in the cortex (Dominguez et al., 2005; Ahmed et al., 2010). However due to the varied side effect profile of many of the BACEi's in development, the interaction of this secondary BACE enzyme is still yet to be fully elucidated.

With there being over 40 known substrates of BACE1 one potential culprit is Seizure Protein 6 (SEZ6) which in rodent models has been shown to maintain in LTP and helps dendritic spine density (Kuhn et al., 2012). This substrate is thought to be a leading candidate for these cognitive impairments shown in BACEi's due to BACE1 knockout mice being impaired in both LTP and spine density. It is hypothesised that in humans these are manifested as cognitive deficits or other neuropsychological symptoms (Filser et al., 2015; Vassar, 2019; Zhu et al., 2019). Another substrate that has been suggested to be interacting with this process is a close homolog of L1, CHL1. It has been shown to guide axons to their targets. Robert Vassar's group (Ou-Yang et al., 2018) recently showed that knocked out BACE1 in adult mice resulted in stunted growth of mossy fiber axons in the hippocampi manifesting in impaired LTP function. The last of three main hypothesised substrates is

neuregulin which helps maintain myelination and muscle spindles; its absence during development causes seizures and could play a role in the cognitive deficits seen in humans.

Overall, the BACEi field is still in a state of unknowns. As the recent failures showing a class effect at the current dosing levels, with these compounds having deleterious cognitive performance in humans. All compounds in this class have been discontinued from any further development (much like the gamma-secretase inhibitors). However, there is some suggestions lower levels may not induce these transient deleterious effects on cognition whilst maintaining reductions in amyloid. Nevertheless, given the outlay involved and direction of travel of the field, any restarting of development seems highly unlikely, for now.

Immunotherapeutics targeting Aβ

The most expounded anti-Ab approach is that of immunotherapy, encompassing both active vaccines that stimulate an individual's innate immune system to produce its own antibodies and passive immunisation through the administration of exogenous antibodies (mAbs). This pillar of development is a class of compounds that is routinely utilised in other areas of drug development and is currently showing the most promising trial results for disease modification in AD. This class of compounds acts directly on the different isoforms of Aβ, whether that be oligomers, plaques or fibrils or a combination of these, inducing essential phagocytosis of these protein formations through microglia or complement activation (Lannfelt et al., 2021; Prins & Scheltens, 2013). This is hypothesised to arrest the amyloid cascade and thus prevent further neurodegenerative, cognitive and functional decline by removing these epitopes.

Table A1.2. Breakdown of the current Aβ antibodies in development at Phase III

Compound	Company	AD Population	Epitope	Aβ Conformations Recognised°	Status
Crenezumab	Roche Genetech	Prodromal	AA13-24	M, O, F	Discontinued
Gantenerumab	Roche	Prodromal to Mild	AA3-12, AA18-27	M (weak), O, F	Ongoing in early AD –

					prior failed trials
Solanezumab	Eli Lilly	Prodromal to Mild	AA16-26	M	Discontinued
Aducanumab*	Biogen/ Eisai	Prodromal	AA3-6	O, F	Approved, Trials halted
Bapinenuzimab	Pfizer	Mild to Moderate	AA1-5	M, O ,F	Discontinued
Donanemab*	Lilly	Prodromal	Pyrogluta mated	O,F	Ongoing
Lecanemab/ BAN2401*	Eisai/ Biogen	Prodromal	Protofibrils	O,F	Ongoing

Trial data gathered from clinicaltrials.gov. *Positive results from prior phase I/II study. Confirmations targeted gathered from trial data & Lannfelt et al., 2022. ºM=Monomer O=Oligomer and F=Fibril

The majority of these compounds have failed at phase III. With the inception of trials for mAbs the population recruited was one much later in the disease course and as such already had widespread disease pathology. Unlike some of the early BACEi trials, mAb phase III trials all included amyloid positivity assessments in order to ensure target engagement. As with the rest of the field of AD there was a consensus that trials needed to target subjects earlier in the disease course in order to effectively interact with the disease pathology. Gantenerumab is one compound has consistently failed at phase III and has some preliminary phase II data that showed no treatment effects on primary or secondary outcomes at either dose suggesting a lack of efficacy. However, Roche have pushed forward with this compound by upping the dosage for the treatment arm at phase III as they saw enough target engagement to warrant continued development (Cummings et al., 2021).

The two most promising compounds currently in development have both shown efficacy in early phase trials. Lecanemab and Donanemab are being developed by Eisai and Lilly respectively. Donanemab targets a specific confirmation of amyloid called pyroglutamated amyloid beta. In trials Donanemab has shown significant clearance of amyloid plaques, which has ultimately led to a cessation of dosing after subjects become amyloid negative on scans (Alawode et al., 2021). However, this antibody targets a specific type of amyloid only found in cerebral amyloid plaques (DeMattos et al., 2012) and not found in any biofluids (CSF or plasma). This suggests that whilst providing robust and quick plaque clearance, it does not target the smaller (still pathological) confirmations of amyloid that cause degeneration/cascade effects. Given the slow build-up of amyloid (~1-2% per year), once cleared plaques will still continue to form, just at a slow rate, akin to early in the disease process. Therefore, whilst disease modifying, it seems to remove “gravestones” of the pathological process rather than treating the underlying reasons for amyloid build up.

Lecanemab is also in phase III trials with previous data also indicating a positive treatment effect on the disease course. Over the 18-month phase II trial fibrillar amyloid was reduced in all treatment groups compared to placebo by up to 93% in the highest dose group (Swanson et al., 2018; 2021). In this group cognitive decline was also reduced by 47% on ADAS-Cog and by 30% on the ADCOMS composite measure. With over 850 subjects with MCI or mild AD undergoing the trial this is the largest one to feature both significant amyloid reduction and a downstream benefit on cognition and function. Nevertheless, the analysis was complicated due to an adaptive randomisation design within the study and due to the EMA halting stratification of ApoE carriers to the highest dosing group due to side effect concerns. Further analysis yielded more details on progression rates by genotype which still strongly supported the treatment effect even with the uneven stratification groupings. A phase III trial is now underway and the possibility of a disease modifying treatment is something that could be closer to approval. On top of this further unblinded data from the open label extension (OLE) study from the phase II trial, has shown continued and sustained disease modifying trajectories (Irizarry et al., 2022). There was an average of a two-year gap period, between the end of the phase II trial and the start of the OLE. Across both placebo and active arms decline occurred across these two years, however the treatment difference between the two groups was maintained. This is what would be expected from a disease modifying therapy. In line with the Donanemab data and dosing regimen, this is also being explored within this OLE as data has shown that whilst plaque levels are fairly stable across this two-year gap, fluid plasma markers of amyloid and tau still increase significantly, which is argued to warrant continued treatment with Lecanemab. This is primarily due to Lecanemab targeting protofibrils and smaller conformation of amyloid beta, on top of amyloid plaques.

This brings us to the discussion of the only approved monoclonal antibody to date, Aducanumab. This is currently only approved in the US after the both phase III trials of the compound were halted by Biogen in 2019. All other regulatory authorities have provided a negative decision on licencing the compound in their region. This is primarily due to an inability for Biogen to demonstrate a consistent meaningful efficacy profile of the compound. Their two phase III trials had divergent results upon final analysis of the trial data (Budd-Haeberlein., 2022; 2020; Knopman et al., 2021). With one trending positive and the other showing no signal of treatment effect at all. The post-hoc analysis showed that given enough of the drug subjects taking aducanumab has a statistically significant slowing of decline on the CDR-SB compared to placebo. It is widely viewed by the field and specifically prescribing physicians in the US that these trials did not satisfactorily provide evidence the aducanumab ameliorates cognitive decline due to AD. This controversial approval has been discussed at length by all sides and full

discussions are not warranted here but can be found (Knopman et al., 2021; Liu & Howard, 2021; Dunn et al 2021; Alexander et al., 2021).

The prior trial data showed in an analysis of the phase I trial, dose dependant reductions in PET SUVR for brain amyloid load and further to this whilst not powered for efficacy on the cognitive endpoints the analysis showed a dose dependant slowing at 1 year in MMSE and CDR (Sevigny et al., 2016). These results support the amyloid hypothesis and the last 20 years of drug development and a lot of hope has been placed on achieving a positive read out at phase III for this compound. However, as part of ongoing monitoring Biogen have added another 20% to their placebo group due to “unexpected variation” in their study population which has somewhat quelled the excitement for the results. The trials of Aducanumab have utilised the RBANS as a key efficacy endpoint as well as the MMSE, CDR and ADAS-Cog, all of which are discussed in detail later. The positive direction of results from the early phase trials of Aducanumab also encourages further trials to continue of Gantenerumab, due to its similar epitope positional engagement (Lasser et al., 2015).

However, there have been two major issues for this immunotherapeutic approach in AD, one being the size of the molecules, which have resulted in some compounds having a distinct lack of brain penetration and needing an increase in the amount of the compound needed (Lemere & Masliah, 2013). And secondly a far more surprising and major issue of ARIA (amyloid related imaging abnormalities) [ARIA-E & ARIA-H] (Sperling et al., 2011; Prins & Shceltens, 2013; van Dyck, 2018). The size of the molecules is something that does not become an issue in other areas of the body but with low levels of antibodies traversing the blood brain barrier (BBB), penetration is argued to not be a necessity to induce desired target engagement (van Dyck, 2018). Novel attempts to overcome this hurdle have been mixed with targets including receptor targeting on the BBB to induce active transportation of mAbs through to the CNS, delivering genes that encode the mAbs and use of ultrasound to widen the BBB (Montoliu-Gaya & Villegas, 2016; Jordao et al., 2013; 2010). However, none of these techniques have been implemented in clinical trials as the target engagement of these compounds at higher doses has been efficacious.

Nevertheless, with higher dosing comes another difficulty, ARIA. ARIA is split into two categories of disturbances, ARIA-E standing for vasogenic edema and ARIA-H for microhaemorrhage. Most commonly with mAbs being found to have subjects who develop ARIA-E at the beginning of treatment (Sperling et al., 2011). These MRI signal alterations on the FLAIR (fluid attenuation inversion recovery) sequence have been seen across most trials of Aβ antibodies (van Dyck, 2018) and profess as large areas of inverted colour on these scans. They are occasionally accompanied by headaches/migraines and is predominantly ARIA-E or ARIA-H. ARIA-E has been found to be strongly

associated with compound dose as well as ApoE status, however in approximately 78% of cases, it is asymptomatic and self-limiting (Sperling et al., 2012; Salloway et al., 2014). In previous trials occurrence of ARIA have resulted in a cessation of treatment but as the field has developed the wider understanding of this phenomenon regulators are taking a more lenient approach and this may now not require temporary suspension of treatment as any serious complications are extremely rare and are balanced against the consequences of untreated AD. Previous trials have come up against this (Bapinenuzimab and Solanezumab), which limited maximum tolerated doses and has been argued to have impacted the study outcome, whilst more recent trials with the same frequencies of ARIA occurrences have been tolerated at higher doses with similar isotope engagement and pharmacology (Aducanumab) (van Dyck, 2018).

Overall, this class of compound has shown the most promising results with two compounds currently showing disease modifying efficacy as well as significant improvements and trends to these in cognition and functional measures (Swanson et al, 2021; Sevigny et al., 2016). The vast number of previous failures tend to not suggest the wrong target for disease modification but for a lack of efficacy due to low dose, lack of brain penetration and poor sensitivity of the endpoints employed. This coupled with a greater understanding of disease progression and earlier disease course engagement has led to better designed trials that have yielded some promising signs of clinical benefit even when not powered to do so. Only once the Lecanemab and Donanemab studies have been fully completed will the immunotherapeutic approach for targeting Aβ be fully understood.

Tau Therapeutics: Modulation, Stabilisation & Immunotherapy

With the ever increasing number of failures in the quest for the amelioration of amyloid, some companies have shifted their pipelines to target the younger brother of AD pathogenesis, tau. Tauopathies are present in many disorders however, in AD the presence of tau tangles are one of the key hallmarks of the disease and its progression (Grundke-Iqbal et al., 1986; Jack et al., 2018). As shown in the Jack graphs (Jack et al., 2010; 2013), tau increases at an exponential rate in AD, but in this model, it does not reach its maximum impairment at the same time as a subject transitions into MCI, thus giving a clinical presentation of symptoms prior to reaching its impairment saturation point. Coupled with the disappointing results of amyloid targeting therapies, this is one of the reasons therapies targeting this abnormality have begun to reach clinical development.

Currently three approaches to tau amelioration are being trialled in patients; phosphorylation modulation, microtubule stabilisation and tau immunotherapy (Medina, 2018). Firstly, the compounds targeting tau phosphorylation modulation are based upon the relationship between phosphorylated tau and disease pathology in AD and as such, a number of

protein kinase inhibitors have been considered as potentially efficacious therapeutic agents (Wang et al., 2007; Hanger et al., 2009; Medina & Avila, 2015). However, as kinases are abundant both intra and inter-cellularly, specificity and toxicity concerns are prominent and highly challenging for developers. Tideglusib, a GSK-3β inhibitor (one target for this modulation hypothesis), is an approved compound that has shown efficacy in preclinical studies (Sereno et al., 2009) but has failed to find efficacy in human clinical trials (Del Ser et al., 2013). Another target of this hypothesis is a tyrosine kinase, Fyn, as this also interacts with the amyloid signalling pathway (Nygaard et al., 2015).

As one of the key properties of tau is its stabilisation of microtubules, as the protein becomes misfolded and phosphorylated, one hypothesis is that the microtubules come apart causing axonal transport defects and synaptic dysfunction (Li & Gotz, 2017). It stands to reason then that compounds that increase the stabilisation of these microtubules may show promise in effecting tau mediated dysfunction. To this degree a number of small molecules have gone into clinical development, however, all of the known compounds have failed to meet any primary or secondary endpoints in MCI and AD populations and have also shown negative side effect profiles (Medina, 2018). To note, methylene blue (TauRX) and its derivatives have also received attention for their trials in AD, with the trial sponsors suggesting potential efficacy in tau disease modification by arresting tau aggregation (Wilcock et al., 2018). However due to a lack of efficacy in all trials with this and similar compounds (Gauthier et al., 2016; Feldman et al., 2016) as well as its publicised pleiotropic effects (Bakota & Brandt, 2016), it is not considered as a tau-based therapy within this review.

The main emphasis of tau therapeutics however, much like Aβ, is immunotherapy. Hyperphosphorylation has widely been shown to be the pre-eminent factor in inducing aggregation of tau tangles (Iqbal et al., 2005; Alonso et al., 1996), many of the compounds described below focus on domains rather than any specific phospho-epitopes with many showing positive results in transgenic animal models (Chai et al., 2011; Yanamandra et al., 2013; Sankaranarayanan et al 2015). This hyperphosphorylation has also been shown to induce prion like behaviour (Hu et al., 2017; Dai et al., 2018) as normal tau molecules can be sequestered by hyperphosphorylated tau into aggregates (Alonso et al., 1996). Nevertheless, during the process from a soluble protein into insoluble aggregates and deposits, tau also undergoes numerous modifications outside phosphorylation, all of which makes it arguably a more challenging target than Aβ.

Table A1.3. Current Tau Antibodies In Clinical Development

	Tau Epitope	**Current Stage**	**Company**	**Status**

ACI-35	P-Ser396,404	Phase I	AC Immune/J&J	On hold
Lu AF87908	Unknown	Phase II	Lundbeck	Ongoing
Tilavonemab	Tau25–30	Phase II	AbbVie	Ongoing
Zagotenemab	7–9, 312–342	Phase II	Eli Lilly	Ongoing
JNJ-63733657	Middle region	Phase II	J&J	Ongoing
BIIB076	Unknown	Phase I	Biogen/Eisai	Ongoing
E2814	Mid-domain	Phase I	Eisai	Ongoing
Bepranemab	235–246	Phase II	UCB	Ongoing

Data adapted from Cummings et al (2021) & Medina (2018).

As shown in the table, all of these compounds are currently undergoing clinical trials in the very earliest stages and very little is known about their efficacy in human subjects. As such a number of key questions are still to be answered, with it primarily unknown which tau species will produce the most efficacious results when targeted and for the majority of these compounds the safety profiles are also still relatively unknown as they are still in the infancy of clinical development.

Appendix II – Pathological Amyloid Biology, Composition & Downstream Processes

Amyloid Composition

$A\beta_{42}$ is the most common protein composition found within the brains of those with AD. However, the length of the protein chain can be much longer, with the increase in length, leading to greater increases in its self-aggregation (Kim et al., 2007). Typically, in AD due to the cleavage sites on APP, this peptide is formed of 42 amino acids and at this length is thought to be the most toxic due to its abundance in the pathology of sporadic AD (Haass & Selkoe, 2007; Benilova et al., 2012). Consequently, understanding the mechanisms that circumscribe Aβ length is highly pertinent for both the fundamental understanding of the disease process and critically for the advancement of efficacious disease modifying strategies for AD. Looking within non demented controls Haass and Selkoe showed that 50% of the Aβ fragment ends at amino acid 40 ($A\beta_{40}$), 16% at amino acid 38 ($A\beta_{38}$) and 10% is $A\beta_{42}$.

The length of the Aβ fragment is dependent upon the position of the two sequential cleavages along the amyloidogenic proteolytic pathway, both of which have been shown to be imprecise (Kummer & Heneka, 2014). This is argued to be due to the APP isoform being present at several lengths ranging from 695 (the most abundant) to 770 amino acids and with the β-secretase (BACE1) step cleaving APP at a minimum of two positions, the γ-secretase then generates a variety of different Aβ fragments from 34 to 50 amino acids by its cleavage of sAPPβ. This process is not fully understood but the most common final cleaving positions are 38, 40 and 42, which is the last stage of the stepwise three stage endo-proteolysis that occurs within the trans-membrane domain (Takami et al., 2009; Qi-Takahara et al., 2005). One suggestion for this toxic aggregation in AD at increased lengths is due to the destabilization and decrease in productivity of γ-secretase due to PSEN mutations (Chavez-Guiterrez et al., 2012; Fernandez et al., 2014). The reduction in processivity amounts to the number of cuts made by γ-secretase decreasing and thus shifting the Aβ toward longer and more amyloidogenic peptides due to these PSEN destabilisations (Szaruga et al., 2017). PSEN1/2 are the catalytic subunits of distinct γ-secretase intra-membrane protease complexes (De Strooper et al., 1998; Wolfe et al 1999) and it is these proteases that have been shown to have a causative role within both, in vitro and murine models, on APP cleavage that mimics the pathogenic process of autosomal dominant AD (Szaruga et al., 2015). However, whilst there is strong evidence for this relationship within animal models there are still a number of unknown questions around this pathogenic process in humans, for example it is still not known what drives the sequential cleavage of APP, or how clinical mutations in PSEN lead to this release of longer Aβ peptides and

hence there is still some debate around the pathogenic role of PSEN in human AD (Veugelen et al., 2017).

Plaques, the large clumps of Aβ clearly visible in post mortem brain tissue, are the final stop on the pathogenesis and aggregation pathway of $A\beta_{42}$. As such these large aggregates were widely thought of as the key components that drive the cascade of neuropathology (Hardy & Higgins, 1992). However, improvements within in vivo and in vitro studies have allowed for a more precise insights to the behaviour of Aβ within models of AD, showing clear differences in $A\beta_{40}$ fibrils and $A\beta_{42}$ fibrils that mimic human AD pathogenesis (Petkova et al., 2002; Schmidt et al.,2009; Kirchnser et al., 1986; Sikorski et al., 2003; Saito et al., 2014). Notably a study by Yoshiike and colleagues (2003) found the most toxic forms of $A\beta_{42}$ aggregates to be formed at the initial stage of fibrilliogenesis with consistent findings also specifically showing Aβ oligomers to have the most potent synaptotoxcity and neurotoxicity when compared to plaques in adult rodents (Shankar et al., 2008).

The strongest quantitative way to measure this cortical damage by AD pathogenesis has long been recognised as decreased numbers of synapses (Selkoe & Hardy, 2016). This has been seen in laboratory studies to be driven by Aβ oligomers which impairs both synaptic structure (dendritic spines) and synaptic function (LTP). These oligomers are considered to be highly toxic yet soluble and have been shown to dose dependently decrease synaptic function in healthy adult rats resulting in impaired memory of learned behaviours (Shankar et al., 2008).

The toxicity of $A\beta_{42}$ aggregates are also equated to the ratio of $A\beta_{42}$: $A\beta_{40}$ peptides (Pauwels et al., 2012). In normal controls the ratio is approximately 1:9 however increased $A\beta_{42}$: $A\beta_{40}$ ratio (1:10) is seen to correspond to more aggressive forms of sporadic AD and alter synaptic activity, neuronal viability and memory formation in murine models (Citron et al., 1997; Mann et al., 1996; Wang et al., 2006; Duff et al., 1996; Scheuner et al., 1996). Subtle shifts such as these in this ratio have also been shown to dramatically influence the formation of neurotoxic oligomers (Kuperstein et al.,2010; Yoshiike at al., 2002). Both $A\beta_{42}$ and $A\beta_{40}$ have analogous chemical natures however their structural and biological properties seem to be significantly heterogeneous, with $A\beta_{42}$ being shown to be highly fibrillogeneic and more prone to neurotoxic aggregations than the shorter $A\beta_{40}$ (Kuperstein et al., 2010; Jarrett et al., 1993; Bitan et al., 2003; Chen & Glabe, 2006). Whilst preliminary evidence has seemed to show the effect $A\beta_{42}$ and $A\beta_{40}$ have on each others aggregation rates, there has been some evidence to suggest that $A\beta_{40}$ formation actually inhibits the aggregation of $A\beta_{42}$ (Wang et al., 2006; Yoshiike et al., 2002; Synder et al., 1994; Frost et al., 2003; Jan et al., 2008; Kim et al., 2007; Yan & Wang, 2007). As suggested earlier with the

alterations in processivity of γ-secretase by PSEN's this increased ratio is modelled to be a direct result of this alteration in homeostatic proteolysis (Selkoe & Hardy, 2016).

Overall, evidence seems to suggest that but while not conclusively, $A\beta_{40}$ has a protective effect against the formation of $A\beta_{42}$ neurotoxic aggregates in humans. With the oligomeric forms of $A\beta_{42}$ also being shown to be the key components of this neurotoxicity rather than the larger more visible plaques. Understanding this biological basis and pathogenesis is critical within drug development as the compounds currently under development have a range of interactions which are not ubiquitous or pervasive in nature and are highly specific to different formations of Aβ. In order to understand why some therapeutics are ineffective, the molecular biology is key to understand before looking for more fundamental, cognitive or methodological defects. Nevertheless, it is only until these hypotheses are consummately engaged by mechanistic altering compounds in humans that the in vivo and in vitro theories can be thoroughly tested. As whilst these mechanistic alterations hypothesised can explain the cascade of pathology from a genetics and biological basis, many more factors interact with the disease course within heterogeneous real-world populations and it is only after a significant amyloidogenic build up that crucial cognitive and clinical deficits occur.

Immune response & TREM2

A secondary effect of these senile plaques is that they are thought to induce an inflammatory response from the brain's innate immune system (McGeer et al., 1989). This cascade of pathology continues with the inflammatory response to these plaques being shown to be the step between plaque accumulation and tangles pathology (Felsky ey al., 2019). At this stage of pathology, the brain's instinctive response to plaques is primarily driven by the microglia. Many researchers currently argue that the development of AD is primarily due to this failed, and/or impaired, clearance mechanisms of toxic Aβ species.

Recent studies have shown the effects that key risk alterations to this system can have, both on plaque formation and clearance (Jones, 2010). Three genes have been prominently researched in this regard and found to be risk factors implicated in alterations in Aβ deposition. These have emerged from numerous GWAS studies and subsequent animal work as primary influences to microglia response to Aβ (Selkoe & Hardy, 2016). Complement Receptor 1 (CR1) has been shown to inhibit microglial activation when blocked and also shown to proliferate microglial phagocytosis (Lambert et al., 2009; Crehan et al., 2013). Secondly is CD33 who's inactivation in primary microglia potentiates microglial uptake of Aβ (Bertram et al., 2008; Griciuc et al., 2013). However, the most prominent of these three is TREM2, a triggering receptor found primarily in microglia and

macrophages. This has been shown to be responsible for the sustaining of Aβ phagocytosis by microglia (Wang et al., 2015). All three of these genetically implicated microglial proteins undergo increased expression during Aβ plaque formation (Griciuc et al., 2013; Wang et al., 2015; Matarin et al., 2015) with CSF levels of TREM2 tracking with Aβ plaque load suggesting it may prove a useful biomarker of this plaque load (Suarez-Calvet et al., 2016).

TREM2 promotes microglia to induce phagocytosis of apoptotic neurons and other neuronal debris (Takahashi et al., 2005, 2007; Hamerman et al., 2006; Turnbull et al., 2006; Neumann & Takahashi, 2007; Hsieh et al., 2009; N'Diaye et al., 2009; Ito & Hamerman, 2012; Hickman & El Khoury, 2013; Jiang et al., 2014). Due to apoptotic neurons, such as those affected by Aβ, expressing a specific ligand for TREM2, reductions in levels of TREM2 could lead to reduced interaction with this specific ligand and thereby attenuating the signal for phagocytosis (Lue et al., 2015). This suggests TREM2's intrinsic nature to this homeostatic process. However, when this has been studied within the context of AD the evidence is less clear. The upregulation in a subset of microglia within amyloid plaques in mice, suggests that this known immunoregulatory function is compromised during plaque development (Guerreiro et al., 2013). One hypothesis of why this increase in function leads to a lack of microglia activation is currently thought to be primarily due to the extremely rare R47H mutation of TREM2 (Selkoe & Hardy, 2016). This is the most studied TREM2 mutation which has been shown to impair TREM2's function in relation to a response to Aβ formations in vivo (Song et al., 2018). It is possible that, although the microglial response to Aβ occurs very early in the progression of AD in these animal and in-vitro models, compared to that of the human form of AD, the Aβ accumulation itself may require much longer to be homologous with the human form of AD (Colonna & Wang, 2016). Furthermore, the incongruities in the effects of TREM2 paucity on Aβ accumulation may be in part due to the varied time points of the analyses indicating methodological refinement is needed for consistent analysis.

Cleavage of TREM2 by ADAM10 or ADAM17 has been shown to produce a soluble TREM2 (sTREM2) protein fragment (Ma et al., 2016; Begum et al., 2004; Thornton et al., 2017; Schlepckow et al., 2017; Jin et al., 2014; Celarain et al., 2016). This has shown to be elevated in the CSF of subjects with AD (Suarez-Calvet et al., 2016a; Heslegrave et al., 2016; Piccio et al., 2016), but the evidence is somewhat conflicting (Kleinberger et al., 2014; Henjum et al., 2016) as such the relationship between elevated sTREM2 and markers of cortical inflammation within AD is currently unclear. In a recent study by Bekris and colleagues (2018) a consistent association of peripheral or central markers of inflammation with CSF sTREM2 levels was not found. This suggests a limited impact of general inflammation on sTREM2 levels. However, it has been posited that sTREM2 actually promotes microglial survival and stimulates increases in production of innate immunity factors

(Zhong et al., 2017). Yet whilst CSF sTREM2 levels have been shown to be influenced by age (Suarez-Calvet et al., 2016), which may underpin longitudinal analysis of this fragment, the function of sTREM2 is still unclear and further research is needed to probe these mechanisms further.

Nevertheless, in two different mouse models, data on showing TREM2 promotes a broad array of microglial functions in response to Aβ deposition, not just phagocytosis; was analysed at 3, 4 and 8 months (Ulrich et al., 2014; Jay et al., 2015; Wang et al., 2015 [respectively]). A time-course analysis of Aβ accumulation in relation to these models is needed to clarify these discrepancies. Also in inflammatory conditions, it is common for microglia-like cells circulating in the blood to generate monocytes that cross the blood–brain barrier (BBB) and act like cortical microglia. It is therefore unclear whether the Aβ-reactive microgliosis that is observed, involves cells that are derived from cortical microglia or infiltrating monocytes. These studies indicated that the blood monocytes express TREM2 and cluster around Aβ deposits, but only after a greater length of time. However, these monocytes fail to modify Aβ deposits, despite adopting characteristics of cortical microglia (Varvel et al., 2015; Prokop et al., 2015) showing peripheral monocytes primarily are unable to become fully functional microglia or may require a longer process of maturation. All studies showed the pathological changes primarily occurred in the hippocampus suggesting a regional role of TREM2 that should be further investigated that is also concurrent with the earliest stages of pathology. Furthermore, in these quoted studies, Aβ accumulation was examined in which the mutations targeted different regions of the TREM2 gene, therefore it is possible that the discrepancies may be related to as-yet-undefined disruptions of the TREM locus engendered during targeting, which may affect the expression of other TREM family members that are also involved in AD.

These findings indicate that TREM2 promotes a broad array of microglial functions in response to Aβ deposition, rather than only phagocytosis. Although the results from all of the mouse studies described above demonstrated the importance of TREM2 in microgliosis, their findings on the effect of TREM2 deficiency on Aβ accumulation were contradictory, which led to disparate interpretations of the mechanisms through which TREM2 deficiency affects Aβ-reactive microgliosis. Thus, although there is a consensus that TREM2 is required for microgliosis, the origin of the microglia surrounding Aβ plaques and their effects on Aβ accumulation remains indistinct (Tanzi, 2015; Colonna & Wang, 2016). However, current research indicates that whilst showing TREM2 plays a role in AD pathology, whether this is detrimental or a beneficial one still remains wholly unclear, with the further aforementioned research directions needed to better understand this facet within AD and to uncover potential therapeutic targets (Deming et al., 2018; Rauchmann et al., 2019).

Appendix III – R Code for all studies

Study 1

```
library(readxl)
E2609_G000_301_COG_AMYLOID_CSF_PET_WCT_20191223_1 <-
read_excel("~/Documents/PhD/Eisai
Data/E2609_G000_301_COG_AMYLOID_CSF_PET_WCT_20191223_1.xls")
View(E2609_G000_301_COG_AMYLOID_CSF_PET_WCT_20191223_1)

library(readxl)
DSData <- read_excel("DSData.xls")
View(DSData)

install.packages("tidyr")

options(digits=5)

getwd()
setwd("~/Documents/PhD/R/")
#Remove rescreens
library(readxl)
DSData <- read_excel("~/Documents/PhD/R/Discrepancy
Scores/DSData.xls")
dsnors <- DSData[is.na( DSData$PREVSCRN | DSData$PRVSCRN1 |
DSData$PRVSCRN2), ]

#Find any outliers
tblDET <- with(dsnors, table(CBDETZ))
barplot(tblDET)
tblIDE <- with(dsnors, table(CBIDEZ))
barplot(tblIDE)
tblOBM <- with(dsnors, table(CBOBMZ))
barplot(tblOBM)
tblOCL <- with(dsnors, table(CBOCLZ))
barplot(tblOCL)
tblMMSETOTS <- with(dsnors, table(MMSETOTS))
barplot(tblMMSETOTS)

boxplot(dsnors$CBDETZ, dsnors$CBIDEZ, dsnors$CBOBMZ, dsnors$CBOCLZ,
  names=c("Detection","Identification", "One Back","One Card Learning"),
  ylab="z-score")

stat.desc(dsnors$CBDETZ)
stat.desc(dsnors$CBIDEZ)
```

```
stat.desc(dsnors$CBOBMZ)
stat.desc(dsnors$CBOCLZ)

#subset dataset by CBB values  `
dsnors <- DSData[is.na( DSData$PREVSCRN | DSData$PRVSCRN1 |
DSData$PRVSCRN2), ]
dsn1<-subset(dsnors, MMSETOTS >(17))
ds1 <-subset(dsnors,CBDETZ >(-6.077))
ds2 <-subset(ds1,CBIDEZ >(-7.607))
ds3 <-subset(ds2,CBOBMZ >(-6.231))
ds4 <-subset(ds3,CBOCLZ >(-4.162))
dsnors1 <-subset(ds4,MMSETOTS >(17))

#compute cyrstallised composite
dsnors1 <- mutate(dsnors1, FComp = (CBDETZ + CBIDEZ + CBOBMZ +
CBOCLZ)/4)

#compute fluid composite (MMSE Language Domain #6,7,8,9,10)
dsnors1 <- mutate(dsnors1, CComp = MMSE06S + MMSE07S + MMSE08S +
MMSE09S + MMSE10S)

#Remove NAs from dataset for CComp & FComp
dsnors2 <- subset(dsnors1, !is.na(CComp))
dsnors3 <- subset(dsnors2, !is.na(FComp))

#CComp Z-Score
summary(dsnors3$CComp)
stat.desc(dsnors3$CComp)
dsnors30 <- mutate(dsnors3, CCompZ = (CComp-7.55)/6.257920e-01
(digits=3))


#compute discrepancy score
dsnors3 <- mutate(dsnors3, DS = FComp - CCompZ)

#descriptive stats for DS
stat.desc(dsnors3$DS)
tblDS <- with(dsnors3, table(DS))
barplot(tblDS)

#Regional analysis of cognitive measures
dsnors3 <-
  dsnors3 %>%
  mutate(
    Region = case_when(
      COUNTRYC == "United States of America" ~ 1,
```

```
    COUNTRYC == "Canada" ~ 1,
    COUNTRYC == "United Kingdom of Great Britain and Northern Ireland" ~
2,
    COUNTRYC == "Spain" ~ 2,
    COUNTRYC == "France" ~ 2,
    COUNTRYC == "Germany" ~ 2,
    COUNTRYC == "Croatia" ~ 2,
    COUNTRYC == "Denmark" ~ 2,
    COUNTRYC == "Poland" ~ 2,
    COUNTRYC == "Czechia" ~ 2,
    COUNTRYC == "Slovakia" ~ 2,
    COUNTRYC == "Austria" ~ 2,
    COUNTRYC == "Italy" ~ 2,
    COUNTRYC == "Portugal" ~ 2,
    COUNTRYC == "South Africa" ~ 2,
    COUNTRYC == "Bulgaria" ~ 2,
    COUNTRYC == "Japan" ~ 3,
    COUNTRYC == "China" ~ 3,
    COUNTRYC == "Korea, Republic of" ~ 3,
    COUNTRYC == "Taiwan, Province of China" ~ 3,
    COUNTRYC == "Australia" ~ 3,
    COUNTRYC == "Argentina" ~ 4,
    COUNTRYC == "Mexico" ~ 4,
    COUNTRYC == "Chile" ~ 4,
    TRUE ~ NA_real_))

#ANOVA
boxplot(MMSETOTS ~ region, data=DSData)
mmse.aov.ds <- aov(MMSETOTS ~ region, data = DSData)
summary(mmse.aov.ds)

#post hocs tests GT2
TukeyHSD(aov(MMSETOTS ~ as.factor(region), data=DSData))

#Regional comparisons of CBB, ISLT & CDR

boxplot(ISLTDRZ ~ region, data=DSData)
ISLTDR.aov.ds <- aov(ISLTDRZ ~ region, data = DSData)
summary(ISLTDR.aov.ds)
TukeyHSD(aov(ISLTDRZ ~ as.factor(region), data=DSData))

boxplot(ISLTTRZ ~ region, data=DSData)
ISLTTR.aov.ds <- aov(ISLTTRZ ~ region, data = DSData)
summary(ISLTTR.aov.ds)
TukeyHSD(aov(ISLTTRZ ~ as.factor(region), data=DSData))
```

```
boxplot(CDR0107S ~ region, data=DSData)
cdr.aov.ds <- aov(CDR0107S ~ region, data = DSData)
summary(cdr.aov.ds)
TukeyHSD(aov(CDR0107S ~ as.factor(region), data=DSData))

boxplot(FComp ~ region, data=DSData)
cbb.aov.ds <- aov(FComp ~ region, data = DSData)
summary(cbb.aov.ds)
TukeyHSD(aov(FComp ~ as.factor(region), data=DSData))

#Language Domain regional comparisons
boxplot(CComp ~ region, data=DSData)
crystal.aov.ds <- aov(CComp ~ region, data = DSData)
summary(crystal.aov.ds)
TukeyHSD(aov(CComp ~ as.factor(region), data=DSData))

#Remove South America
dsnors4 <- subset(dsnors3, Region != 4)

#Language correlation with years of education
ggscatter(DSData, x = "CComp", y = "EDUYRNUM",
        add = "reg.line", conf.int = TRUE,
        cor.coef = TRUE, cor.method = "pearson",
        xlab = "Crystallised Composite", ylab = "Years of Education")
res.cor <- cor.test(DSData$CComp, DSData$EDUYRNUM,
                method = "pearson")

#Crystalised composite zscored analysis
boxplot(CCompZ ~ region, data=DSData)
crystalz.aov.ds <- aov(CCompZ ~ region, data = DSData)
summary(crystalz.aov.ds)
TukeyHSD(aov(CCompZ ~ as.factor(region), data=DSData))

ggscatter(DSData, x = "CCompZ", y = "EDUYRNUM",
        add = "reg.line", conf.int = TRUE,
        cor.coef = TRUE, cor.method = "pearson",
        xlab = "Crystallised Composite Z-Score", ylab = "Years of Education")
res.cor <- cor.test(DSData$CCompZ, DSData$EDUYRNUM,
                method = "pearson")

#new ccomp yoe graph, pirateplot
ds10 <-subset(dsnors,EDUYRNUM <(40))
ds10 <- mutate(ds10, CCompZ = (CComp-7.55)/6.257920e-01)
format(round(ds10$CCompZ,4),nsmall=4)
pirateplot(formula = EDUYRNUM ~ CCompZ,
         data = ds10,
```

```
          pal="decision",
          point.o = .3,
          theme = 1,
          main = "Years of Education & Crystallised Composite Z-Score")

#descriptive stats for DS
stat.desc(dsn1$DS)
tblDS <- with(dsn1, table(DS))
barplot(tblDS)

#YOE & CCompZ
YOres <- cor.test(dsn1$EDUYRNUM, dsn1$CCompZ,
                  method = "pearson")

#correlations

#DR&DS
ggplot(dsnors3, aes(x=DS, y=ISLTDRZ)) +
  geom_point(color='#2980B9', size = 1) +
  geom_smooth(method=lm, color='#2C3E50')

DRres <- cor.test(dsnors3$DS, dsnors3$ISLTDRZ,
                  method = "pearson")

#IR&DS
ggplot(dsnors3, aes(x=DS, y=ISLTTRZ)) +
  geom_point(color='#2980B9', size = 1) +
  geom_smooth(method=lm, color='#2C3E50')
IRres <- cor.test(dsnors3$DS, dsnors3$ISLTTRZ,
                  method = "pearson")

ggplot(dsnors3, aes(x=CCompZ, y=FComp)) +
  geom_point(color='#2980B9', size = 1) +
  geom_smooth(method=lm, color='#2C3E50')
FCres <- cor.test(dsnors3$CCompZ, dsnors3$FComp,
                  method = "pearson")

#Final Analysis by Jutten CDR stages

#CDR Stage Assignment
dsn3 <-
  dsn2%>%
  mutate(
    CDRGroup = case_when(
      CDR0107S <=0.5 ~ 1,
      CDR0107S >=1 & CDR0107S<1.5 ~ 2,
```

```
    CDR0107S >=1.5 & CDR0107S<4.5 ~ 3,
    CDR0107S >=4.5 ~ 4,))


describeBy(dsn3$DS, group=dsn3$CDRGroup)
describeBy(dsn3$ISLTTRZ, group=dsn3$CDRGroup)
describeBy(dsn3$ISLTDRZ, group=dsn3$CDRGroup)


#CDR Staging correlations
ggplot(dsn3, aes(x=DS, y=ISLTDRZ))+
  geom_point(color='#2980B9', size = 1) +
  facet_wrap(~CDRGroup)+
  stat_cor(method = "pearson", label.x = -7.5, label.y = 2.1)+
  geom_smooth(method=lm, color='#2C3E50')

ggplot(dsn3, aes(x=DS, y=ISLTTRZ))+
  geom_point(color='#2980B9', size = 1) +
  facet_wrap(~CDRGroup)+
  stat_cor(method = "pearson", label.x = -7.5, label.y = 2.7)+
  geom_smooth(method=lm, color='#2C3E50')

plots.dir.path <- list.files(tempdir(), pattern="rs-graphics", full.names =
TRUE);
plots.png.paths <- list.files(plots.dir.path, pattern=".png", full.names =
TRUE)
file.copy(from=plots.png.paths, to="/Users/Tom/Documents/PhD/Chapter III -
Discrepancy Scores/Graphs/V2")

#ANOVA for DS by group
boxplot(DS ~ CDRGroup, data=DSData)
DS.aov.ds <- aov(DS ~ CDRGroup, data = DSData)
summary(DS.aov.ds)

#post hocs tests GT2
TukeyHSD(aov(DS ~ as.factor(CDRGroup), data=DSData))
```

Study 2

```
install.packages("githubinstall")
install.packages("devtools")
gh_install_packages("blandr")
library(blandr)

plot(Alternate_Paragraphs$APII, Alternate_Paragraphs$LMII,
    xlab="Alternate Story", ylab="Logical Memory",
    abline(lm(Alternate_Paragraphs$APII~Alternate_Paragraphs$LMII),
col="blue"))
cor.test(Alternate_Paragraphs$APII, Alternate_Paragraphs$LMII,
method="pearson")

ggscatter(Alternate_Paragraphs, x = "APII", y = "LMII",
        add = "reg.line", conf.int = TRUE,
        cor.coef = TRUE, cor.method = "pearson",
        xlab = "Alternate Paragraphs", ylab = "Logical Memory")

#testing for differences in correlations
cocor.result1 <- cocor.indep.groups(0.3865049, 0.6002357, 49, 67,
alternative = "two.sided",
                            test = "all", alpha = 0.05, conf.level = 0.95, null.value
= 0,
                            data.name = NULL, var.labels = NULL, return.htest =
FALSE)


#t-test
t.test(Alternate_Paragraphs$APII,Alternate_Paragraphs$LMII, paired=TRUE)

#B&A Stats
Alternate_Paragraphs <- transform(Alternate_Paragraphs,
                            average=((APII+LMII)/2),
                            difference=(APII-LMII)
                            )

blandr.output.text (Alternate_Paragraphs$APII, Alternate_Paragraphs$LMII ,
sig.level=0.95 )

blandr.draw(Alternate_Paragraphs$APII, Alternate_Paragraphs$LMII ,
ciDisplay = FALSE , ciShading = FALSE )

blandr.draw(Alternate_Paragraphs$APII, Alternate_Paragraphs$LMII)
```

```
#Split by group
split(Alternate_Paragraphs, as.factor(Alternate_Paragraphs$Group))

df1WI <- Alternate_Paragraphs[which(Alternate_Paragraphs$Group == 1),]
df2OI <- Alternate_Paragraphs[which(Alternate_Paragraphs$Group == 2),]



plot(df1WI$APII, df1WI$LMII,
   xlab="Alternate Story", ylab="Logical Memory",
   abline(lm(df1WI$APII~df1WI$LMII), col="blue"))
cor.test(df1WI$APII, df1WI$LMII, method="pearson")

plot(df2OI$APII, df2OI$LMII,
    xlab="Alternate Story", ylab="Logical Memory",
    abline(lm(df2OI$APII~df2OI$LMII), col="blue"))
cor.test(df2OI$APII, df2OI$LMII, method="pearson")

ggscatter(df1WI, x = "APII", y = "LMII",
        add = "reg.line", conf.int = TRUE,
        cor.coef = TRUE, cor.method = "pearson",
        xlab = "Alternate Paragraphs", ylab = "Logical Memory")

ggscatter(df2OI, x = "APII", y = "LMII",
        add = "reg.line", conf.int = TRUE,
        cor.coef = TRUE, cor.method = "pearson",
        xlab = "Alternate Paragraphs", ylab = "Logical Memory")

#t-test
t.test(df1WI$APII,df1WI$LMII, paired=TRUE)

t.test(df2OI$APII,df2OI$LMII, paired=TRUE)

#B&A Stats

blandr.output.text (df1WI$APII, df1WI$LMII , sig.level=0.95 )

blandr.draw(df1WI$APII, df1WI$LMII , ciDisplay = FALSE , ciShading = FALSE
)

blandr.draw(df1WI$APII, df1WI$LMII)


blandr.output.text (df2OI$APII, df2OI$LMII , sig.level=0.95 )

blandr.draw(df2OI$APII, df2OI$LMII , ciDisplay = FALSE , ciShading = FALSE )
```

```
blandr.draw(df2OI$APII, df2OI$LMII)
```

Study 3

```
install.packages("githubinstall")
install.packages("devtools")
gh_install_packages("blandr")
library(blandr)

#immediate recall comparison
MissionRandOnly <- mutate(MissionRandOnly, ADCIR = ADCRL*3)
round(MissionRandOnly$ADCIR, digits=0)

#descriptive stats
stat.desc(MissionRandOnly$ISLTDR)
stat.desc(MissionRandOnly$ISLTTR)
stat.desc(MissionRandOnly$ADCIR)
stat.desc(MissionRandOnly$ADCDRL)

hist(MissionRandOnly$ISLTDR)
hist(MissionRandOnly$ISLTTR)
hist(MissionRandOnly$ADCIR)
hist(MissionRandOnly$ADCDRL)

table(MissionRandOnly$ISLTTR,MissionRandOnly$ACIR)

#correlations

IRPlot <- ggplot(MissionRandOnly, aes(ISLTTR,ACIR))+
  geom_count()+
  scale_size_area()+
  geom_smooth(method="lm", se=TRUE)+
  xlab("ISLT Immediate Recall")+ylab("ADAS-Cog Immediate Recall")

IRPlot + stat_cor(method = "pearson", label.x = 3, label.y = 30)

DRPlot <- ggplot(MissionRandOnly, aes(ISLTDR,ACDR))+
  geom_count()+
  scale_size_area()+
  geom_smooth(method="lm", se=TRUE)+
  xlab("ISLT Delayed Recall")+ylab("ADAS-Cog Delayed Recall")+
  scale_x_continuous(breaks = c(0,2,4,6,8,10))+
  scale_y_continuous(breaks = c(0,2,4,6,8,10,12,14))
DRPlot + stat_cor(method = "pearson", label.x = 0.1, label.y = 12)

ggscatter(MissionRandOnly, x = "ISLTTR", y = "ACIR",
```

```
        add = "reg.line", conf.int = TRUE,
        cor.coef = TRUE, cor.method = "pearson",
        xlab = "ISLT", ylab = "ADAS-Cog")

ggscatter(MissionRandOnly, x = "ISLTDR", y = "ACDR",
        add = "reg.line", conf.int = TRUE,
        cor.coef = TRUE, cor.method = "pearson",
        xlab = "ISLT", ylab = "ADAS-Cog")

#B&A
#Same scale the variables
MissionRandOnly <- mutate(MissionRandOnly, ACIR = (((ADCIR/5)*6)-36)*-1)

MissionRandOnly <- mutate(MissionRandOnly, ACDR = (((ADCDRL/5)*6)-
12)*-1)

round(MissionRandOnly$ACIR, digits=1)

#Immediate Recall
MissionRandOnly <- transform(MissionRandOnly,
                        average=((ACIR+ISLTTR)/2),
                        difference=(ACIR-ISLTTR)
)
blandr.output.text (MissionRandOnly$ACIR, MissionRandOnly$ISLTTR ,
sig.level=0.95 )

blandr.draw(MissionRandOnly$ACIR, MissionRandOnly$ISLTTR , ciDisplay =
FALSE , ciShading = FALSE )

blandr.draw(MissionRandOnly$ACIR, MissionRandOnly$ISLTTR)

#Delayed Recall
MissionRandOnly <- transform(MissionRandOnly,
                        average=((ACDR+ISLTDR)/2),
                        difference=(ACDR-ISLTDR)
)

blandr.output.text (MissionRandOnly$ACDR, MissionRandOnly$ISLTDR ,
sig.level=0.95 )

blandr.draw(MissionRandOnly$ACDR, MissionRandOnly$ISLTDR , ciDisplay =
FALSE , ciShading = FALSE )

blandr.draw(MissionRandOnly$ACDR, MissionRandOnly$ISLTDR)

cocor.result1 <- cocor(~ACDR + ISLTDR | ACIR + ISLTTR,
```

```
                    MissionRandOnly)
as.htest(cocor.result)


#RM ANOVAS
library(tidyverse)
library(rstatix)
library(broom)

anova(lm(ACDR ~ APOE4ST * ISLTDR, data = MissionRandOnly))

anova(lm(ISLTDR ~ APOE4ST * ACDR, data = MissionRandOnly))

anova(lm(ACIR ~ APOE4ST * ISLTTR, data = MissionRandOnly))

anova(lm(ISLTTR ~ APOE4ST * ACIR, data = MissionRandOnly))

summary(manova(cbind(ISLTDR, ACDR) ~ APOE4ST, data =
MissionRandOnly))

summary(manova(cbind(ISLTTR, ACIR) ~ APOE4ST, data =
MissionRandOnly))

glm()
```

Study 4

```
#Import data #Limit to Tier 5 subjects

#Remove SA
  MALL <-
  MissionAllT5 %>%
  mutate(
    Region = case_when(
      COUNTRYC == "United States of America" ~ 1,
      COUNTRYC == "Canada" ~ 1,
      COUNTRYC == "United Kingdom of Great Britain and Northern Ireland" ~
2,
      COUNTRYC == "Spain" ~ 2,
      COUNTRYC == "France" ~ 2,
      COUNTRYC == "Germany" ~ 2,
      COUNTRYC == "Croatia" ~ 2,
      COUNTRYC == "Denmark" ~ 2,
      COUNTRYC == "Poland" ~ 2,
      COUNTRYC == "Czechia" ~ 2,
      COUNTRYC == "Slovakia" ~ 2,
      COUNTRYC == "Austria" ~ 2,
      COUNTRYC == "Italy" ~ 2,
      COUNTRYC == "Portugal" ~ 2,
      COUNTRYC == "South Africa" ~ 2,
      COUNTRYC == "Bulgaria" ~ 2,
      COUNTRYC == "Japan" ~ 3,
      COUNTRYC == "China" ~ 3,
      COUNTRYC == "Korea, Republic of" ~ 3,
      COUNTRYC == "Taiwan, Province of China" ~ 3,
      COUNTRYC == "Australia" ~ 3,
      COUNTRYC == "Argentina" ~ 4,
      COUNTRYC == "Mexico" ~ 4,
      COUNTRYC == "Chile" ~ 4,
      TRUE ~ NA_real_))
MALL <- subset(MALL, Region != 4)

#Assign value to APOE status
MALL <- mutate(MALL, APOEGroup = ifelse(APOE4ST=="Positive", 1, 0))

#Compute Discrepancy Scores
MALL <-subset(MALL, MMSETOTS >(17))
MALL <-subset(MALL,CBDETZ >(-6.077))
MALL <-subset(MALL,CBIDEZ >(-7.607))
MALL <-subset(MALL,CBOBMZ >(-6.231))
MALL <-subset(MALL,CBOCLZ >(-4.162))
```

```
MALL <- mutate(MALL, FComp = (CBDETZ + CBIDEZ + CBOBMZ +
CBOCLZ)/4)

#compute fluid composite (MMSE Language Domain #6,7,8,9,10)
MALL <- mutate(MALL, CComp = MMSE06S + MMSE07S + MMSE08S +
MMSE09S + MMSE10S)

#Memory Difference Score
MALL <- mutate(MALL, MD=ISLTTRZ-ISLTDRZ)

#Remove NAs from dataset for CComp & FComp
MALL <- subset(MALL, !is.na(CComp))
MALL <- subset(MALL, !is.na(CDR0107S))
MALL <- subset(MALL, !is.na(FComp))
MALL <- subset(MALL, !is.na(AmyTest))
MALL <- subset(MALL, !is.na(ISLTDRZ))
MALL <- subset(MALL, !is.na(ISLTTRZ))
MALL <- subset(MALL, !is.na(MMSETOTS))
MALL <- subset(MALL, !is.na(CBDETZ))
MALL <- subset(MALL, !is.na(CBOCLZ))
MALL <- subset(MALL, !is.na(CBOBMZ))
MALL <- subset(MALL, !is.na(CBIDEZ))

#CComp Z-Score

summary(MALL$CComp)
stat.desc(MALL$CComp)
MALL <- mutate(MALL, CCompZ = (CComp-7.55)/6.257920e-01)
MALL2 <-select(MALL, c('FComp', 'CCompZ', 'DS', 'AmyRes', 'MD'))

#Discrepancy Score
MALL <- mutate(MALL, DS = FComp - CCompZ)
MALL <- mutate(MALL, across(where(is.numeric), round, 3))

install.packages('e1071')
library(e1071)

#Run t-tests between key variables in both datasets

sapply(MALL,mean, na.rm=TRUE)
summary(MALL)
stat.desc(MALL)
describe(MALL$FComp)
describe(MALL$CCompZ)
describe(MALL$DS)
describe(MALL$MD)
```

```
describe.by(MALL, group="AmyRes")
describe.by(MALL2, group="AmyRes")

t.test(AGE ~ AmyRes, data = MALL)
t.test(SEX ~ AmyRes, data = MALL)
t.test(EDUYRNUM ~ AmyRes, data = MALL)
t.test(APOE4ST ~ AmyRes, data = MALL)
t.test(ISLTDRZ ~ AmyRes, data = MALL)
t.test(ISLTTRZ ~ AmyRes, data = MALL)
t.test(CBDETZ ~ AmyRes, data = MALL)
t.test(CBIDEZ ~ AmyRes, data = MALL)
t.test(CBOBMZ ~ AmyRes, data = MALL)
t.test(CBOCLZ ~ AmyRes, data = MALL)
t.test(CDR0107S ~ AmyRes, data = MALL)
t.test(MMSETOTS ~ AmyRes, data = MALL)
t.test(FComp ~ AmyRes, data = MALL)
t.test(CCompZ ~ AmyRes, data = MALL)
t.test(DS ~ AmyRes, data = MALL)
t.test(MD ~ AmyRes, data = MALL)

#SVM Model 1 ISLT, CBB, MMSE, CDR totals

#restrict dataset to required variables
M1<-select(MALL, c('ISLTDRZ', 'ISLTTRZ', 'CDR0107S', 'MMSETOTS',
'CBDETZ', 'CBOCLZ', 'CBOBMZ', 'CBIDEZ', 'AmyRes'))

#Split
dt = sort(sample(nrow(M1), nrow(M1)*.7))
M1train<-M1[dt,]
M1test<-M1[-dt,]

#SVM
#First go
svmfit <- svm(AmyRes~., data = M1train, kernel = "radial", cost = .1, scale
= FALSE, type='C-classification')
summary(svmfit)
print(svmfit, M1train)
tuned <- tune(svm, AmyRes~., data = M1train, kernel = "radial", ranges =
list(cost=c(0.001,0.01,.1,1,10,100)))
summary(tuned)
ypred=predict(svmfit ,M1test)
table(predict =ypred , truth= M1test$AmyRes )
ypred1=predict(svmfit ,M1train)
cm <- table(predict =ypred1 , truth= M1train$AmyRes )

#second go
```

```
set.seed(2)
ind <- sample(2, nrow(M1), replace = TRUE, prob=c(0.7, 0.3))
x.svm <- svm(AmyRes~., data = M1[ind == 1,], kernel = "radial",gamma=4,
cost = .1, scale = FALSE, type='C-classification', probability=TRUE)
summary(x.svm)
tuned <- tune(svm, AmyRes~., data = M1train, kernel = "radial", ranges =
list(gamma=2^(-8:1), cost=c(0.001,0.01,.1,1,10,100)))
summary(tuned)
x.svm <- svm(AmyRes~., data = M1[ind == 1,], kernel = "radial", cost =
0.1,scale = FALSE, type='C-classification', probability=TRUE)
x.svm.prob <- predict(x.svm, type="prob", newdata=M1[ind == 2,],
probability = TRUE)
x.svm.prob.rocr <- prediction(attr(x.svm.prob, "probabilities")[,2], M1[ind ==
2,'AmyRes'])
x.svm.perf <- performance(x.svm.prob.rocr, "tpr","fpr")
x.svm.perf2 <- performance(x.svm.prob.rocr, measure="auc")
acc.perf = performance(x.svm.prob.rocr, measure = "acc")
ind = which.max( slot(acc.perf, "y.values")[[1]] )
acc = slot(acc.perf, "y.values")[[1]][ind]
cutoff = slot(acc.perf, "x.values")[[1]][ind]
print(c(accuracy= acc, cutoff = cutoff))
svm1.plot <-plot(x.svm.perf, col="red", xlab="1-Specificity",
ylab="Sensitivity", abline(a = 0, b = 1), grid(col = "lightgray", lty = "dotted",
lwd = par("lwd"), equilogs = TRUE))
par(pty="s")




##########################################
#Model 2

M2<-select(MALL, c('DS', 'MD','MMSETOTS', 'CDR0107S', 'AmyRes'))
M2<-na.omit(M2)
set.seed(2)
ind <- sample(2, nrow(M2), replace = TRUE, prob=c(0.7, 0.3))
x.svm2 <- svm(AmyRes~., data = M2[ind == 1,], kernel = "radial", cost =
.1, scale = FALSE, type='C-classification', probability=TRUE)
summary(x.svm2)
tuned2 <- tune(svm, AmyRes~., data = M2[ind == 1,], kernel = "radial",
ranges = list(gamma=2^(-8:1), cost=c(0.001,0.01,.1,1,10,100)))
summary(tuned2)

x.svm2 <- svm(AmyRes~., data = M2[ind == 1,], kernel = "radial",
gamma=0.078125, cost = 1,scale = FALSE, type='C-classification',
probability=TRUE)
```

```
ypred22=predict(x.svm2 ,M2[ind == 1,])
table(predict= ypred22 , truth= M2[ind == 1,'AmyRes'] )

x.svm2 <- svm(AmyRes~., data = M2[ind == 2,], kernel = "radial",
gamma=0.078125, cost = 1,scale = FALSE, type='C-classification',
probability=TRUE)
ypred2=predict(x.svm2 ,M2[ind == 2,])
cm2<-table(predict=ypred2,truth= M2[ind == 2,'AmyRes'] )


x.svm2.prob <- predict(x.svm2, type="prob", newdata=M2[ind == 2,],
probability = TRUE)
x.svm2.prob.rocr <- prediction(attr(x.svm2.prob, "probabilities")[,2], M2[ind
== 2,'AmyRes'])
x.svm2.perf <- performance(x.svm2.prob.rocr, "tpr","fpr")
x.svm2.perf2 <- performance(x.svm2.prob.rocr, measure="auc")
acc.perf = performance(x.svm2.prob.rocr, measure = "acc")
ind = which.max(slot(acc.perf2, "y.values")[[1]])
acc = slot(acc.perf, "y.values")[[1]][ind]
cutoff = slot(acc.perf, "x.values")[[1]][ind]
print(c(accuracy= acc, cutoff = cutoff))



#####################################
#Model 3

M3<-select(MALL, c('AGE', 'SEX', 'EDUYRNUM', 'APOEGroup', 'ISLTDRZ',
'ISLTTRZ', 'CDR0107S', 'MMSETOTS', 'CBDETZ', 'CBOCLZ', 'CBOBMZ',
'CBIDEZ', 'AmyRes'))
M3<-na.omit(M3)
set.seed(2)
ind <- sample(2, nrow(M3), replace = TRUE, prob=c(0.7, 0.3))
x.svm3 <- svm(AmyRes~., data = M3[ind == 1,], kernel = "radial", cost =
.1, scale = FALSE, type='C-classification', probability=TRUE)
summary(x.svm3)
tuned <- tune(svm, AmyRes~., data = M3[ind == 1,], kernel = "radial",
ranges = list(gamma=2^(-8:1), cost=c(0.001,0.01,.1,1,10,100)))
summary(tuned)
x.svm3 <- svm(AmyRes~., data = M3[ind == 1,], kernel = "radial", cost =
0.1,gamma=0.00390625, scale = FALSE, type='C-classification',
probability=TRUE)

x.svm3.prob <- predict(x.svm3, type="prob", newdata=M3[ind == 2,],
probability = TRUE)
```

```
x.svm3.prob.rocr <- prediction(attr(x.svm3.prob, "probabilities")[,2], M3[ind
== 2,'AmyRes'])
x.svm3.perf <- performance(x.svm3.prob.rocr, "tpr","fpr")
x.svm3.perf2 <- performance(x.svm3.prob.rocr, measure="auc")
acc.perf = performance(x.svm3.prob.rocr, measure = "acc")
ind = which.max(slot(acc.perf, "y.values")[[1]] )
acc = slot(acc.perf, "y.values")[[1]][ind]
cutoff = slot(acc.perf, "x.values")[[1]][ind]
print(c(accuracy= acc, cutoff = cutoff))

dt3 = sort(sample(nrow(M3), nrow(M3)*.7))

M3train<-M3[dt,]
M3test<-M3[-dt,]
ypred33=predict(x.svm3 ,M3train)
table(predict= ypred22 , truth= M3train$AmyRes)
x.svm3 <- svm(AmyRes~., data = M3test, kernel = "radial", cost =
0.1,gamma=0.00390625, scale = FALSE, type='C-classification',
probability=TRUE)
ypred3=predict(x.svm3 ,M3test)
table(predict=ypred3,trut= M3test$AmyRes)



#AUC PLOT

plot(x.svm.perf, col=1, main="ROC curves of different machine learning
classifier", abline(a = 0, b = 1))
plot(x.svm2.perf, col=2, add=TRUE)
plot(x.svm3.perf, col=3, add=TRUE)

svm1.plot <-plot(x.svm.perf, col=1, abline(a = 0, b = 1), grid(col =
"lightgray", lty = "dotted", lwd = par("lwd"), equilogs = TRUE))
svm1.plot <-plot(x.svm.perf2, col=6, add=TRUE)
svm1.plot <-plot(x.svm.perf, col=8, add=TRUE)

svm.all.plot <-plot(x.svm.perf, col=1, main="ROC curves of different
machine learning classifier", abline(a = 0, b = 1))
svm.all.plot <-plot(x.svm2.perf, col=2, add=TRUE)
svm.all.plot <-plot(x.svm3.perf, col=3, add=TRUE)
```

Study 4 Additional Console Output

Model 1 – ‘ISLTDRZ', 'ISLTTRZ', 'CDR0107S', 'MMSETOTS', 'CBDETZ', 'CBOCLZ', 'CBOBMZ', 'CBIDEZ', 'AmyRes'

```
Call:
svm(formula = AmyRes ~ ., data = M1train, kernel = "radial", cost = 0.01,
type = "C-classification",
    scale = FALSE)

Parameters:
   SVM-Type:  C-classification
 SVM-Kernel:  radial
       cost:  0.01
Number of Support Vectors:  2276
 ( 1131 1145 )

Number of Classes:  2
Levels:
 0 1
Parameter tuning of ‘svm’:
- sampling method: 10-fold cross validation
- best parameters:
 cost
  0.1

- best performance: 0.2378208

- Detailed performance results:
        gamma  cost     error dispersion
1  0.00390625 1e-03 0.4036277 0.02207063
2  0.00781250 1e-03 0.4034516 0.02208725
3  0.01562500 1e-03 0.4031274 0.02211346
4  0.03125000 1e-03 0.4023496 0.02217296
5  0.06250000 1e-03 0.4010796 0.02220646
6  0.12500000 1e-03 0.4000306 0.02221270
7  0.25000000 1e-03 0.4000939 0.02214300
8  0.50000000 1e-03 0.4010707 0.02210481
9  1.00000000 1e-03 0.4022756 0.02207500
10 2.00000000 1e-03 0.4030790 0.02205516
11 0.00390625 1e-02 0.4023687 0.02220801
12 0.00781250 1e-02 0.4006561 0.02234014
13 0.01562500 1e-02 0.3969513 0.02267287
```

```
14 0.03125000 1e-02 0.3889418 0.02293331
15 0.06250000 1e-02 0.3767756 0.02333949
16 0.12500000 1e-02 0.3675339 0.02343957
17 0.25000000 1e-02 0.3685054 0.02271030
18 0.50000000 1e-02 0.3767315 0.02227221
19 1.00000000 1e-02 0.3874627 0.02201744
20 2.00000000 1e-02 0.3948561 0.02193160

Call:
svm(formula = AmyRes ~ ., data = M1train, kernel = "radial", cost = 0.1,
type = "C-classification",
    scale = FALSE)
Parameters:
   SVM-Type:  C-classification
 SVM-Kernel:  radial
       cost:  0.1
Number of Support Vectors:  2154
 ( 1065 1089 )

Number of Classes:  2
Levels:
 0 1

TRAINING DATASET OUTCOME
> ypred1=predict(svmfit ,M1train)
> table(predict =ypred1 , truth= M1train$AmyRes )
       truth
predict    0    1
      0  564  293
      1  567 1148

TEST DATASET OUTCOME
> ypred=predict(svmfit ,M1test)
> table(predict =ypred , truth= M1test$AmyRes )
       truth
predict   0   1
      0 230 136
      1 258 479
> acc.perf = performance(x.svm.prob.rocr, measure = "acc")
> ind = which.max( slot(acc.perf, "y.values")[[1]] )
> acc = slot(acc.perf, "y.values")[[1]][ind]
> cutoff = slot(acc.perf, "x.values")[[1]][ind]
> print(c(accuracy= acc, cutoff = cutoff))
accuracy cutoff.224
 0.6452489  0.4755230
```

Model 2 – 'DS', 'MD','MMSETOTS', 'CDR0107S', 'AmyRes'

```
Call:
svm(formula = AmyRes ~ ., data = M2[ind == 1, ], kernel = "radial", cost =
0.1,
    type = "C-classification", probability = TRUE, scale = FALSE)


Parameters:
   SVM-Type:  C-classification
 SVM-Kernel:  radial
       cost:  0.1

Number of Support Vectors:  2278

 ( 1130 1148 )


Number of Classes:  2

Levels:
 0 1

Parameter tuning of ‘svm’:

- sampling method: 10-fold cross validation

- best parameters:
     gamma cost
 0.0078125    1

- best performance: 0.2869819

- Detailed performance results:
        gamma  cost     error dispersion
1  0.00390625 1e-03 0.4036277 0.02207063
2  0.00781250 1e-03 0.4034516 0.02208725
3  0.01562500 1e-03 0.4031274 0.02211346
4  0.03125000 1e-03 0.4023496 0.02217296
5  0.06250000 1e-03 0.4010796 0.02220646
6  0.12500000 1e-03 0.4000306 0.02221270
7  0.25000000 1e-03 0.4000939 0.02214300
8  0.50000000 1e-03 0.4010707 0.02210481
9  1.00000000 1e-03 0.4022756 0.02207500
10 2.00000000 1e-03 0.4030790 0.02205516
11 0.00390625 1e-02 0.4023687 0.02220801
```

```
12 0.00781250 1e-02 0.4006561 0.02234014
13 0.01562500 1e-02 0.3969513 0.02267287
14 0.03125000 1e-02 0.3889418 0.02293331
15 0.06250000 1e-02 0.3767756 0.02333949
16 0.12500000 1e-02 0.3675339 0.02343957
17 0.25000000 1e-02 0.3685054 0.02271030
18 0.50000000 1e-02 0.3767315 0.02227221
19 1.00000000 1e-02 0.3874627 0.02201744
20 2.00000000 1e-02 0.3948561 0.02193160

Call:
svm(formula = AmyRes ~ ., data = M2[ind == 1, ], kernel = "radial", gamma
= 0.078125,
    cost = 1, type = "C-classification", probability = TRUE, scale = FALSE)

Parameters:
   SVM-Type:  C-classification
 SVM-Kernel:  radial
       cost:  1

Number of Support Vectors:  2193

 ( 1093 1100 )
Number of Classes:  2
Levels:
 0 1

table(predict= ypred22 , truth= M2[ind == 1,'AmyRes'] )
     truth
predict    0    1
      0  393  303
      1  741 1135

table(predict=ypred2,truth= M2[ind == 2,'AmyRes'] )
  truth
predict   0   1
      0 150 103
      1 335 515
ind = which.max(slot(acc.perf2, "y.values")[[1]])
> acc = slot(acc.perf, "y.values")[[1]][ind]
> cutoff = slot(acc.perf, "x.values")[[1]][ind]
> print(c(accuracy= acc, cutoff = cutoff))
  accuracy cutoff.728
 0.5918552  0.3961233
```

Model 3 – ‘AGE’, 'SEX', ‘EDUYRNUM’, ‘APOEGroup’'ISLTDRZ', 'ISLTTRZ', 'CDR0107S', 'MMSETOTS', 'CBDETZ', 'CBOCLZ', 'CBOBMZ', 'CBIDEZ', 'MMSETOTS', 'CDR0107S', 'AmyRes'

```
Call:
svm(formula = AmyRes ~ ., data = M3[ind == 1, ], kernel = "radial", cost = 0.1,
    type = "C-classification", probability = TRUE, scale = FALSE)


Parameters:
   SVM-Type:  C-classification
 SVM-Kernel:  radial
       cost:  0.1

Number of Support Vectors:  2245

 ( 1096 1149 )


Number of Classes:  2

Levels:
 0 1
Parameter tuning of ‘svm’:

- sampling method: 10-fold cross validation

- best parameters:
      gamma cost
 0.00390625  0.1

- best performance: 0.1768172

- Detailed performance results:
       gamma  cost     error  dispersion
1  0.00390625 1e-03 0.3973412 0.020308004
2  0.00781250 1e-03 0.3923247 0.020228557
3  0.01562500 1e-03 0.3850083 0.020132793
4  0.03125000 1e-03 0.3773245 0.019986676
5  0.06250000 1e-03 0.3756828 0.020005291
6  0.12500000 1e-03 0.3847868 0.020297802
7  0.25000000 1e-03 0.3967004 0.020494503
8  0.50000000 1e-03 0.4020537 0.020458420
9  1.00000000 1e-03 0.4031687 0.020443919
10 2.00000000 1e-03 0.4032916 0.020441508
```

```
11 0.00390625 1e-02 0.3460287 0.018953802
12 0.00781250 1e-02 0.3049806 0.017434177
13 0.01562500 1e-02 0.2545752 0.014662003
14 0.03125000 1e-02 0.2146571 0.010624657
15 0.06250000 1e-02 0.2083003 0.010334287
16 0.12500000 1e-02 0.2539961 0.015554849
17 0.25000000 1e-02 0.3409806 0.020317116
18 0.50000000 1e-02 0.3895341 0.020524415
19 1.00000000 1e-02 0.3993143 0.020316706
20 2.00000000 1e-02 0.4002759 0.020303776


accuracy cutoff.684
 0.6623256  0.6196396

train
    truth
predict    0    1
     0  300  369
    1   807 1015


Test
   truth
predict   0   1
     0 150  67
     1 304 544
```

Appendix IV - References

A

Abbott, R. A., Skirrow, C., Jokisch, M., Timmers, M., Streffer, J., van Nueten, L., ... & Weimar, C. (2019). Normative data from linear and nonlinear quantile regression in CANTAB: Cognition in mid-to-late life in an epidemiological sample. *Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring*, *11*, 36-44.

Ahmed, R. R., Holler, C. J., Webb, R. L., Li, F., Beckett, T. L., & Murphy, M. P. (2010). BACE1 and BACE2 enzymatic activities in Alzheimer's disease. *Journal of neurochemistry*, *112*(4), 1045-1053.

Ahmad, T. *et al.* (2018) "Machine Learning Methods Improve Prognostication, Identify Clinically Distinct Phenotypes, and Detect Heterogeneity in Response to Therapy in a Large Cohort of Heart Failure Patients," *Journal of the American Heart Association*, 7(8).

Aisen, P. S. (2019). IV immunoglobulin for Alzheimer disease: Implications for immunotherapy strategies.

Albert, M. S. (2011). Changes in cognition. *Neurobiology of aging*, *32*, S58-S63.

Albert, M. S., DeKosky, S. T., Dickson, D., Dubois, B., Feldman, H. H., Fox, N. C., ... & Phelps, C. H. (2011). The diagnosis of mild cognitive impairment due to 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), 270-279.

Alonso, A. D. C., Zaidi, T., Grundke-Iqbal, I., & Iqbal, K. (1994). Role of abnormally phosphorylated tau in the breakdown of microtubules in Alzheimer disease. *Proceedings of the National Academy of Sciences*, *91*(12), 5562-5566.

Alonso, A. D. C., Grundke-Iqbal, I., & Iqbal, K. (1996). Alzheimer's disease hyperphosphorylated tau sequesters normal tau into tangles of filaments and disassembles microtubules. *Nature medicine*, *2*(7), 783.

Alonso, A. D. C., Grundke-Iqbal, I., Barra, H. S., & Iqbal, K. (1997). Abnormal phosphorylation of tau and the mechanism of Alzheimer neurofibrillary degeneration: sequestration of microtubule-associated proteins 1 and 2 and the disassembly of microtubules by the abnormal tau. *Proceedings of the National Academy of Sciences*, *94*(1), 298-303.

Alonso, A. D. C., Mederlyova, A., Novak, M., Grundke-Iqbal, I., & Iqbal, K. (2004). Promotion of hyperphosphorylation by frontotemporal dementia tau mutations. *Journal of Biological Chemistry*, *279*(33), 34873-34881.

Altman, D. G., & Bland, J. M. (1994). Diagnostic tests. 1: Sensitivity and specificity. *BMJ: British Medical Journal*, *308*(6943), 1552.

Alzheimer's Association. (2019). https://www.alz.org/media/Documents/alzheimers-facts-and-figures-2019-r.pdf

Amariglio, R. E., Becker, J. A., Carmasin, J., Wadsworth, L. P., Lorius, N., Sullivan, C., ... & Johnson, K. A. (2012). Subjective cognitive complaints and amyloid burden in cognitively normal older individuals. *Neuropsychologia*, *50*(12), 2880-2886.

American Psychiatric Association (APA). (2000). *Diagnostic and statistical manual of mental disorders* (4th ed., text rev.).

Amieva, H., Jacqmin-Gadda, H., Orgogozo, J. M., Le Carret, N., Helmer, C., Letenneur, L., ... & Dartigues, J. F. (2005). The 9 year cognitive decline before dementia of the Alzheimer type: a prospective population-based study. *Brain*, *128*(5), 1093-1101.

Amieva, H., Le Goff, M., Millet, X., Orgogozo, J. M., Pérès, K., Barberger-Gateau, P., ... & Dartigues, J. F. (2008). Prodromal Alzheimer's disease: successive emergence of the clinical symptoms. *Annals of Neurology: Official Journal of the American Neurological Association and the Child Neurology Society*, *64*(5), 492-498.

Andersson, C., Blennow, K., Almkvist, O., Andreasen, N., Engfeldt, P., Johansson, S. E., ... & Eriksdotter-Jönhagen, M. (2008). Increasing CSF phospho-tau levels during cognitive decline and progression to dementia. *Neurobiology of aging*, *29*(10), 1466-1473.

Andreasen, N., Vanmechelen, E., Van de Voorde, A., Davidsson, P., Hesse, C., Tarvonen, S., ... & Blennow, K. (1998). Cerebrospinal fluid tau protein as a biochemical marker for Alzheimer's disease: a community based follow up study. *Journal of Neurology, Neurosurgery & Psychiatry*, *64*(3), 298-305.

Andreasson, U., Vanmechelen, E., Shaw, L. M., Zetterberg, H., & Vanderstichele, H. (2012). Analytical aspects of molecular Alzheimer's disease biomarkers. *Biomarkers in medicine*, *6*(4), 377-389.

Andrews, J. S., Desai, U., Kirson, N. Y., Zichlin, M. L., Ball, D. E., & Matthews, B. R. (2019). Disease severity and minimal clinically important differences in clinical outcome assessments for Alzheimer's disease clinical trials. *Alzheimer's & Dementia: Translational Research & Clinical Interventions*, *5*(1), 354-363.

Arevalo-Rodriguez, I., Pedraza, O. L., Rodríguez, A., Sánchez, E., Gich, I., Solà, I., ... & Alonso-Coello, P. (2013). Alzheimer's disease dementia guidelines for diagnostic testing: a systematic review. *American Journal of Alzheimer's Disease & Other Dementias®*, *28*(2), 111-119.

Arevalo-Rodriguez, I., Smailagic, N., i Figuls, M. R., Ciapponi, A., Sanchez-Perez, E., Giannakou, A., ... & Cullum, S. (2015). Mini-Mental State Examination (MMSE) for the detection of Alzheimer's disease and other dementias in people with mild cognitive impairment (MCI). *Cochrane Database of Systematic Reviews*, (3).

Ashford, J. W., Kolm, P., Colliver, J. A., Bekian, C., & Hsu, L. N. (1989). Alzheimer patient evaluation and the mini-mental state: item characteristic curve analysis. *Journal of Gerontology*, *44*(5), P139-P146.

B

Bäckman, L., Jones, S., Berger, A. K., Laukka, E. J., & Small, B. J. (2005). Cognitive impairment in preclinical Alzheimer's disease: a meta-analysis. *Neuropsychology*, *19*(4), 520.

Bacskai, B. J., Frosch, M. P., Freeman, S. H., Raymond, S. B., Augustinack, J. C., Johnson, K. A., ... & Greenberg, S. M. (2007). Molecular imaging with Pittsburgh Compound B confirmed at autopsy: a case report. *Archives of neurology*, *64*(3), 431-434.

Baker, J. E., Lim, Y. Y., Pietrzak, R. H., Hassenstab, J., Snyder, P. J., Masters, C. L., & Maruff, P. (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*, 108-121.

Bakota, L., & Brandt, R. (2016). Tau biology and tau-directed therapies for Alzheimer's disease. *Drugs*, *76*(3), 301-313.

Baltes, P. B., Cornelius, S. W., Spiro, A., Nesselroade, J. R., & Willis, S. L. (1980). Integration versus differentiation of fluid/crytallized intelligence in old age. *Developmental Psychology*, *16*(6), 625.

Bancher, C., Brunner, C., Lassmann, H., Budka, H., Jellinger, K., Seitelberger, F., ... & Wisniewski, H. M. (1989). Tau and ubiquitin immunoreactivity at

different stages of formation of Alzheimer neurofibrillary tangles. *Progress in clinical and biological research*, *317*, 837-848.

Bansal, D., Chhikara, R., Khanna, K., & Gupta, P. (2018). Comparative analysis of various machine learning algorithms for detecting dementia. *Procedia computer science*, *132*, 1497-1502.

Barnes, D. E., Yaffe, K., Satariano, W. A., & Tager, I. B. (2003). A longitudinal study of cardiorespiratory fitness and cognitive function in healthy older adults. *Journal of the American Geriatrics Society*, *51*(4), 459-465.

Barthel, H., Gertz, H. J., Dresel, S., Peters, O., Bartenstein, P., Buerger, K., ... & Sabri, O. (2011). Cerebral amyloid-β PET with florbetaben (18F) in patients with Alzheimer's disease and healthy controls: a multicentre phase 2 diagnostic study. *The Lancet Neurology*, *10*(5), 424-435.

Barthélemy, N. R., Li, Y., Joseph-Mathurin, N., Gordon, B. A., Hassenstab, J., Benzinger, T., ... & McDade, E. (2020). A soluble phosphorylated tau signature links tau, amyloid and the evolution of stages of dominantly inherited Alzheimer’s disease. *Nature medicine*, *26*(3), 398-407.

Basaia, S., Agosta, F., Wagner, L., Canu, E., Magnani, G., Santangelo, R., ... & Alzheimer's Disease Neuroimaging Initiative. (2019). Automated classification of Alzheimer's disease and mild cognitive impairment using a single MRI and deep neural networks. *NeuroImage: Clinical*, *21*, 101645.

Bastin, C., & Salmon, E. (2014). Early neuropsychological detection of Alzheimer's disease. *European journal of clinical nutrition*, *68*(11), 1192-1199.

Bateman, R. J., Xiong, C., Benzinger, T. L., Fagan, A. M., Goate, A., Fox, N. C., ... & Holtzman, D. M. (2012). Clinical and biomarker changes in dominantly inherited Alzheimer's disease. *New England Journal of Medicine*, *367*(9), 795-804.

Bateman, R. J., Benzinger, T. L., Berry, S., Clifford, D. B., Duggan, C., Fagan, A. M., ... & Mills, S. (2017). The DIAN-TU Next Generation Alzheimer's prevention trial: adaptive design and disease progression model. *Alzheimer's & Dementia*, *13*(1), 8-19.

Baudier, J., Deloulme, J. C., Van Dorsselaer, A., Black, D., & Matthes, H. W. (1991). Purification and characterization of a brain-specific protein kinase C substrate, neurogranin (p17). Identification of a consensus amino acid sequence between neurogranin and neuromodulin (GAP43) that corresponds to the protein kinase C phosphorylation site and the calmodulin-binding domain. *Journal of Biological Chemistry*, *266*(1), 229-237.

Beeri, M. S., Schmeidler, J., Sano, M., Wang, J., Lally, R., Grossman, H., & Silverman, J. M. (2006). Age, gender, and education norms on the CERAD neuropsychological battery in the oldest old. *Neurology*, *67*(6), 1006-1010.

Begum, N. A., Ishii, K., Kurita-Taniguchi, M., Tanabe, M., Kobayashi, M., Moriwaki, Y., ... & Seya, T. (2004). Mycobacterium bovis BCG cell wall-specific differentially expressed genes identified by differential display and cDNA subtraction in human macrophages. *Infection and immunity*, *72*(2), 937-948.

Beitzke, M., Enzinger, C., Wünsch, G., Asslaber, M., Gattringer, T., & Fazekas, F. (2015). Contribution of convexal subarachnoid hemorrhage to disease progression in cerebral amyloid angiopathy. *Stroke*, *46*(6), 1533-1540.

Bekris, L. M., Khrestian, M., Dyne, E., Shao, Y., Pillai, J. A., Rao, S. M., ... & Leverenz, J. B. (2018). Soluble TREM2 and biomarkers of central and peripheral inflammation in neurodegenerative disease. *Journal of neuroimmunology*, *319*, 19-27.

Benilova, I., Karran, E., & De Strooper, B. (2012). The toxic Aβ oligomer and Alzheimer's disease: an emperor in need of clothes. *Nature neuroscience*, *15*(3), 349.

Bennett, D. A., Wilson, R. S., Schneider, J. A., Evans, D. A., Aggarwal, N. T., Arnold, S. E., ... & Bienias, J. L. (2003). Apolipoprotein E ε4 allele, AD pathology, and the clinical expression of Alzheimer’s disease. *Neurology*, *60*(2), 246-252.

Berry, D. A., & Stangl, D. K. (1996). Bayesian methods in health-related research. *Bayesian biostatistics*, 3-66.

Bertram, L., & Tanzi, R. E. (2008). Thirty years of Alzheimer's disease genetics: the implications of systematic meta-analyses. *Nature Reviews Neuroscience*, *9*(10), 768.

Besson, J. A. O., Crawford, J. R., Parker, D. M., Ebmeier, K. P., Best, P. V., Gemmell, H. G., ... & Smith, F. W. (1990). Multimodal imaging in Alzheimer's disease: the relationship between MRI, SPECT, cognitive and pathological changes. *The British Journal of Psychiatry*, *157*(2), 216-220.

Besson, F. L., La Joie, R., Doeuvre, L., Gaubert, M., Mézenge, F., Egret, S., ... & de La Sayette, V. (2015). Cognitive and brain profiles associated with current neuroimaging biomarkers of preclinical Alzheimer's disease. *Journal of Neuroscience*, *35*(29), 10402-10411.

BERTONI-FREDDARI, C. A. R. L. O., Fattoretti, P., Paoloni, R., Caselli, U., & MEIER-RUGE, W. I. L. L. I. A. M. (1997). Impaired dynamic morphology of cerebellar mitochondria in physiological aging and Alzheimer's disease. *Annals of the New York Academy of Sciences*, *826*(1), 479-482.

Bertram L, Lange C, Mullin K, Parkinson M, Hsiao M, Hogan MF, Schjeide BM, Hooli B, Divito J, Ionita I *et al* (2008) Genome-wide association analysis reveals putative Alzheimer's disease susceptibility loci in addition to APOE. Am J Hum Genet 83: 623–632

Beyreuther K, Masters CL (1991) Amyloid precursor protein (APP) and beta A4 amyloid in the etiology of Alzheimer's disease: precursor-product relationships in the derangement of neuronal function. Brain Pathol 1: 241–251

Birks, J. S. (2006). Cholinesterase inhibitors for Alzheimer's disease. *Cochrane database of systematic reviews*, (1).

Birks, J., & Flicker, L. (2006). Donepezil for mild cognitive impairment. *Cochrane Database of Systematic Reviews*, (3).

Birks, J. S., & Harvey, R. J. (2018). Donepezil for dementia due to Alzheimer's disease. *Cochrane Database of systematic reviews*, (6).

Bitan, G., Kirkitadze, M. D., Lomakin, A., Vollers, S. S., Benedek, G. B., & Teplow, D. B. (2003). Amyloid β-protein (Aβ) assembly: Aβ40 and Aβ42 oligomerize through distinct pathways. *Proceedings of the National Academy of Sciences*, *100*(1), 330-335.

Bjerke, M., Portelius, E., Minthon, L., Wallin, A., Anckarsäter, H., Anckarsäter, R., ... & Blennow, K. (2010). Confounding factors influencing amyloid beta concentration in cerebrospinal fluid. *International journal of Alzheimer's disease*, *2010*.

Blair, M., Kertesz, A., Davis-Faroque, N., Hsiung, G. Y., Black, S. E., Bouchard, R. W., ... & Feldman, H. (2007). Behavioural measures in frontotemporal lobar dementia and other dementias: the utility of the frontal behavioural inventory and the neuropsychiatric inventory in a national cohort study. *Dementia and geriatric cognitive disorders*, *23*(6), 406-415.

Blair, J. R., & Spreen, O. (1989). Predicting premorbid IQ: a revision of the National Adult Reading Test. *The Clinical Neuropsychologist*, *3*(2), 129-136.

Bland JM, Altman DG. (1986). Statistical methods for assessing agreement between two methods of clinical measurement. Lancet. i: 307-10

Bleecker, M. L., Bolla-Wilson, K., Kawas, C., & Agnew, J. (1988). Age-specific norms for the mini-mental state exam. *Neurology*, *38*(10), 1565-1565.

Blennow, K., Zetterberg, H., Minthon, L., Lannfelt, L., Strid, S., Annas, P., ... & Andreasen, N. (2007). Longitudinal stability of CSF biomarkers in Alzheimer's disease. *Neuroscience letters*, *419*(1), 18-22.

Blennow, K. (2010). Biomarkers in Alzheimer's disease drug development. *Nature medicine*, *16*(11), 1218.

Blennow, K., Zetterberg, H., & Fagan, A. M. (2012). Fluid biomarkers in Alzheimer disease. *Cold Spring Harbor perspectives in medicine*, *2*(9), a006221.

Bloudek, L. M., Spackman, D. E., Blankenburg, M., & Sullivan, S. D. (2011). Review and meta-analysis of biomarkers and diagnostic imaging in Alzheimer's disease. *Journal of Alzheimer's Disease*, *26*(4), 627-645.

Bogdanovic, N., Davidsson, P., Gottfries, J., Volkman, I., Winblad, B., & Blennow, K. (2002). Regional and cellular distribution of synaptic proteins in the normal human brain. *Brain Aging*, *2*, 18-30.

Bolognani, S. A. P., Miranda, M. C., Martins, M., Rzezak, P., Bueno, O. F. A., Camargo, C. H. P. D., & Pompeia, S. (2015). Development of alternative versions of the Logical Memory subtest of the WMS-R for use in Brazil. *Dementia & neuropsychologia*, *9*, 136-148.

Boraxbekk, C. J., Ames, D., Kochan, N. A., Lee, T., Thalamuthu, A., Wen, W., ... & Mather, K. A. (2015). Investigating the influence of KIBRA and CLSTN2 genetic polymorphisms on cross-sectional and longitudinal measures of memory performance and hippocampal volume in older individuals. *Neuropsychologia*, *78*, 10-17.

Braak, H., & Braak, E. (1991). Neuropathological stageing of Alzheimer-related changes. *Acta neuropathologica*, *82*(4), 239-259.

Braak, H., Alafuzoff, I., Arzberger, T., Kretzschmar, H., & Del Tredici, K. (2006). Staging of Alzheimer disease-associated neurofibrillary pathology using paraffin sections and immunocytochemistry. *Acta neuropathologica*, *112*(4), 389-404.

Brandt, J. (1991). The Hopkins Verbal Learning Test: Development of a new memory test with six equivalent forms. *The clinical neuropsychologist*, *5*(2), 125-142.

Bransby, L., Lim, Y. Y., Ames, D., Fowler, C., Roberston, J., Harrington, K., ... & Maruff, P. (2019). Sensitivity of a Preclinical Alzheimer's Cognitive Composite (PACC) to amyloid β load in preclinical Alzheimer's disease. *Journal of Clinical and Experimental Neuropsychology*, 1-10.

Brayne, C., & Calloway, P. (1990). The association of education and socioeconomic status with the Mini Mental State Examination and the clinical diagnosis of dementia in elderly people. *Age and ageing*, *19*(2), 91-96.

Brier, M. R., Gordon, B., Friedrichsen, K., McCarthy, J., Stern, A., Christensen, J., ... & Cairns, N. J. (2016). Tau and Aβ imaging, CSF measures, and cognition in Alzheimer's disease. *Science translational medicine*, *8*(338), 338ra66-338ra66.

Brown, G. G., Rahill, A. A., Gorell, J. M., McDonald, C., Brown, S. J., Sillanpaa, M., & Shults, C. (1999). Validity of the Dementia Rating Scale in assessing cognitive function in Parkinson's disease. *Journal of geriatric psychiatry and neurology*, *12*(4), 180-188.

Brundel, M., De Bresser, J., Van Dillen, J. J., Kappelle, L. J., & Biessels, G. J. (2012). Cerebral microinfarcts: a systematic review of neuropathological studies. *Journal of Cerebral Blood Flow & Metabolism*, *32*(3), 425-436.

Bu, G. (2009). Apolipoprotein E and its receptors in Alzheimer's disease: pathways, pathogenesis and therapy. *Nature Reviews Neuroscience*, *10*(5), 333.

Bucholc, M., Ding, X., Wang, H., Glass, D. H., Wang, H., Prasad, G., ... & Alzheimer's Disease Neuroimaging Initiative. (2019). A practical computerized decision support system for predicting the severity of Alzheimer's disease of an individual. *Expert systems with applications*, *130*, 157-171.

Buckley, R. F., Schultz, A. P., Hedden, T., Papp, K. V., Hanseeuw, B. J., Marshall, G., ... & Sperling, R. A. (2017). Functional network integrity presages cognitive decline in preclinical Alzheimer disease. *Neurology*, *89*(1), 29-37.

Budd Haeberlein, S., Aisen, P. S., Barkhof, F., Chalkias, S., Chen, T., Cohen, S., ... & Sandrock, A. (2022). Two randomized phase 3 studies of aducanumab in early Alzheimer's disease. *The Journal of Prevention of Alzheimer's Disease*, *9*(2), 197-210.

Buerger, K., Ewers, M., Pirttilä, T., Zinkowski, R., Alafuzoff, I., Teipel, S. J., ... & Hampel, H. (2006). CSF phosphorylated tau protein correlates with

neocortical neurofibrillary pathology in Alzheimer's disease. *Brain*, *129*(11), 3035-3041.

Bullock, R., Touchon, J., Bergman, H., Gambina, G., He, Y., Rapatz, G., ... & Lane, R. (2005). Rivastigmine and donepezil treatment in moderate to moderately-severe Alzheimer's disease over a 2-year period. *Current medical research and opinion*, *21*(8), 1317-1327.

C

Caldwell, J. Z., Berg, J. L., Cummings, J. L., & Banks, S. J. (2017). Moderating effects of sex on the impact of diagnosis and amyloid positivity on verbal memory and hippocampal volume. *Alzheimer's research & therapy*, *9*(1), 72.

Calvin, C. M., de Boer, C., Raymont, V., Gallacher, J., & Koychev, I. (2020). Prediction of Alzheimer's disease biomarker status defined by the 'ATN framework'among cognitively healthy individuals: results from the EPAD longitudinal cohort study. *Alzheimer's Research & Therapy*, *12*(1), 1-16.

Carrascal, L. M., Galván, I., & Gordo, O. (2009). Partial least squares regression as an alternative to current regression methods used in ecology. *Oikos*, *118*(5), 681-690.

Castillo-Barnes, D., Su, L., Ramírez, J., Salas-Gonzalez, D., Martinez-Murcia, F. J., Illan, I. A., ... & Dominantly Inherited Alzheimer Network. (2020). Autosomal dominantly inherited alzheimer disease: Analysis of genetic subgroups by machine learning. *Information Fusion*, *58*, 153-167.

Cattell, R. B. (1971). Abilities: Their structure, growth, and action.

Cedarbaum, J. M., Jaros, M., Hernandez, C., Coley, N., Andrieu, S., Grundman, M., ... & Alzheimer's Disease Neuroimaging Initiative. (2013). Rationale for use of the Clinical Dementia Rating Sum of Boxes as a primary outcome measure for Alzheimer's disease clinical trials. *Alzheimer's & Dementia*, *9*(1), S45-S55.

Celarain, N., de Gordoa, J. S. R., Zelaya, M. V., Roldán, M., Larumbe, R., Pulido, L., ... & Mendioroz, M. (2016). TREM2 upregulation correlates with 5-hydroxymethycytosine enrichment in Alzheimer's disease hippocampus. *Clinical epigenetics*, *8*(1), 37.

Chai, X., Wu, S., Murray, T. K., Kinley, R., Cella, C. V., Sims, H., ... & Hutton, M. L. (2011). Passive immunization with anti-tau antibodies in two transgenic models reduction of tau pathology and delay of disease progression. *Journal of Biological Chemistry*, *286*(39), 34457-34467.

Chapman, K. R., Bing-Canar, H., Alosco, M. L., Steinberg, E. G., Martin, B., Chaisson, C., ... & Stern, R. A. (2016). Mini Mental State Examination and Logical Memory scores for entry into Alzheimer's disease trials. *Alzheimer's research & therapy*, *8*(1), 9.

Chávez-Gutiérrez, L., Bammens, L., Benilova, I., Vandersteen, A., Benurwar, M., Borgers, M., ... & Wiltfang, J. (2012). The mechanism of γ-secretase dysfunction in familial Alzheimer disease. *The EMBO journal*, *31*(10), 2261-2274.

Chen, Y. R., & Glabe, C. G. (2006). Distinct Early Folding and Aggregation Properties of Alzheimer Amyloid-β Peptides Aβ40 and Aβ42 STABLE TRIMER OR TETRAMER FORMATION BY Aβ42. *Journal of Biological Chemistry*, *281*(34), 24414-24422.

Chhatwal, J. P., Schultz, A. P., Marshall, G. A., Boot, B., Gomez-Isla, T., Dumurgier, J., ... & Sperling, R. A. (2016). Temporal T807 binding correlates with CSF tau and phospho-tau in normal elderly. *Neurology*, *87*(9), 920-926.

Chien, D. T., Bahri, S., Szardenings, A. K., Walsh, J. C., Mu, F., Su, M. Y., ... & Kolb, H. C. (2013). Early clinical PET imaging results with the novel PHF-tau radioligand [F-18]-T807. *Journal of Alzheimer's Disease*, *34*(2), 457-468.

Chien, D. T., Szardenings, A. K., Bahri, S., Walsh, J. C., Mu, F., Xia, C., ... & Kolb, H. C. (2014). Early clinical PET imaging results with the novel PHF-tau radioligand [F18]-T808. *Journal of Alzheimer's Disease*, *38*(1), 171-184.

Choe, Y. M., Lee, B. C., Choi, I. G., Suh, G. H., Lee, D. Y., Kim, J. W., & Alzheimer's Disease Neuroimaging Initiative. (2020). MMSE subscale scores as useful predictors of AD conversion in mild cognitive impairment. *Neuropsychiatric Disease and Treatment*, *16*, 1767.

Choi, J. Y., Cho, H., Ahn, S. J., Lee, J. H., Ryu, Y. H., Lee, M. S., & Lyoo, C. H. (2018). Off-Target 18F-AV-1451 binding in the basal ganglia correlates with age-related iron accumulation. *Journal of Nuclear Medicine*, *59*(1), 117-120.

Chow, V. W., Mattson, M. P., Wong, P. C., & Gleichmann, M. (2010). An overview of APP processing enzymes and products. *Neuromolecular medicine*, *12*(1), 1-12.

Christensen, H., & Mackinnon, A. (1993). The association between mental, social and physical activity and cognitive performance in young and old subjects. *Age and Ageing*, *22*(3), 175-182.

Citron, M., Westaway, D., Xia, W., Carlson, G., Diehl, T., Levesque, G., ... & Kholodenko, D. (1997). Mutant presenilins of Alzheimer's disease increase

production of 42-residue amyloid β-protein in both transfected cells and transgenic mice. *Nature medicine*, *3*(1), 67.

Clark, C. M., Sheppard, L., Fillenbaum, G. G., Galasko, D., Morris, J. C., Koss, E., ... & Cerad Investigators. (1999). Variability in annual Mini-Mental State Examination score in patients with probable Alzheimer disease: a clinical perspective of data from the Consortium to Establish a Registry for Alzheimer's Disease. *Archives of neurology*, *56*(7), 857-862.

Clark, C. M., Schneider, J. A., Bedell, B. J., Beach, T. G., Bilker, W. B., Mintun, M. A., ... & Krautkramer, M. J. (2011). Use of florbetapir-PET for imaging β-amyloid pathology. *Jama*, *305*(3), 275-283.

Clark, C. M., Pontecorvo, M. J., Beach, T. G., Bedell, B. J., Coleman, R. E., Doraiswamy, P. M., ... & Schneider, J. A. (2012). Cerebral PET with florbetapir compared with neuropathology at autopsy for detection of neuritic amyloid-β plaques: a prospective cohort study. *The Lancet Neurology*, *11*(8), 669-678.

Coblentz, J. M., Mattis, S., Zingesser, L. H., Kasoff, S. S., Wiśniewski, H. M., & Katzman, R. (1973). Presenile dementia: Clinical aspects and evaluation of cerebrospinal fluid dynamics. *Archives of Neurology*, *29*(5), 299-308.

Coley, N., Andrieu, S., Jaros, M., Weiner, M., Cedarbaum, J., & Vellas, B. (2011). Suitability of the Clinical Dementia Rating-Sum of Boxes as a single primary endpoint for Alzheimer's disease trials. *Alzheimer's & Dementia*, *7*(6), 602-610.

Collette, F., & Van der Linden, M. (2002). Brain imaging of the central executive component of working memory. *Neuroscience & Biobehavioral Reviews*, *26*(2), 105-125.

Colonna, M., & Wang, Y. (2016). TREM2 variants: new keys to decipher Alzheimer disease pathogenesis. *Nature Reviews Neuroscience*, *17*(4), 201.

Corder, E. H., Saunders, A. M., Strittmatter, W. J., Schmechel, D. E., Gaskell, P. C., Small, G., ... & Pericak-Vance, M. A. (1993). Gene dose of apolipoprotein E type 4 allele and the risk of Alzheimer's disease in late onset families. *Science*, *261*(5123), 921-923.

Cotman, C. W., Poon, W. W., Rissman, R. A., & Blurton-Jones, M. (2005). The role of caspase cleavage of tau in Alzheimer disease neuropathology. *Journal of neuropathology & experimental neurology*, *64*(2), 104-112.

Crane, P. K., Carle, A., Gibbons, L. E., Insel, P., Mackin, R. S., Gross, A., ... & Mungas, D. (2012). Development and assessment of a composite score for

memory in the Alzheimer’s Disease Neuroimaging Initiative (ADNI). *Brain imaging and behavior*, *6*(4), 502-516.

Creavin, S. T., Wisniewski, S., Noel-Storr, A. H., Trevelyan, C. M., Hampton, T., Rayment, D., ... & Patel, A. S. (2016). Mini-Mental State Examination (MMSE) for the detection of dementia in clinically unevaluated people aged 65 and over in community and primary care populations. *Cochrane Database of Systematic Reviews*, (1).

Crehan, H., Hardy, J., & Pocock, J. (2013). Blockage of CR1 prevents activation of rodent microglia. *Neurobiology of disease*, *54*, 139-149.

Crook, T., Bartus, R. T., Ferris, S. H., Whitehouse, P., Cohen, G. D., & Gershon, S. (1986). Age-associated memory impairment: Proposed diagnostic criteria and measures of clinical change—report of a national institute of mental health work group.

Cruchaga, C., Chakraverty, S., Mayo, K., Vallania, F. L., Mitra, R. D., Faber, K., ... & Boeve, B. F. (2012). Rare variants in APP, PSEN1 and PSEN2 increase risk for AD in late-onset Alzheimer's disease families. *PloS one*, *7*(2), e31039.

Crum, R. M., Anthony, J. C., Bassett, S. S., & Folstein, M. F. (1993). Population-based norms for the Mini-Mental State Examination by age and educational level. *Jama*, *269*(18), 2386-2391.

Cummings, J., Lee, G., Mortsdorf, T., Ritter, A., & Zhong, K. (2017). Alzheimer's disease drug development pipeline: 2017. *Alzheimer's & Dementia: Translational Research & Clinical Interventions*, *3*(3), 367-384.

Cummings, J. (2018). Lessons learned from Alzheimer disease: clinical trials with negative outcomes. *Clinical and translational science*, *11*(2), 147-152.

Cummings, J., Lee, G., Zhong, K., Fonseca, J., & Taghva, K. (2021). Alzheimer's disease drug development pipeline: 2021. Alzheimer's & Dementia: Translational Research & Clinical Interventions, 7(1), e12179.

Cunningham, W. R., Clayton, V., & Overton, W. (1975). Fluid and crystallized intelligence in young adulthood and old age. *Journal of Gerontology*, *30*(1), 53-55.

Cutler, A., Cutler, D. R., & Stevens, J. R. (2012). Random forests. In *Ensemble machine learning* (pp. 157-175). Springer, Boston, MA.

D

Dai, C. L., Hu, W., Tung, Y. C., Liu, F., Gong, C. X., & Iqbal, K. (2018). Tau passive immunization blocks seeding and spread of Alzheimer hyperphosphorylated Tau-induced pathology in 3× Tg-AD mice. *Alzheimer's research & therapy*, *10*(1), 13.

Darby, D. G., Pietrzak, R. H., Fredrickson, J., Woodward, M., Moore, L., Fredrickson, A., ... & Maruff, P. (2012). Intraindividual cognitive decline using a brief computerized cognitive screening test. *Alzheimer's & Dementia*, *8*(2), 95-104.

Datamonitor Healthcare, Pharma Intelligence, 2016. https://pharmaintelligence.informa.com/2016

Davidsson, P., & Blennow, K. (1998). Neurochemical dissection of synaptic pathology in Alzheimer's disease. *International psychogeriatrics*, *10*(1), 11-23.

Davies, C. A., Mann, D. M. A., Sumpter, P. Q., & Yates, P. O. (1987). A quantitative morphometric analysis of the neuronal and synaptic content of the frontal and temporal cortex in patients with Alzheimer's disease. *Journal of the neurological sciences*, *78*(2), 151-164.

De Jager, P. L., Ma, Y., McCabe, C., Xu, J., Vardarajan, B. N., Felsky, D., ... & Bennett, D. A. (2018). A multi-omic atlas of the human frontal cortex for aging and Alzheimer's disease research. *Scientific data*, *5*(1), 1-13.

De Strooper, B., Saftig, P., Craessaerts, K., Vanderstichele, H., Guhde, G., Annaert, W., ... & Van Leuven, F. (1998). Deficiency of presenilin-1 inhibits the normal cleavage of amyloid precursor protein. *Nature*, *391*(6665), 387.

DeKosky, S. T., & Scheff, S. W. (1990). Synapse loss in frontal cortex biopsies in Alzheimer's disease: correlation with cognitive severity. *Annals of Neurology: Official Journal of the American Neurological Association and the Child Neurology Society*, *27*(5), 457-464.

del Ser, T., Steinwachs, K. C., Gertz, H. J., Andres, M. V., Gomez-Carrillo, B., Medina, M., ... & Leon, T. (2013). Treatment of Alzheimer's disease with the GSK-3 inhibitor tideglusib: a pilot study. *Journal of Alzheimer's Disease*, *33*(1), 205-215.

Delacourte, A., Sergeant, N., Champain, D., Wattez, A., Maurage, C. A., Lebert, F., ... & David, J. P. (2002). Nonoverlapping but synergetic tau and APP pathologies in sporadic Alzheimer's disease. *Neurology*, *59*(3), 398-407.

Delano-Wood, L., Bondi, M. W., Sacco, J., Abeles, N. O. R. M., Jak, A. J., Libon, D. J., & Bozoki, A. (2009). Heterogeneity in mild cognitive impairment:

Differences in neuropsychological profile and associated white matter lesion pathology. *Journal of the International Neuropsychological Society*, *15*(6), 906-914.

Delis, D. C, Kramer, J. H., Kaplan, E., & Obei; B. A. (1987). The California Verbal Learning Test-Research edition. New York: Psychological Corporation.

Delobel, P., Lavenir, I., Fraser, G., Ingram, E., Holzer, M., Ghetti, B., ... & Goedert, M. (2008). Analysis of tau phosphorylation and truncation in a mouse model of human tauopathy. *The American journal of pathology*, *172*(1), 123-131.

Deming, Y., Li, Z., Benitez, B. A., & Cruchaga, C. (2018). Triggering receptor expressed on myeloid cells 2 (TREM2): a potential therapeutic target for Alzheimer disease?. *Expert opinion on therapeutic targets*, *22*(7), 587-598.

Dickerson, B. C., Feczko, E., Augustinack, J. C., Pacheco, J., Morris, J. C., Fischl, B., & Buckner, R. L. (2009). Differential effects of aging and Alzheimer's disease on medial temporal lobe cortical thickness and surface area. *Neurobiology of aging*, *30*(3), 432-440.

Dierckx, E., Engelborghs, S., De Raedt, R., Van Buggenhout, M., De Deyn, P. P., Verleye, G., ... & Ponjaert-Kristoffersen, I. (2008). Differentiation between dementia and depression among older persons: can the difference between actual and premorbid intelligence be useful?. *Journal of geriatric psychiatry and neurology*, *21*(4), 242-249.

Díez-Guerra, F. J. (2010). Neurogranin, a link between calcium/calmodulin and protein kinase C signaling in synaptic plasticity. *IUBMB life*, *62*(8), 597-606.

Dimitrov, M., Phipps, M., Zahn, T. P., & Grafman, J. (1999). A thoroughly modern Gage. *Neurocase*, *5*(4), 345-354.

Diniz, B. S., Yassuda, M. S., Nunes, P. V., Radanovic, M., & Forlenza, O. V. (2007). Mini-mental State Examination performance in mild cognitive impairment subtypes. *International Psychogeriatrics*, *19*(4), 647-656.

Doherty, B. M., Schultz, S. A., Oh, J. M., Koscik, R. L., Dowling, N. M., Barnhart, T. E., ... & LaRue, A. (2015). Amyloid burden, cortical thickness, and cognitive function in the Wisconsin Registry for Alzheimer's Prevention. *Alzheimer's & dementia: diagnosis, assessment & disease monitoring*, *1*(2), 160-169.

Dominguez, D., Tournoy, J., Hartmann, D., Huth, T., Cryns, K., Deforce, S., ... & Roebroek, A. J. (2005). Phenotypic and biochemical analyses of BACE1-and

BACE2-deficient mice. *Journal of Biological Chemistry*, *280*(35), 30797-30806.

Donohue, M. C., Sperling, R. A., Salmon, D. P., Rentz, D. M., Raman, R., Thomas, R. G., ... & Aisen, P. S. (2014). The preclinical Alzheimer cognitive composite: measuring amyloid-related decline. *JAMA neurology*, *71*(8), 961-970.

Donohue, M. C., Sperling, R. A., Petersen, R., Sun, C. K., Weiner, M. W., & Aisen, P. S. (2017). Association between elevated brain amyloid and subsequent cognitive decline among cognitively normal persons. *Jama*, *317*(22), 2305-2316.

Doraiswamy, P. M., Kaiser, L., Bieber, F., & Garman, R. L. (2001). The Alzheimer's Disease Assessment Scale: evaluation of psychometric properties and patterns of cognitive decline in multicenter clinical trials of mild to moderate Alzheimer's disease. *Alzheimer Disease & Associated Disorders*, *15*(4), 174-183.

Doraiswamy, P. M., Sperling, R. A., Coleman, R. E., Johnson, K. A., Reiman, E. M., Davis, M. D., ... & Carpenter, A. (2012). Amyloid-β assessed by florbetapir F 18 PET and 18-month cognitive decline: a multicenter study. *Neurology*, *79*(16), 1636-1644.

Doré, V., Villemagne, V. L., Bourgeat, P., Fripp, J., Acosta, O., Chetélat, G., ... & Ames, D. (2013). Cross-sectional and longitudinal analysis of the relationship between Aβ deposition, cortical thickness, and memory in cognitively unimpaired individuals and in Alzheimer disease. *JAMA neurology*, *70*(7), 903-911.

Dubois, B., & Albert, M. L. (2004). Amnestic MCI or prodromal Alzheimer's disease?. *The Lancet Neurology*, *3*(4), 246-248.

Dubois, B., Feldman, H. H., Jacova, C., DeKosky, S. T., Barberger-Gateau, P., Cummings, J., ... & Scheltens, P. (2007). Research criteria for the diagnosis of Alzheimer's disease: revising the NINCDS–ADRDA criteria. *The Lancet Neurology*, *6*(8), 734-746.

Dubois, B., Feldman, H. H., Jacova, C., Cummings, J. L., DeKosky, S. T., Barberger-Gateau, P., ... & Scheltens, P. (2010). Revising the definition of Alzheimer's disease: a new lexicon. *The Lancet Neurology*, *9*(11), 1118-1127.

Dubois, B., Feldman, H. H., Jacova, C., Hampel, H., Molinuevo, J. L., Blennow, K., ... & Cappa, S. (2014). Advancing research diagnostic criteria for Alzheimer's disease: the IWG-2 criteria. *The Lancet Neurology*, *13*(6), 614-629.

Dubois, B., Hampel, H., Feldman, H. H., Scheltens, P., Aisen, P., Andrieu, S., ... & Broich, K. (2016). Preclinical Alzheimer's disease: definition, natural history, and diagnostic criteria. *Alzheimer's & Dementia*, *12*(3), 292-323.

Dubois, B., Epelbaum, S., Nyasse, F., Bakardjian, H., Gagliardi, G., Uspenskaya, O., ... & Bertrand, A. (2018). Cognitive and neuroimaging features and brain β-amyloidosis in individuals at risk of Alzheimer's disease (INSIGHT-preAD): a longitudinal observational study. *The Lancet Neurology*, *17*(4), 335-346.

Duchesne, S., Caroli, A., Geroldi, C., Frisoni, G. B., & Collins, D. L. (2005, October). Predicting clinical variable from MRI features: application to MMSE in MCI. In *International Conference on Medical Image Computing and Computer-Assisted Intervention* (pp. 392-399). Springer, Berlin, Heidelberg.

Duering, M., Gesierich, B., Seiler, S., Pirpamer, L., Gonik, M., Hofer, E., ... & Dichgans, M. (2014). Strategic white matter tracts for processing speed deficits in age-related small vessel disease. *Neurology*, *82*(22), 1946-1950.

Duff, K., Eckman, C., Zehr, C., Yu, X., Prada, C. M., Perez-Tur, J., ... & Morgan, D. (1996). Increased amyloid-β42 (43) in brains of mice expressing mutant presenilin 1. *Nature*, *383*(6602), 710.

Duff, K., Beglinger, L. J., Moser, D. J., Paulsen, J. S., Schultz, S. K., & Arndt, S. (2010). Predicting cognitive change in older adults: the relative contribution of practice effects. *Archives of Clinical Neuropsychology*, *25*(2), 81-88.

Duff, K., Foster, N. L., & Hoffman, J. M. (2014). Practice effects and amyloid deposition: preliminary data on a method for enriching samples in clinical trials. *Alzheimer disease and associated disorders*, *28*(3), 247.

Duff, K., Atkinson, T. J., Suhrie, K. R., Dalley, B. C. A., Schaefer, S. Y., & Hammers, D. B. (2017). Short-term practice effects in mild cognitive impairment: Evaluating different methods of change. *Journal of clinical and experimental neuropsychology*, *39*(4), 396-407.

E

Egan, M. F., Kost, J., Voss, T., Mukai, Y., Aisen, P. S., Cummings, J. L., ... & Zhang, Y. (2019). Randomized Trial of Verubecestat for Prodromal Alzheimer’s Disease. *New England Journal of Medicine*, *380*(15), 1408-1420.

Eisdorfer, C., Busse, E. W., & Cohen, L. D. (1959). The WAIS performance of an aged sample: The relationship between verbal and performance IQ's. *Journal of gerontology*.

Ekstrom, R.B., French, J.W. & Harman, H.H. (1976). Fit of factor-referenced cognitive tests. Princeton, NJ: Educational Testing Service.

Elliott, R. (2003). Executive functions and their disorders: Imaging in clinical neuroscience. *British medical bulletin*, *65*(1), 49-59.

Ellis, K. A., Lim, Y. Y., Harrington, K., Ames, D., Bush, A. I., Darby, D., ... & Szoeke, C. (2013). Decline in cognitive function over 18 months in healthy older adults with high amyloid-β. *Journal of Alzheimer's Disease*, *34*(4), 861-871.

Ellul, J., Archer, N., Foy, C. M. L., Poppe, M., Boothby, H., Nicholas, H., ... & Lovestone, S. (2007). The effects of commonly prescribed drugs in patients with Alzheimer's disease on the rate of deterioration. *Journal of neurology, neurosurgery & psychiatry*, *78*(3), 233-239.

Emre, M., & Qizilbash, N. (2001). Experimental approaches and drugs in development for the treatment of dementia. *Expert opinion on investigational drugs*, *10*(4), 607-617.

European Medicines Association (2004). https://www.ema.europa.eu/en/documents/product-information/ebixa-epar-product-information_en.pdf

Ezzati, A., Zammit, A. R., Harvey, D. J., Habeck, C., Hall, C. B., Lipton, R. B., & Alzheimer's Disease Neuroimaging Initiative. (2019). Optimizing machine learning methods to improve predictive models of Alzheimer's disease. *Journal of Alzheimer's Disease*, *71*(3), 1027-1036.

F

Fagan, A. M., Roe, C. M., Xiong, C., Mintun, M. A., Morris, J. C., & Holtzman, D. M. (2007). Cerebrospinal fluid tau/β-amyloid42 ratio as a prediction of cognitive decline in nondemented older adults. *Archives of neurology*, *64*(3), 343-349.

Fagan, A. M., Head, D., Shah, A. R., Marcus, D., Mintun, M., Morris, J. C., & Holtzman, D. M. (2009). Decreased cerebrospinal fluid Aβ42 correlates with brain atrophy in cognitively normal elderly. *Annals of Neurology: Official Journal of the American Neurological Association and the Child Neurology Society*, *65*(2), 176-183.

Fagan, A. M., Xiong, C., Jasielec, M. S., Bateman, R. J., Goate, A. M., Benzinger, T. L., ... & Ringman, J. M. (2014). Longitudinal change in CSF

biomarkers in autosomal-dominant Alzheimer’s disease. *Science translational medicine*, *6*(226).

Farrell, M. E., Kennedy, K. M., Rodrigue, K. M., Wig, G., Bischof, G. N., Rieck, J. R., ... & Park, D. C. (2017). Association of longitudinal cognitive decline with amyloid burden in middle-aged and older adults: evidence for a dose-response relationship. *JAMA neurology*, *74*(7), 830-838.

Farrer, L. A., Cupples, L. A., Haines, J. L., Hyman, B., Kukull, W. A., Mayeux, R., ... & Van Duijn, C. M. (1997). Effects of age, sex, and ethnicity on the association between apolipoprotein E genotype and Alzheimer disease: a meta-analysis. *Jama*, *278*(16), 1349-1356.

Farrimond, L. E., Roberts, E., & McShane, R. (2012). Memantine and cholinesterase inhibitor combination therapy for Alzheimer's disease: a systematic review. *BMJ open*, *2*(3), e000917.

Feldman, H., Gauthier, S., Schneider, L., Wilcock, G., Frisoni, G., Bentham, P., ... & Storey, J. (2016). A phase 3 trial of the tau and TDP-43 aggregation inhibitor, leuco-methylthioninium-bis (hydromethanesulfonate)(LMTM), for behavioural variant frontotemporal dementia (bvFTD). *Journal of Neurochemistry*, *138*.

Felsky, D., Roostaei, T., Nho, K., Risacher, S. L., Bradshaw, E. M., Petyuk, V., ... & De Jager, P. L. (2019). Neuropathological correlates and genetic architecture of microglial activation in elderly human brain. *Nature communications*, *10*(1), 409.

Fernandez, M. A., Klutkowski, J. A., Freret, T., & Wolfe, M. S. (2014). Alzheimer presenilin-1 mutations dramatically reduce trimming of long amyloid β-peptides (Aβ) by γ-secretase to increase 42-to-40-residue Aβ. *Journal of Biological Chemistry*, *289*(45), 31043-31052.

Ferreira, D., Nordberg, A., & Westman, E. (2020). Biological subtypes of Alzheimer disease: A systematic review and meta-analysis. *Neurology*, *94*(10), 436-448.

Ferro, D. A., Mutsaerts, H. J., Hilal, S., Kuijf, H. J., Petersen, E. T., Petr, J., ... & Chen, C. (2020). Cortical microinfarcts in memory clinic patients are associated with reduced cerebral perfusion. *Journal of Cerebral Blood Flow & Metabolism*, *40*(9), 1869-1878.

Fillenbaum, G. G., Peterson, B., & Morris, J. C. (1996). Estimating the validity of the Clinical Dementia Rating scale: the CERAD experience. *Aging Clinical and Experimental Research*, *8*(6), 379-385.

Filser, S., Ovsepian, S. V., Masana, M., Blazquez-Llorca, L., Elvang, A. B., Volbracht, C., ... & Herms, J. (2015). Pharmacological inhibition of BACE1 impairs synaptic plasticity and cognitive functions. *Biological psychiatry*, *77*(8), 729-739.

Fine, E. M., Delis, D. C., Wetter, S. R., Jacobson, M. W., Jak, A. J., McDonald, C. R., ... & Bondi, M. W. (2008). Cognitive discrepancies versus APOE genotype as predictors of cognitive decline in normal-functioning elderly individuals: A longitudinal study. *The American Journal of Geriatric Psychiatry*, *16*(5), 366-374.

Fisher, C. K., Smith, A. M., & Walsh, J. R. (2019). Machine learning for comprehensive forecasting of Alzheimer's Disease progression. *Scientific reports*, *9*(1), 1-14.

Food & Drug Administration (FDA), 2003. https://www.accessdata.fda.gov/drugsatfda_docs/label/2010/022525s000lbl.pdf

Food & Drug Administration (FDA), 2012. Appendix 2. ADAS-Cog Administration And Scoring Manual. https://www.fda.gov/media/122843/download

Food & Drug Administration (FDA), 2017. https://www.fda.gov/regulatory-information/search-fda-guidance-documents/multiple-endpoints-clinical-trials-guidance-industry

Food & Drug Administration (FDA), 2018. https://www.fda.gov/regulatory-information/search-fda-guidance-documents/alzheimers-disease-developing-drugs-treatment-guidance-industy

Folstein, M. F., Folstein, S. E., & 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), 189-198.

Fox, N. C., Crum, W. R., Scahill, R. I., Stevens, J. M., Janssen, J. C., & Rossor, M. N. (2001). Imaging of onset and progression of Alzheimer's disease with voxel-compression mapping of serial magnetic resonance images. *The Lancet*, *358*(9277), 201-205.

Francis, P. T., Palmer, A. M., Snape, M., & Wilcock, G. K. (1999). The cholinergic hypothesis of Alzheimer's disease: a review of progress. *Journal of Neurology, Neurosurgery & Psychiatry*, *66*(2), 137-147.

Fratiglioni, L., Paillard-Borg, S., & Winblad, B. (2004). An active and socially integrated lifestyle in late life might protect against dementia. *The Lancet Neurology*, *3*(6), 343-353.

Fredrickson, J., Maruff, P., Woodward, M., Moore, L., Fredrickson, A., Sach, J., & Darby, D. (2010). Evaluation of the usability of a brief computerized cognitive screening test in older people for epidemiological studies. *Neuroepidemiology*, *34*(2), 65-75.

Frost, D., Gorman, P. M., Yip, C. M., & Chakrabartty, A. (2003). Co-incorporation of Aβ40 and Aβ42 to form mixed pre-fibrillar aggregates. *European journal of biochemistry*, *270*(4), 654-663.

G

Galvin, J. E., Powlishta, K. K., Wilkins, K., McKeel, D. W., Xiong, C., Grant, E., ... & Morris, J. C. (2005). Predictors of preclinical Alzheimer disease and dementia: a clinicopathologic study. *Archives of neurology*, *62*(5), 758-765.

Gamblin, T. C., Chen, F., Zambrano, A., Abraha, A., Lagalwar, S., Guillozet, A. L., ... & Miller, R. (2003). Caspase cleavage of tau: linking amyloid and neurofibrillary tangles in Alzheimer's disease. *Proceedings of the national academy of sciences*, *100*(17), 10032-10037.

Gauthier, S., Feldman, H. H., Schneider, L. S., Wilcock, G. K., Frisoni, G. B., Hardlund, J. H., ... & Schelter, B. O. (2016). Efficacy and safety of tau-aggregation inhibitor therapy in patients with mild or moderate Alzheimer's disease: a randomised, controlled, double-blind, parallel-arm, phase 3 trial. *The Lancet*, *388*(10062), 2873-2884.

Giacobini, E., & Gold, G. (2013). Alzheimer disease therapy—moving from amyloid-β to tau. *Nature Reviews Neurology*, *9*(12), 677.

Giovacchini, G., Giovannini, E., Borsò, E., Lazzeri, P., Riondato, M., Leoncini, R., ... & Ciarmiello, A. (2019). The brain cognitive reserve hypothesis: A review with emphasis on the contribution of nuclear medicine neuroimaging techniques. *Journal of cellular physiology*.

Glenner GG, Wong CW (1984) Alzheimer's disease: initial report of the purification and characterization of a novel cerebrovascular amyloid protein. Biochem Biophys Res Commun 120: 885–890

Gobbi, L. C., Knust, H., Körner, M., Honer, M., Czech, C., Belli, S., ... & Grall-Ulsemer, S. (2017). Identification of three novel radiotracers for imaging aggregated tau in Alzheimer’s disease with positron emission tomography. *Journal of medicinal chemistry*, *60*(17), 7350-7370.

Goedert, M., Spillantini, M. G., Cairns, N. J., & Crowther, R. A. (1992). Tau proteins of Alzheimer paired helical filaments: abnormal phosphorylation of all six brain isoforms. *Neuron*, *8*(1), 159-168.

Goedert, M. (2018). Tau filaments in neurodegenerative diseases. *FEBS letters*, *592*(14), 2383-2391.

Golby, A., Silverberg, G., Race, E., Gabrieli, S., O'shea, J., Knierim, K., ... & Gabrieli, J. (2005). Memory encoding in Alzheimer's disease: an fMRI study of explicit and implicit memory. *Brain*, *128*(4), 773-787.

Gonneaud, J., Arenaza-Urquijo, E. M., Mézenge, F., Landeau, B., Gaubert, M., Bejanin, A., ... & Abbas, A. (2017). Increased florbetapir binding in the temporal neocortex from age 20 to 60 years. *Neurology*, *89*(24), 2438-2446.

Gordon, B. A., Blazey, T. M., Su, Y., Hari-Raj, A., Dincer, A., Flores, S., ... & Cairns, N. J. (2018). Spatial patterns of neuroimaging biomarker change in individuals from families with autosomal dominant Alzheimer's disease: a longitudinal study. *The Lancet Neurology*, *17*(3), 241-250.

Graff-Radford, J., Yong, K. X., Apostolova, L. G., Bouwman, F. H., Carrillo, M., Dickerson, B. C., ... & Murray, M. E. (2021). New insights into atypical Alzheimer's disease in the era of biomarkers. *The Lancet Neurology*, *20*(3), 222-234.

Graham, D. P., Cully, J. A., Snow, A. L., Massman, P., & Doody, R. (2004). The Alzheimer's Disease Assessment Scale-Cognitive subscale: normative data for older adult controls. *Alzheimer Disease & Associated Disorders*, *18*(4), 236-240.

Grassi, M., Perna, G., Caldirola, D., Schruers, K., Duara, R., & Loewenstein, D. A. (2018). A clinically-translatable machine learning algorithm for the prediction of Alzheimer’s disease conversion in individuals with mild and premild cognitive impairment. *Journal of Alzheimer's Disease*, *61*(4), 1555-1573.

Greenberg, S. M., Vernooij, M. W., Cordonnier, C., Viswanathan, A., Salman, R. A. S., Warach, S., ... & Microbleed Study Group. (2009). Cerebral microbleeds: a guide to detection and interpretation. *The Lancet Neurology*, *8*(2), 165-174.

Griciuc A, Serrano-Pozo A, Parrado AR, Lesinski AN, Asselin CN, Mullin K, Hooli B, Choi SH, Hyman BT, Tanzi RE (2013) Alzheimer's disease risk gene CD33 inhibits microglial uptake of amyloid beta. Neuron 78: 631–643

Grober, E., Buschke, H. C. H. B. S., Crystal, H., Bang, S., & Dresner, R. (1988). Screening for dementia by memory testing. *Neurology*, *38*(6), 900-900.

Grober, E., Hall, C., Sanders, A. E., & Lipton, R. B. (2008). Free and cued selective reminding distinguishes Alzheimer's disease from vascular dementia. *Journal of the American Geriatrics Society*, *56*(5), 944.

Grober, E., Ocepek-Welikson, K. A. T. J. A., & Teresi, J. A. (2009). The free and cued selective reminding test: evidence of psychometric adequacy. *Psychology Science Quarterly*, *51*(3), 266-282.

Grundke-Iqbal, I., Iqbal, K., Tung, Y. C., Quinlan, M., Wisniewski, H. M., & Binder, L. I. (1986a). Abnormal phosphorylation of the microtubule-associated protein tau (tau) in Alzheimer cytoskeletal pathology. *Proceedings of the National Academy of Sciences*, *83*(13), 4913-4917.

Grundke-Iqbal, I., Iqbal, K., Quinlan, M., Tung, Y. C., Zaidi, M. S., & Wisniewski, H. M. (1986b). Microtubule-associated protein tau. A component of Alzheimer paired helical filaments. *Journal of Biological Chemistry*, *261*(13), 6084-6089.

Grundman, M., Petersen, R. C., Ferris, S. H., Thomas, R. G., Aisen, P. S., Bennett, D. A., ... & Kaye, J. (2004). Mild cognitive impairment can be distinguished from Alzheimer disease and normal aging for clinical trials. *Archives of neurology*, *61*(1), 59-66.

Guerreiro, R., Wojtas, A., Bras, J., Carrasquillo, M., Rogaeva, E., Majounie, E., ... & Hazrati, L. (2013). TREM2 variants in Alzheimer's disease. *New England Journal of Medicine*, *368*(2), 117-127.

Guerrero-Berroa, E., Luo, X., Schmeidler, J., Rapp, M.A., Dahlman, K., Grossman, H. T., (2010). The MMSE orientation for time domain is a strong predictor of subsequent cognitive decline in the elderly. International Journal of Geriatric Psychiatry, 24(12), 1429–1437.

H

Haan, M. N., Shemanski, L., Jagust, W. J., Manolio, T. A., & Kuller, L. (1999). The role of apoe∈ 4 in modulating effects of other risk factors for cognitive decline in elderly persons. *Jama*, *282*(1), 40-46.

Haass, C., & Selkoe, D. J. (2007). Soluble protein oligomers in neurodegeneration: lessons from the Alzheimer's amyloid β-peptide. *Nature reviews Molecular cell biology*, *8*(2), 101.

Hahn, A., Strandberg, T. O., Stomrud, E., Nilsson, M., van Westen, D., Palmqvist, S., ... & Hansson, O. (2019). Association Between Earliest Amyloid Uptake and Functional Connectivity in Cognitively Unimpaired Elderly. *Cerebral Cortex*.

Hanger, D. P., Anderton, B. H., & Noble, W. (2009). Tau phosphorylation: the therapeutic challenge for neurodegenerative disease. *Trends in molecular medicine*, *15*(3), 112-119.

Hanseeuw, B. J., Betensky, R. A., Jacobs, H. I., Schultz, A. P., Sepulcre, J., Becker, J. A., ... & Johnson, K. (2019). Association of amyloid and tau with cognition in preclinical Alzheimer disease: a longitudinal study. *JAMA neurology*, *76*(8), 915-924.

Hansen, A. K., Brooks, D. J., & Borghammer, P. (2018). MAO-B inhibitors do not block in vivo flortaucipir ([18 F]-AV-1451) binding. *Molecular Imaging and Biology*, *20*(3), 356-360.

Hansson, O. (2021). Biomarkers for neurodegenerative diseases. *Nature medicine*, *27*(6), 954-963.

Hamerman, J. A., Jarjoura, J. R., Tchao, N. K., Humphrey, M. B., Nakamura, M. C., Lowell, C. A., ... & Lanier, L. L. (2007). Inhibition of Toll-like receptor and Fc receptor responses in macrophages by TREM-2 and DAP12 (44.7).

Hampel, H., Vergallo, A., Aguilar, L. F., Benda, N., Broich, K., Cuello, A. C., ... & Genthon, R. (2018). Precision pharmacology for Alzheimer's disease. *Pharmacological research*, *130*, 331-365.

Hampel, H., Cummings, J., Blennow, K., Gao, P., Jack, C. R., & Vergallo, A. (2021). Developing the ATX (N) classification for use across the Alzheimer disease continuum. *Nature Reviews Neurology*, *17*(9), 580-589.

Hampel, H., Au, R., Mattke, S., van der Flier, W. M., Aisen, P., Apostolova, L., ... & Cummings, J. (2022). Designing the next-generation clinical care pathway for Alzheimer's disease. *Nature Aging*, *2*(8), 692-703.

Harada, R., Okamura, N., Furumoto, S., Furukawa, K., Ishiki, A., Tomita, N., ... & Miyake, M. (2016). 18F-THK5351: a novel PET radiotracer for imaging neurofibrillary pathology in Alzheimer disease. *Journal of Nuclear Medicine*, *57*(2), 208-214.

Hardy J, Allsop D (1991) Amyloid deposition as the central event in the aetiology of Alzheimer's disease. Trends in Pharmac 12: 383–388

Hardy JA, Higgins G (1992) Alzheimer's disease: the amyloid cascade hypothesis. Science 256: 184–185

Harold, D., Abraham, R., Hollingworth, P., Sims, R., Gerrish, A., Hamshere, M. L., ... & Jones, N. (2009). Genome-wide association study identifies variants at CLU and PICALM associated with Alzheimer's disease. *Nature genetics*, *41*(10), 1088.

Harrison, J. E., Rentz, D. M., Brashear, H. R., Arrighi, H. M., Ropacki, M. T., & Liu, E. (2018). Psychometric Evaluation of the Neuropsychological Test Battery in Individuals with Normal Cognition, Mild Cognitive Impairment, or Mild to Moderate Alzheimer's Disease: Results from a Longitudinal Study. *The journal of prevention of Alzheimer's disease*, *5*(4), 236-244.

Hassenstab, J., Ruvolo, D., Jasielec, M., Xiong, C., Grant, E., & Morris, J. C. (2015). Absence of practice effects in preclinical Alzheimer's disease. *Neuropsychology*, *29*(6), 940.

Hedden, T., Oh, H., Younger, A. P., & Patel, T. A. (2013). Meta-analysis of amyloid-cognition relations in cognitively normal older adults. *Neurology*, *80*(14), 1341-1348.

Heinonen, O., Soininen, H., Sorvari, H., Kosunen, O., Palja, L., Koivisto, E., & Riekkinen Sr, P. J. (1995). Loss of synaptophysin-like immunoreactivity in the hippocampal formation is an early phenomenon in Alzheimer's disease. *Neuroscience*, *64*(2), 375-384.

Heneka, M. T., Carson, M. J., El Khoury, J., Landreth, G. E., Brosseron, F., Feinstein, D. L., ... & Herrup, K. (2015). Neuroinflammation in Alzheimer's disease. *The Lancet Neurology*, *14*(4), 388-405.

Henjum, K., Almdahl, I. S., Årskog, V., Minthon, L., Hansson, O., Fladby, T., & Nilsson, L. N. (2016). Cerebrospinal fluid soluble TREM2 in aging and Alzheimer's disease. *Alzheimer's research & therapy*, *8*(1), 17.

Hensel, A., Angermeyer, M. C., & Riedel-Heller, S. G. (2007). Measuring cognitive change in older adults: reliable change indices for the Mini-Mental State Examination. *Journal of Neurology, Neurosurgery & Psychiatry*, *78*(12), 1298-1303.

Heslegrave, A., Heywood, W., Paterson, R., Magdalinou, N., Svensson, J., Johansson, P., ... & Mills, K. (2016). Increased cerebrospinal fluid soluble TREM2 concentration in Alzheimer's disease. *Molecular neurodegeneration*, *11*(1), 3.

Herholz, K., Salmon, E., Perani, D., Baron, J. C., Holthoff, V., Frölich, L., ... & Zündorf, G. (2002). Discrimination between Alzheimer dementia and controls by automated analysis of multicenter FDG PET. *Neuroimage*, *17*(1), 302-316.

Herrup, K. (2015). The case for rejecting the amyloid cascade hypothesis. *Nature neuroscience*, *18*(6), 794.

Hickman, S. E., Howieson, D. B., Dame, A., Sexton, G., & Kaye, J. (2000). Longitudinal analysis of the effects of the aging process on neuropsychological test performance in the healthy young-old and oldest-old. *Developmental neuropsychology*, *17*(3), 323-337.

Hickman, S. E., & El Khoury, J. (2014). TREM2 and the neuroimmunology of Alzheimer's disease. *Biochemical pharmacology*, *88*(4), 495-498.

Hirni, D. I., Kivisaari, S. L., Monsch, A. U., & Taylor, K. I. (2013). Distinct neuroanatomical bases of episodic and semantic memory performance in Alzheimer's disease. *Neuropsychologia*, *51*(5), 930-937.

Hippius, H. & Neundörfer, G., (2003). Historischer Hintergrund der Alzheimer-Demenz. Alois Alzheimer–sein Name im Kontext der Erkrankung. *Alzheimer-Demenz. Klinische Verläufe, diagnostische Möglichkeiten, moderne Therapiestrategien. Wissenschaftliche Verlagsges. Stuttgart*, 1-13.

Hobart, J., Cano, S., Posner, H., Selnes, O., Stern, Y., Thomas, R., ... & Alzheimer's Disease Neuroimaging Initiative. (2013). Putting the Alzheimer's cognitive test to the test II: Rasch Measurement Theory. *Alzheimer's & Dementia*, *9*(1), S10-S20.

Hoenig, M. C., Bischof, G. N., Hammes, J., Faber, J., Fliessbach, K., van Eimeren, T., & Drzezga, A. (2017). Tau pathology and cognitive reserve in Alzheimer's disease. *Neurobiology of aging*, *57*, 1-7.

Hollands, S., Lim, Y. Y., Buckley, R., Pietrzak, R. H., Snyder, P. J., Ames, D., ... & Masters, C. L. (2015). Amyloid-β related memory decline is not associated with subjective or informant rated cognitive impairment in healthy adults. *Journal of Alzheimer's Disease*, *43*(2), 677-686.

Hollingworth, P., Harold, D., Sims, R., Gerrish, A., Lambert, J. C., Carrasquillo, M. M., ... & Dowzell, K. (2011). Common variants at ABCA7, MS4A6A/MS4A4E, EPHA1, CD33 and CD2AP are associated with Alzheimer's disease. *Nature genetics*, *43*(5), 429.

Honer, M., Gobbi, L., Knust, H., Kuwabara, H., Muri, D., Koerner, M., ... & Borroni, E. (2018). Preclinical evaluation of 18F-RO6958948, 11C-RO6931643, and 11C-RO6924963 as novel PET radiotracers for imaging tau

aggregates in Alzheimer disease. *Journal of Nuclear Medicine*, *59*(4), 675-681.

Houston, W. S., Delis, D. C., Lansing, A., Jacobson, M. W., Cobell, K. R., Salmon, D. P., & Bondi, M. W. (2005). Executive function asymmetry in older adults genetically at-risk for Alzheimer's disease: verbal versus design fluency. *Journal of the International Neuropsychological Society*, *11*(7), 863-870.

Hsieh, C. L., Koike, M., Spusta, S. C., Niemi, E. C., Yenari, M., Nakamura, M. C., & Seaman, W. E. (2009). A role for TREM2 ligands in the phagocytosis of apoptotic neuronal cells by microglia. *Journal of neurochemistry*, *109*(4), 1144-1156.

Hu, W., Wu, F., Zhang, Y., Gong, C. X., Iqbal, K., & Liu, F. (2017). Expression of tau pathology-related proteins in different brain regions: a molecular basis of tau pathogenesis. *Frontiers in aging neuroscience*, *9*, 311.

Hubbard, B. M., FENTONM, G., & Anderson, J. M. (1990). A quantitative histological study of early clinical and preclinical Alzheimer's disease. *Neuropathology and applied neurobiology*, *16*(2), 111-121.

Huang, K. P., Huang, F. L., Jäger, T., Li, J., Reymann, K. G., & Balschun, D. (2004). Neurogranin/RC3 enhances long-term potentiation and learning by promoting calcium-mediated signaling. *Journal of Neuroscience*, *24*(47), 10660-10669.

Huang, Y., & Mucke, L. (2012). Alzheimer mechanisms and therapeutic strategies. *Cell*, *148*(6), 1204-1222.

Huang, S., Li, J., Sun, L., Ye, J., Fleisher, A., Wu, T., ... & Alzheimer's Disease NeuroImaging Initiative. (2010). Learning brain connectivity of Alzheimer's disease by sparse inverse covariance estimation. *NeuroImage*, *50*(3), 935-949.

Hughes, C. P., Berg, L., Danziger, W., Coben, L. A., & Martin, R. L. (1982). A new clinical scale for the staging of dementia. *The British journal of psychiatry*, *140*(6), 566-572.

Huijbers, W., Mormino, E. C., Wigman, S. E., Ward, A. M., Vannini, P., McLaren, D. G., ... & Sperling, R. A. (2014). Amyloid deposition is linked to aberrant entorhinal activity among cognitively normal older adults. *Journal of Neuroscience*, *34*(15), 5200-5210.

Hyman, B. T., Phelps, C. H., Beach, T. G., Bigio, E. H., Cairns, N. J., Carrillo, M. C., ... & Mirra, S. S. (2012). National Institute on Aging–Alzheimer's

Association guidelines for the neuropathologic assessment of Alzheimer's disease. *Alzheimer's & dementia*, *8*(1), 1-13.

I

Ihara, R., Iwata, A., Suzuki, K., Ikeuchi, T., Kuwano, R., Iwatsubo, T., & Japanese Alzheimer's Disease Neuroimaging Initiative. (2018). Clinical and cognitive characteristics of preclinical Alzheimer's disease in the Japanese Alzheimer's Disease Neuroimaging Initiative cohort. *Alzheimer's & Dementia: Translational Research & Clinical Interventions*, *4*, 645-651.

Ikonomovic, M. D., Klunk, W. E., Abrahamson, E. E., Mathis, C. A., Price, J. C., Tsopelas, N. D., ... & Debnath, M. L. (2008). Post-mortem correlates of in vivo PiB-PET amyloid imaging in a typical case of Alzheimer's disease. *Brain*, *131*(6), 1630-1645.

Ingala, S., de Boer, C., Masselink, L. A., Vergari, I., Lorenzini, L., Blennow, K., ... & EPAD Consortium. (2020). Operationalization of the ATN classification scheme in preclinical AD: Findings from EPAD V500. 0 data release: Biomarkers (non-neuroimaging)/Use in clinical trial design and evaluation. *Alzheimer's & Dementia*, *16*, e037912.

Iqbal, K., Zaidi, T., Wen, G., Grundke-Iqbal, I., Merz, P., Shaikh, S., ... & Winblad, B. (1986). Defective brain microtubule assembly in Alzheimer's disease. *The Lancet*, *328*(8504), 421-426.

Iqbal, K., Grundke-Iqbal, I., Smith, A. J., George, L., Tung, Y. C., & Zaidi, T. (1989). Identification and localization of a tau peptide to paired helical filaments of Alzheimer disease. *Proceedings of the National Academy of Sciences*, *86*(14), 5646-5650.

Iqbal, K., Alonso, A. D. C., Chen, S., Chohan, M. O., El-Akkad, E., Gong, C. X., ... & Tanimukai, H. (2005). Tau pathology in Alzheimer disease and other tauopathies. *Biochimica et Biophysica Acta (BBA)-Molecular Basis of Disease*, *1739*(2-3), 198-210.

Isaacs, B., & Akhtar, A. J. (1972). The set test: a rapid test of mental function in old people. *Age and ageing*, *1*(4), 222-226.

Ithapu, V.K. *et al.* (2015) “Imaging-based enrichment criteria using deep learning algorithms for efficient clinical trials in mild cognitive impairment,” *Alzheimer’s & Dementia*, 11(12), pp. 1489–1499

Ito, H., & Hamerman, J. A. (2012). TREM-2, triggering receptor expressed on myeloid cell-2, negatively regulates TLR responses in dendritic cells. *European journal of immunology*, *42*(1), 176-185.

J

Jack, C. R., Petersen, R. C., Xu, Y. C., O'Brien, P. C., Smith, G. E., Ivnik, R. J., ... & Kokmen, E. (1999). Prediction of AD with MRI-based hippocampal volume in mild cognitive impairment. *Neurology*, *52*(7), 1397-1397.

Jack, C. R., Shiung, M. M., Weigand, S. D., O'brien, P. C., Gunter, J. L., Boeve, B. F., ... & Petersen, R. C. (2005). Brain atrophy rates predict subsequent clinical conversion in normal elderly and amnestic MCI. *Neurology*, *65*(8), 1227-1231.

Jack Jr, C. R., Bernstein, M. A., Fox, N. C., Thompson, P., Alexander, G., Harvey, D., ... & Dale, A. M. (2008). The Alzheimer's disease neuroimaging initiative (ADNI): MRI methods. *Journal of Magnetic Resonance Imaging: An Official Journal of the International Society for Magnetic Resonance in Medicine*, *27*(4), 685-691.

Jack Jr, C. R., Knopman, D. S., Jagust, W. J., Shaw, L. M., Aisen, P. S., Weiner, M. W., ... & Trojanowski, J. Q. (2010). Hypothetical model of dynamic biomarkers of the Alzheimer's pathological cascade. *The Lancet Neurology*, *9*(1), 119-128.

Jack Jr, C. R., Knopman, D. S., Jagust, W. J., Petersen, R. C., Weiner, M. W., Aisen, P. S., ... & Lesnick, T. G. (2013). Tracking pathophysiological processes in Alzheimer's disease: an updated hypothetical model of dynamic biomarkers. *The Lancet Neurology*, *12*(2), 207-216.

Jack, C. R., Bennett, D. A., Blennow, K., Carrillo, M. C., Feldman, H. H., Frisoni, G. B., ... & Petersen, R. C. (2016). A/T/N: an unbiased descriptive classification scheme for Alzheimer disease biomarkers. *Neurology*, *87*(5), 539-547.

Jack Jr, C. R., Wiste, H. J., Weigand, S. D., Therneau, T. M., Lowe, V. J., Knopman, D. S., ... & Machulda, M. M. (2017). Defining imaging biomarker cut points for brain aging and Alzheimer's disease. *Alzheimer's & Dementia*, *13*(3), 205-216.

Jack Jr, C. R., Bennett, D. A., Blennow, K., Carrillo, M. C., Dunn, B., Haeberlein, S. B., ... & Liu, E. (2018). NIA-AA Research Framework: Toward a biological definition of Alzheimer's disease. *Alzheimer's & Dementia*, *14*(4), 535-562.

Jack, C. R., Wiste, H. J., Therneau, T. M., Weigand, S. D., Knopman, D. S., Mielke, M. M., ... & Petersen, R. C. (2019). Associations of amyloid, tau, and neurodegeneration biomarker profiles with rates of memory decline among individuals without dementia. *Jama*, *321*(23), 2316-2325.

Jacobson, M. W., Delis, D. C., Bondi, M. W., & Salmon, D. P. (2002). Do neuropsychological tests detect preclinical Alzheimer's disease: individual-test versus cognitive-discrepancy score analyses. *Neuropsychology*, *16*(2), 132.

Jacobson, M. W., Delis, D. C., Lansing, A., Houston, W., Olsen, R., Wetter, S., ... & Salmon, D. P. (2005). Asymmetries in Global-Local Processing Ability in Elderly People With the Apolipoprotein E-ε4 Allele. *Neuropsychology*, *19*(6), 822.

Jagust, W. (2018). Imaging the evolution and pathophysiology of Alzheimer disease. *Nature Reviews Neuroscience*, *19*(11), 687.

Jagust, W., Jack Jr, C. R., Bennett, D. A., Blennow, K., Haeberlein, S. B., Holtzman, D. M., ... & Sperling, R. (2019). Alzheimer's disease" is neither" Alzheimer's clinical syndrome" nor" dementia. *Alzheimer's & dementia: the journal of the Alzheimer's Association*, *15*(1), 153-157.

James, B. D., Wilson, R. S., Boyle, P. A., Trojanowski, J. Q., Bennett, D. A., & Schneider, J. A. (2016). TDP-43 stage, mixed pathologies, and clinical Alzheimer’s-type dementia. *Brain*, *139*(11), 2983-2993.

James, H. J., Van Houtven, C. H., Lippmann, S., Burke, J. R., Shepherd-Banigan, M., Belanger, E., ... & Plassman, B. L. (2020). How Accurately Do Patients and Their Care Partners Report Results of Amyloid-β PET Scans for Alzheimer’s Disease Assessment?. *Journal of Alzheimer's Disease*, *74*(2), 625-636.

Jan, A., Gokce, O., Luthi-Carter, R., & Lashuel, H. A. (2008). The ratio of monomeric to aggregated forms of Aβ40 and Aβ42 is an important determinant of amyloid-β aggregation, fibrillogenesis, and toxicity. *Journal of Biological Chemistry*, *283*(42), 28176-28189.

Janelidze, S., Mattsson, N., Palmqvist, S., Smith, R., Beach, T. G., Serrano, G. E., ... & Hansson, O. (2020). Plasma P-tau181 in Alzheimer’s disease: relationship to other biomarkers, differential diagnosis, neuropathology and longitudinal progression to Alzheimer’s dementia. *Nature medicine*, *26*(3), 379-386.

Janelidze, S., Berron, D., Smith, R., Strandberg, O., Proctor, N. K., Dage, J. L., ... & Hansson, O. (2021). Associations of plasma phospho-Tau217 levels with tau positron emission tomography in early Alzheimer disease. *JAMA neurology*, *78*(2), 149-156.

Jang, Y. K., Lyoo, C. H., Park, S., Oh, S. J., Cho, H., Oh, M., ... & Moon, S. H. (2018). Head to head comparison of [18 F] AV-1451 and [18 F] THK5351 for tau imaging in Alzheimer's disease and frontotemporal dementia. *European journal of nuclear medicine and molecular imaging*, *45*(3), 432-442.

Jansen, I. E., van der Lee, S. J., Gomez-Fonseca, D., de Rojas, I., Dalmasso, M. C., Grenier-Boley, B., ... & van der Flier, W. M. (2022). Genome-wide meta-analysis for Alzheimer´ s disease cerebrospinal fluid biomarkers. *medRxiv*.

Jarrett, J. T., & Lansbury Jr, P. T. (1993). Seeding "one-dimensional crystallization" of amyloid: a pathogenic mechanism in Alzheimer's disease and scrapie?. *Cell*, *73*(6), 1055-1058.

Jay, T. R., Miller, C. M., Cheng, P. J., Graham, L. C., Bemiller, S., Broihier, M. L., ... & Cotleur, A. C. (2015). TREM2 deficiency eliminates TREM2+ inflammatory macrophages and ameliorates pathology in Alzheimer's disease mouse models. *Journal of Experimental Medicine*, *212*(3), 287-295.

Jellinger, K. A. (2005). Understanding the pathology of vascular cognitive impairment. *Journal of the neurological sciences*, *229*, 57-63.

Jellinger, K. A. (2007). The enigma of vascular cognitive disorder and vascular dementia. *Acta neuropathologica*, *113*(4), 349-388.

Jellinger, K. A. (2013). Pathology and pathogenesis of vascular cognitive impairment—a critical update. *Frontiers in aging neuroscience*, *5*, 17.

Jiang, T., Tan, L., Zhu, X. C., Zhang, Q. Q., Cao, L., Tan, M. S., ... & Yu, J. T. (2014). Upregulation of TREM2 ameliorates neuropathology and rescues spatial cognitive impairment in a transgenic mouse model of Alzheimer's disease. *Neuropsychopharmacology*, *39*(13), 2949.

Jicha, G. A., Lane, E., Vincent, I., Otvos Jr, L., Hoffmann, R., & Davies, P. (1997a). A conformation-and phosphorylation-dependent antibody recognizing the paired helical filaments of Alzheimer's disease. *Journal of neurochemistry*, *69*(5), 2087-2095.

Jicha, G. A., Berenfeld, B., & Davies, P. (1999a). Sequence requirements for formation of conformational variants of tau similar to those found in Alzheimer's disease. *Journal of neuroscience research*, *55*(6), 713-723.

Jicha, G. A., Weaver, C., Lane, E., Vianna, C., Kress, Y., Rockwood, J., & Davies, P. (1999b). cAMP-dependent protein kinase phosphorylations on tau in Alzheimer's disease. *Journal of Neuroscience*, *19*(17), 7486-7494.

Jin, S. C., Benitez, B. A., Karch, C. M., Cooper, B., Skorupa, T., Carrell, D., ... & Bertelsen, S. (2014). Coding variants in TREM2 increase risk for Alzheimer's disease. *Human molecular genetics*, *23*(21), 5838-5846.

Jin, K. H., McCann, M. T., Froustey, E., & Unser, M. (2017). Deep convolutional neural network for inverse problems in imaging. *IEEE Transactions on Image Processing*, *26*(9), 4509-4522.

Jin, M., O'Nuallain, B., Hong, W., Boyd, J., Lagomarsino, V. N., O'Malley, T. T., ... & Walsh, D. M. (2018). An in vitro paradigm to assess potential anti-Aβ antibodies for Alzheimer's disease. *Nature communications*, *9*(1), 1-13.

Johnson, S. K., & Anderson, M. C. (2004). The role of inhibitory control in forgetting semantic knowledge. *Psychological science*, *15*(7), 448-453.

Johnson, S. C., Christian, B. T., Okonkwo, O. C., Oh, J. M., Harding, S., Xu, G., ... & Hall, L. T. (2014). Amyloid burden and neural function in people at risk for Alzheimer's disease. *Neurobiology of aging*, *35*(3), 576-584.

Josephs, K. A., Murray, M. E., Whitwell, J. L., Tosakulwong, N., Weigand, S. D., Petrucelli, L., ... & Dickson, D. W. (2016). Updated TDP-43 in Alzheimer's disease staging scheme. *Acta neuropathologica*, *131*(4), 571-585.

Jones, N. (2010). Alzheimer disease: risk of dementia and Alzheimer disease increases with occupational pesticide exposure. *Nature Reviews Neurology*, *6*(7), 353.

Jonsson, T., Stefansson, H., Steinberg, S., Jonsdottir, I., Jonsson, P. V., Snaedal, J., ... & Rujescu, D. (2013). Variant of TREM2 associated with the risk of Alzheimer's disease. *New England Journal of Medicine*, *368*(2), 107-116.

Jordão, J. F., Ayala-Grosso, C. A., Markham, K., Huang, Y., Chopra, R., McLaurin, J., ... & Aubert, I. (2010). Antibodies targeted to the brain with image-guided focused ultrasound reduces amyloid-β plaque load in the TgCRND8 mouse model of Alzheimer's disease. *PloS one*, *5*(5), e10549.

Jordão, J. F., Thévenot, E., Markham-Coultes, K., Scarcelli, T., Weng, Y. Q., Xhima, K., ... & Aubert, I. (2013). Amyloid-β plaque reduction, endogenous antibody delivery and glial activation by brain-targeted, transcranial focused ultrasound. *Experimental neurology*, *248*, 16-29.

Jutten, R. J., Sikkes, S. A., Amariglio, R. E., Buckley, R. F., Properzi, M. J., Marshall, G. A., ... & Alzheimer Disease Neuroimaging Initiative. (2020). Identifying sensitive measures of cognitive decline at different clinical stages

of Alzheimer’s disease. *Journal of the International Neuropsychological Society*, *27*(5), 426-438.

K

Kalaria, R. N. (2016). Neuropathological diagnosis of vascular cognitive impairment and vascular dementia with implications for Alzheimer’s disease. *Acta neuropathologica*, *131*(5), 659-685.

Kanekiyo, T., Xu, H., & Bu, G. (2014). ApoE and Aβ in Alzheimer’s disease: accidental encounters or partners?. *Neuron*, *81*(4), 740-754.

Kang, Y., Na, D. L., & Hahn, S. (1997). A validity study on the Korean Mini-Mental State Examination (K-MMSE) in dementia patients. *Journal of the Korean neurological association*, *15*(2), 300-308.

Kaplan, E., Goodglass, H. & Weintraub, S. (1976). Boston Naming Test, Experimental Edition, Veterans Administration Hospital, Boston.

Kaplow J, Vandijck M, Gray J, et al. Concordance of Lumipulse cerebrospinalfluid t-tau/Aβ42 ratio with amyloid PET status. Alzheimer’sDement. 2020;16:144–152.

Karch, C. M., & Goate, A. M. (2015). Alzheimer’s disease risk genes and mechanisms of disease pathogenesis. *Biological psychiatry*, *77*(1), 43-51.

Karp, A., Paillard-Borg, S., Wang, H. X., Silverstein, M., Winblad, B., & Fratiglioni, L. (2006). Mental, physical and social components in leisure activities equally contribute to decrease dementia risk. *Dementia and geriatric cognitive disorders*, *21*(2), 65-73.

Karr, J. E., Graham, R. B., Hofer, S. M., & Muniz-Terrera, G. (2018). When does cognitive decline begin? A systematic review of change point studies on accelerated decline in cognitive and neurological outcomes preceding mild cognitive impairment, dementia, and death. *Psychology and aging*, *33*(2), 195.

Karran, E., & De Strooper, B. (2016). The amyloid cascade hypothesis: are we poised for success or failure?. *Journal of neurochemistry*, *139*, 237-252.

Kaszniak, A. W., Poon, L. W., & Riege, W. H. (1986). Assessing memory deficits: an information-processing approach.

Katzman, R. (1993). Education and the prevalence of dementia and Alzheimer's disease. *Neurology*.

Kaufman, A. S., Kaufman, J. C., Liu, X., & Johnson, C. K. (2009). How do educational attainment and gender relate to fluid intelligence, crystallized intelligence, and academic skills at ages 22–90 years?. *Archives of Clinical Neuropsychology*, *24*(2), 153-163.

Kavé, G., Sapir-Yogev, S., Bregman, N., & Shiner, T. (2021). On the importance of using local tests and local norms in the assessment of memory. *Applied Neuropsychology: Adult*, *29*(6), 1492-1498.

Kemppainen, N. M., Aalto, S., Karrasch, M., Någren, K., Savisto, N., Oikonen, V., ... & Rinne, J. O. (2008). Cognitive reserve hypothesis: Pittsburgh Compound B and fluorodeoxyglucose positron emission tomography in relation to education in mild Alzheimer's disease. *Annals of Neurology: Official Journal of the American Neurological Association and the Child Neurology Society*, *63*(1), 112-118.

Keskin, A. D., Kekuš, M., Adelsberger, H., Neumann, U., Shimshek, D. R., Song, B., ... & Nelken, I. (2017). BACE inhibition-dependent repair of Alzheimer's pathophysiology. *Proceedings of the National Academy of Sciences*, *114*(32), 8631-8636.

Kettaneh, N., Berglund, A., & Wold, S. (2005). PCA and PLS with very large data sets. *Computational Statistics & Data Analysis*, *48*(1), 69-85.

Khan, A., & Usman, M. (2019). Alzheimer's Disease Prediction Model Using Demographics and Categorical Data.

Khatoon, S., Grundke-Iqbal, I., & Iqbal, K. (1992). Brain levels of microtubule-associated protein τ are elevated in Alzheimer's disease: A radioimmuno-slot-blot assay for nanograms of the protein. *Journal of neurochemistry*, *59*(2), 750-753.

Kim J, Onstead L, Randle S, Price R, Smithson L, Zwizinski C, Dickson DW, Golde T, McGowan E (2007) Abeta40 inhibits amyloid deposition in vivo. J Neurosci 27: 627–633

Kirschner, D. A., Abraham, C., & Selkoe, D. J. (1986). X-ray diffraction from intraneuronal paired helical filaments and extraneuronal amyloid fibers in Alzheimer disease indicates cross-beta conformation. *Proceedings of the National Academy of Sciences*, *83*(2), 503-507.

Kleinberger, G., Yamanishi, Y., Suárez-Calvet, M., Czirr, E., Lohmann, E., Cuyvers, E., ... & Tahirovic, S. (2014). TREM2 mutations implicated in neurodegeneration impair cell surface transport and phagocytosis. *Science translational medicine*, *6*(243)

Klunk, W. E., Engler, H., Nordberg, A., Wang, Y., Blomqvist, G., Holt, D. P., ... & Ausén, B. (2004). Imaging brain amyloid in Alzheimer's disease with Pittsburgh Compound-B. *Annals of Neurology: Official Journal of the American Neurological Association and the Child Neurology Society*, *55*(3), 306-319.

Knopman, D. S., Jack Jr, C. R., Wiste, H. J., Weigand, S. D., Vemuri, P., Lowe, V. J., ... & Roberts, R. O. (2013). Brain injury biomarkers are not dependent on β-amyloid in normal elderly. *Annals of neurology*, *73*(4), 472-480.

Koedam, E. L., Lauffer, V., van der Vlies, A. E., van der Flier, W. M., Scheltens, P., & Pijnenburg, Y. A. (2010). Early-versus late-onset Alzheimer's disease: more than age alone. *Journal of Alzheimer's Disease*, *19*(4), 1401-1408.

Kohannim, O. *et al.* (2010) “Boosting power for clinical trials using classifiers based on multiple biomarkers,” *Neurobiology of Aging*, 31(8), pp. 1429–1442.

Köpke, E., Tung, Y. C., Shaikh, S., Alonso, A. D. C., Iqbal, K., & Grundke-Iqbal, I. (1993). Microtubule-associated protein tau. Abnormal phosphorylation of a non-paired helical filament pool in Alzheimer disease. *Journal of Biological Chemistry*, *268*(32), 24374-24384.

Kowall, N. W., & Kosik, K. S. (1987). Axonal disruption and aberrant localization of tau protein characterize the neuropil pathology of Alzheimer's disease. *Annals of neurology*, *22*(5), 639-643.

Kral, V. A. (1962). Senescent forgetfulness: benign and malignant. *Canadian Medical Association Journal*, *86*(6), 257.

Kruggel, F., Turner, J., Muftuler, L. T., & Alzheimer's Disease Neuroimaging Initiative. (2010). Impact of scanner hardware and imaging protocol on image quality and compartment volume precision in the ADNI cohort. *Neuroimage*, *49*(3), 2123-2133.

Kueper, J. K., Speechley, M., & Montero-Odasso, M. (2018). The Alzheimer’s Disease Assessment Scale–Cognitive Subscale (ADAS-Cog): Modifications and Responsiveness in Pre-Dementia Populations. A Narrative Review. *Journal of Alzheimer's Disease*, (Preprint), 1-22.

Kuhl, D. E., Metter, E. J., Riege, W. H., & Phelps, M. E. (1982). Effects of human aging on patterns of local cerebral glucose utilization determined by the [18F] fluorodeoxyglucose method. *Journal of Cerebral Blood Flow & Metabolism*, *2*(2), 163-171.

Kuhn, P. H., Koroniak, K., Hogl, S., Colombo, A., Zeitschel, U., Willem, M., ... & Haass, C. (2012). Secretome protein enrichment identifies physiological

BACE1 protease substrates in neurons. *The EMBO journal*, *31*(14), 3157-3168.

Kulisevsky, J., & Pagonabarraga, J. (2009). Cognitive impairment in Parkinson's disease: tools for diagnosis and assessment. *Movement disorders*, *24*(8), 1103-1110.

Kummer, M. P., & Heneka, M. T. (2014). Truncated and modified amyloid-beta species. *Alzheimer's research & therapy*, *6*(3), 28.

Kuperstein, I., Broersen, K., Benilova, I., Rozenski, J., Jonckheere, W., Debulpaep, M., ... & Braeken, D. (2010). Neurotoxicity of Alzheimer's disease Aβ peptides is induced by small changes in the Aβ42 to Aβ40 ratio. *The EMBO journal*, *29*(19), 3408-3420.

Kvartsberg, H., Duits, F. H., Ingelsson, M., Andreasen, N., Öhrfelt, A., Andersson, K., ... & Andreasson, U. (2015). Cerebrospinal fluid levels of the synaptic protein neurogranin correlates with cognitive decline in prodromal Alzheimer's disease. *Alzheimer's & Dementia*, *11*(10), 1180-1190.

L

Lahoz, C., Schaefer, E. J., Cupples, L. A., Wilson, P. W., Levy, D., Osgood, D., ... & Ordovas, J. M. (2001). Apolipoprotein E genotype and cardiovascular disease in the Framingham Heart Study. *Atherosclerosis*, *154*(3), 529-537.

Lambert, J. C., Heath, S., Even, G., Campion, D., Sleegers, K., Hiltunen, M., ... & Letenneur, L. (2009). Genome-wide association study identifies variants at CLU and CR1 associated with Alzheimer's disease. *Nature genetics*, *41*(10), 1094.

Lambert, J. C., Ibrahim-Verbaas, C. A., Harold, D., Naj, A. C., Sims, R., Bellenguez, C., ... & Grenier-Boley, B. (2013). Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. *Nature genetics*, *45*(12), 1452.

Landau, S. M., Harvey, D., Madison, C. M., Koeppe, R. A., Reiman, E. M., Foster, N. L., ... & Alzheimer's Disease Neuroimaging Initiative. (2011). Associations between cognitive, functional, and FDG-PET measures of decline in AD and MCI. *Neurobiology of aging*, *32*(7), 1207-1218.

Landau, S. M., Mintun, M. A., Joshi, A. D., Koeppe, R. A., Petersen, R. C., Aisen, P. S., ... & Alzheimer's Disease Neuroimaging Initiative. (2012). Amyloid deposition, hypometabolism, and longitudinal cognitive decline. *Annals of neurology*, *72*(4), 578-586.

Landau, S. M., Fero, A., Baker, S. L., Koeppe, R., Mintun, M., Chen, K., ... & Jagust, W. J. (2015). Measurement of longitudinal β-amyloid change with 18F-florbetapir PET and standardized uptake value ratios. *Journal of Nuclear Medicine*, *56*(4), 567-574.

Lasser, R., Ostrowitzki, S., Scheltens, P., Boada, M., Dubois, B., Dorflinger, E., ... & Edgar, C. (2015). Efficacy and safety of gantenerumab in prodromal Alzheimer's disease: Results from scarlet road—a global, multicenter trial. *Alzheimer's & Dementia: The Journal of the Alzheimer's Association*, *11*(7), P331-P332.

Lee, V. M., Balin, B. J., Otvos, L., & Trojanowski, J. Q. (1991). A68: a major subunit of paired helical filaments and derivatized forms of normal Tau. *Science*, *251*(4994), 675-678.

Lee, M. K., Xu, Z., Wong, P. C., & Cleveland, D. W. (1993). Neurofilaments are obligate heteropolymers in vivo. *The Journal of cell biology*, *122*(6), 1337-1350.

Lee, S. M., & McCarthy, G. (2016). Functional heterogeneity and convergence in the right temporoparietal junction. *Cerebral Cortex*, *26*(3), 1108-1116.

Lemere, C. A., & Masliah, E. (2010). Can Alzheimer disease be prevented by amyloid-β immunotherapy?. *Nature Reviews Neurology*, *6*(2), 108.

Lezak, M. D. (1995). Executive functions and motor performance. *In neuropsychological assessment*, 650-685.

Li, B., Chohan, M. O., Grundke-Iqbal, I., & Iqbal, K. (2007). Disruption of microtubule network by Alzheimer abnormally hyperphosphorylated tau. *Acta neuropathologica*, *113*(5), 501-511.

Li, C., & Götz, J. (2017). Tau-based therapies in neurodegeneration: opportunities and challenges. *Nature Reviews Drug Discovery*, *16*(12), 863.

Li, C., Loewenstein, D. A., Duara, R., Cabrerizo, M., Barker, W., & Adjouadi, M. (2017). The relationship of brain amyloid load and APOE status to regional cortical thinning and cognition in the ADNI Cohort. *Journal of Alzheimer's Disease*, *59*(4), 1269-1282.

Li, Z., Shue, F., Zhao, N., Shinohara, M., & Bu, G. (2020). APOE2: protective mechanism and therapeutic implications for Alzheimer's disease. *Molecular neurodegeneration*, *15*(1), 1-19.

Libon, D. J., Xie, S. X., Eppig, J., Wicas, G., Lamar, M., Lippa, C., ... & Wambach, D. M. (2010). The heterogeneity of mild cognitive impairment: A

neuropsychological analysis. *Journal of the International Neuropsychological Society*, *16*(1), 84-93.

Lim, Y. Y., Ellis, K. A., Harrington, K., Ames, D., Martins, R. N., Masters, C. L., ... & AIBL Research Group. (2012). Use of the CogState Brief Battery in the assessment of Alzheimer's disease related cognitive impairment in the Australian Imaging, Biomarkers and Lifestyle (AIBL) study. *Journal of clinical and experimental neuropsychology*, *34*(4), 345-358.

Lim, Y. Y., Maruff, P., Pietrzak, R. H., Ames, D., Ellis, K. A., Harrington, K., ... & Villemagne, V. L. (2013). Effect of amyloid on memory and non-memory decline from preclinical to clinical Alzheimer’s disease. *Brain*, *137*(1), 221-231.

Lim, H. K., Nebes, R., Snitz, B., Cohen, A., Mathis, C., Price, J., ... & Aizenstein, H. J. (2014). Regional amyloid burden and intrinsic connectivity networks in cognitively normal elderly subjects. *Brain*, *137*(12), 3327-3338.

Lim, Y. Y., Villemagne, V. L., Pietrzak, R. H., Ames, D., Ellis, K. A., Harrington, K., ... & Maruff, P. (2015). APOE ε4 moderates amyloid-related memory decline in preclinical Alzheimer's disease. *Neurobiology of aging*, *36*(3), 1239-1244.

Lim, Y. Y., Baker, J. E., Bruns, L., Mills, A., Fowler, C., Fripp, J., ... & Maruff, P. (2020). Association of deficits in short-term learning and Aβ and hippocampal volume in cognitively normal adults. *Neurology*, *95*(18), e2577-e2585.

Lindenberger, U., & Baltes, P. B. (1997). Intellectual functioning in old and very old age: cross-sectional results from the Berlin Aging Study. *Psychology and aging*, *12*(3), 410.

Lista, S., & Hampel, H. (2017). Synaptic degeneration and neurogranin in the pathophysiology of Alzheimer’s disease. *Expert review of neurotherapeutics*, *17*(1), 47-57.

Liu, C. C., Kanekiyo, T., Xu, H., & Bu, G. (2013). Apolipoprotein E and Alzheimer disease: risk, mechanisms and therapy. *Nature Reviews Neurology*, *9*(2), 106.

Llano, D. A., Laforet, G., & Devanarayan, V. (2011). Derivation of a new ADAS-cog composite using tree-based multivariate analysis: prediction of conversion from mild cognitive impairment to Alzheimer disease. *Alzheimer Disease & Associated Disorders*, *25*(1), 73-84.

Logsdon, R. G., Gibbons, L. E., McCurry, S. M., & Teri, L. (1999). Quality of life in Alzheimer's disease: patient and caregiver reports. *Journal of Mental health and Aging*, *5*, 21-32.

Logsdon, R. G., Gibbons, L. E., McCurry, S. M., & Teri, L. (2002). Assessing quality of life in older adults with cognitive impairment. *Psychosomatic medicine*, *64*(3), 510-519.

Loitfelder, M., Pinter, D., Langkammer, C., Jehna, M., Ropele, S., Fazekas, F., ... & Enzinger, C. (2013). Connectivity patterns obtained by emulated vs. conventional resting state fMRI in clinical cohorts—/INS; Can parts tell the whole story?. *Journal of the Neurological Sciences*, *333*, e388.

Lopera, F., Ardilla, A., Martínez, A., Madrigal, L., Arango-Viana, J. C., Lemere, C. A., ... & Behrens, I. M. (1997). Clinical features of early-onset Alzheimer disease in a large kindred with an E280A presenilin-1 mutation. *Jama*, *277*(10), 793-799.

López, M., Ramírez, J., Górriz, J. M., Salas-Gonzalez, D., Alvarez, I., Segovia, F., & Puntonet, C. G. (2009). Automatic tool for Alzheimer's disease diagnosis using PCA and Bayesian classification rules. *Electronics letters*, *45*(8), 389-391.

Lourida, I., Soni, M., Thompson-Coon, J., Purandare, N., Lang, I. A., Ukoumunne, O. C., & Llewellyn, D. J. (2013). Mediterranean diet, cognitive function, and dementia: a systematic review. *Epidemiology*, *24*(4), 479-489.

Lowe, D. A., Balsis, S., Miller, T. M., Benge, J. F., & Doody, R. S. (2012). Greater precision when measuring dementia severity: establishing item parameters for the clinical dementia rating scale. *Dementia and geriatric cognitive disorders*, *34*(2), 128-134.

Lowe, V. J., Curran, G., Fang, P., Liesinger, A. M., Josephs, K. A., Parisi, J. E., ... & Knopman, D. S. (2016). An autoradiographic evaluation of AV-1451 Tau PET in dementia. *Acta neuropathologica communications*, *4*(1), 58.

Lu, K., Nicholas, J. M., Pertzov, Y., Grogan, J., Husain, M., Pavisic, I. M., ... & Crutch, S. J. (2021). Dissociable effects of APOE ε4 and β-amyloid pathology on visual working memory. *Nature aging*, *1*(11), 1002-1009.

Lucas, J. A., Ivnik, R. J., Smith, G. E., Bohac, D. L., Tangalos, E. G., Kokmen, E., ... & Petersen, R. C. (1998). Normative data for the Mattis dementia rating scale. *Journal of clinical and experimental neuropsychology*, *20*(4), 536-547.

Lue, L. F., Schmitz, C., & Walker, D. G. (2015). What happens to microglial TREM2 in Alzheimer's disease: Immunoregulatory turned into immunopathogenic?. *Neuroscience*, *302*, 138-150.

M

Ma, L., Allen, M., Sakae, N., Ertekin-Taner, N., Graff-Radford, N. R., Dickson, D. W., ... & Sevlever, D. (2016). Expression and processing analyses of wild type and p. R47H TREM2 variant in Alzheimer's disease brains. *Molecular neurodegeneration*, *11*(1), 72.

Maass, A., Lockhart, S. N., Harrison, T. M., Bell, R. K., Mellinger, T., Swinnerton, K., ... & Jagust, W. J. (2018). Entorhinal tau pathology, episodic memory decline, and neurodegeneration in aging. *Journal of Neuroscience*, *38*(3), 530-543.

Majid O, Gee M, Albala B, et al. Dose-related reductions of CSF amyloid β (1-x) by E2609, a novel BACE1 inhibitor, in patients with mild cognitive impairment due to Alzheimer's disease (AD) and mild-to-moderate AD dementia. CTAD 2016;San Diego: P3-28 (poster presentation).

Mahley, R. W., & Rall Jr, S. C. (2000). Apolipoprotein E: far more than a lipid transport protein. *Annual review of genomics and human genetics*, *1*(1), 507-537.

Malouf, R., & Evans, J. G. (2008). Folic acid with or without vitamin B12 for the prevention and treatment of healthy elderly and demented people. *Cochrane Database of Systematic Reviews*, (4).

Mann, D. M. A., Iwatsubo, T., Cairns, N. J., Lantos, P. L., Nochlin, D., Sumi, S. M., ... & Prihar, G. (1996). Amyloid β protein (Aβ) deposition in chromosome 14–linked Alzheimer's disease: Predominance of Aβ42 (43). *Annals of Neurology: Official Journal of the American Neurological Association and the Child Neurology Society*, *40*(2), 149-156.

Marcus, C., Mena, E., & Subramaniam, R. M. (2014). Brain PET in the diagnosis of Alzheimer's disease. *Clinical nuclear medicine*, *39*(10), e413.

Maroco, J., Silva, D., Rodrigues, A., Guerreiro, M., Santana, I., & de Mendonça, A. (2011). Data mining methods in the prediction of Dementia: A real-data comparison of the accuracy, sensitivity and specificity of linear discriminant analysis, logistic regression, neural networks, support vector machines, classification trees and random forests. *BMC research notes*, *4*(1), 1-14.

Marquié, M., Normandin, M. D., Vanderburg, C. R., Costantino, I. M., Bien, E. A., Rycyna, L. G., ... & Vasdev, N. (2015). Validating novel tau positron

emission tomography tracer [F-18]-AV-1451 (T807) on postmortem brain tissue. *Annals of neurology*, *78*(5), 787-800.

Marquié, M., Normandin, M. D., Meltzer, A. C., Siao Tick Chong, M., Andrea, N. V., Antón-Fernández, A., ... & Bien, E. A. (2017a). Pathological correlations of [F-18]-AV-1451 imaging in non-alzheimer tauopathies. *Annals of neurology*, *81*(1), 117-128.

Marquié, M., Verwer, E. E., Meltzer, A. C., Kim, S. J. W., Agüero, C., Gonzalez, J., ... & Normandin, M. D. (2017b). Lessons learned about [F-18]-AV-1451 off-target binding from an autopsy-confirmed Parkinson's case. *Acta neuropathologica communications*, *5*(1), 75.

Maruff, P., Lim, Y. Y., Darby, D., Ellis, K. A., Pietrzak, R. H., Snyder, P. J., ... & Masters, C. L. (2013). Clinical utility of the cogstate brief battery in identifying cognitive impairment in mild cognitive impairment and Alzheimer's disease. *BMC psychology*, *1*(1), 1-11.

Masliah, E., Mallory, M., Alford, M., DeTeresa, R., Hansen, L. A., McKeel, D. W., & Morris, J. C. (2001). Altered expression of synaptic proteins occurs early during progression of Alzheimer's disease. *Neurology*, *56*(1), 127-129.

Matarin M, Salih DA, Yasvoina M, Cummings DM, Guelfi S, Liu W, Nahaboo Solim MA, Moens TG, Paublete RM, Ali SS *et al* (2015) A genome-wide gene-expression analysis and database in transgenic mice during development of amyloid or tau pathology. Cell Rep 10: 633–644

Mathis, C. A., Bacskai, B. J., Kajdasz, S. T., McLellan, M. E., Frosch, M. P., Hyman, B. T., ... & Klunk, W. E. (2002). A lipophilic thioflavin-T derivative for positron emission tomography (PET) imaging of amyloid in brain. *Bioorganic & medicinal chemistry letters*, *12*(3), 295-298.

Mathotaarachchi, S. *et al.* (2017) "Identifying incipient dementia individuals using machine learning and amyloid imaging," *Neurobiology of Aging*, 59, pp. 80–90.

Mathuranath, P. S., Nestor, P. J., Berrios, G. E., Rakowicz, W., & Hodges, J. R. (2000). A brief cognitive test battery to differentiate Alzheimer's disease and frontotemporal dementia. *Neurology*, *55*(11), 1613-1620.

Mattis, S. (1976). Dementia Rating Scale Manual. Odessa, FL: Psychological Assessment Resources.

Mattis, S. (1988). *Dementia rating scale: professional manual*. Psychological Assessment Resources, Incorporated.

Mattsson, N. (2011). CSF biomarkers in neurodegenerative diseases. *Clinical chemistry and laboratory medicine*, *49*(3), 345-352.

Mattsson, N., Zetterberg, H., Hansson, O., Andreasen, N., Parnetti, L., Jonsson, M., ... & Rich, K. (2009). CSF biomarkers and incipient Alzheimer disease in patients with mild cognitive impairment. *Jama*, *302*(4), 385-393.

Mattsson, N., Andreasson, U., Persson, S., Carrillo, M. C., Collins, S., Chalbot, S., ... & Hsiung, G. Y. R. (2013). CSF biomarker variability in the Alzheimer's Association quality control program. *Alzheimer's & Dementia*, *9*(3), 251-261.

Mattsson, N., Insel, P. S., Donohue, M., Landau, S., Jagust, W. J., Shaw, L. M., ... & Alzheimer's Disease Neuroimaging Initiative*. (2014). Independent information from cerebrospinal fluid amyloid-β and florbetapir imaging in Alzheimer's disease. *Brain*, *138*(3), 772-783.

Mattsson, N., Andreasson, U., Zetterberg, H., & Blennow, K. (2017). Association of plasma neurofilament light with neurodegeneration in patients with Alzheimer disease. *JAMA neurology*, *74*(5), 557-566.

Mattsson-Carlgren, N., Andersson, E., Janelidze, S., Ossenkoppele, R., Insel, P., Strandberg, O., ... & Hansson, O. (2020). Aβ deposition is associated with increases in soluble and phosphorylated tau that precede a positive Tau PET in Alzheimer's disease. *Science advances*, *6*(16), eaaz2387.

Maurer K, Volk S, Gerbaldo H. Auguste D. and Alzheimer's disease. Lancet. 1997;349:1546-1549

McCarthy, F., Burns, W. J., & Sellers, A. H. (2005). Discrepancies between premorbid and current IQ as a function of progressive mental deterioration. *Perceptual and motor skills*, *100*(1), 69-76.

McDade, E., Wang, G., Gordon, B. A., Hassenstab, J., Benzinger, T. L., Buckles, V., ... & Marcus, D. S. (2018). Longitudinal cognitive and biomarker changes in dominantly inherited Alzheimer disease. *Neurology*, *91*(14), e1295-e1306.

McDonough, I. M., Bischof, G. N., Kennedy, K. M., Rodrigue, K. M., Farrell, M. E., & Park, D. C. (2016). Discrepancies between fluid and crystallized ability in healthy adults: a behavioral marker of preclinical Alzheimer's disease. *Neurobiology of aging*, *46*, 68-75.

McDonough, I. M., & Popp, T. E. (2020). Linear and nonlinear relationships between cognitive subdomains of ability discrepancy and Alzheimer's disease biomarkers. *Neuropsychology*, *34*(2), 211.

McGeer, P. L., Akiyama, H., Itagaki, S., & McGeer, E. G. (1989). Immune system response in Alzheimer's disease. *Canadian Journal of Neurological Sciences*, *16*(S4), 516-527.

McHugh, R. B., & Owens, W. A. (1954). Age changes in mental organization—a longitudinal study. *Journal of Gerontology*.

McKhann, G., Drachman, D., Folstein, M., Katzman, R., Price, D., & Stadlan, E. M. (1984). Clinical diagnosis of Alzheimer's disease: Report of the NINCDS-ADRDA Work Group* under the auspices of Department of Health and Human Services Task Force on Alzheimer's Disease. *Neurology*, *34*(7), 939-939.

McKhann, G. M., Knopman, D. S., Chertkow, H., Hyman, B. T., Jack Jr, C. R., Kawas, C. H., ... & Mohs, R. C. (2011). The diagnosis of dementia due to 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), 263-269.

McShane, R., Sastre, A. A., & Minakaran, N. (2006). Memantine for dementia. *Cochrane Database of Systematic Reviews*, (2).

Medina, M. (2018). An Overview on the clinical development of tau-based therapeutics. *International journal of molecular sciences*, *19*(4), 1160.

Medina, M., & Avila, J. (2015). Further understanding of tau phosphorylation: implications for therapy. *Expert review of neurotherapeutics*, *15*(1), 115-122.

Mielke, M. M., Vemuri, P., & Rocca, W. A. (2014). Clinical epidemiology of Alzheimer’s disease: assessing sex and gender differences. *Clinical epidemiology*, *6*, 37.

Milà-Alomà, M., Ashton, N. J., Shekari, M., Salvadó, G., Ortiz-Romero, P., Montoliu-Gaya, L., ... & Blennow, K. (2022). Plasma p-tau231 and p-tau217 as state markers of amyloid-β pathology in preclinical Alzheimer’s disease. *Nature medicine*, 1-5.

Minoshima, S., Giordani, B., Berent, S., Frey, K. A., Foster, N. L., & Kuhl, D. E. (1997). Metabolic reduction in the posterior cingulate cortex in very early Alzheimer's disease. *Annals of Neurology: Official Journal of the American Neurological Association and the Child Neurology Society*, *42*(1), 85-94.

Mistridis, P., Krumm, S., Monsch, A. U., Berres, M., & Taylor, K. I. (2015). The 12 years preceding mild cognitive impairment due to Alzheimer’s disease: the temporal emergence of cognitive decline. *Journal of Alzheimer's Disease*, *48*(4), 1095-1107.

Mitchell, A. J. (2009). A meta-analysis of the accuracy of the mini-mental state examination in the detection of dementia and mild cognitive impairment. *Journal of psychiatric research*, *43*(4), 411-431.

Mohs, R. C., Knopman, D., Petersen, R. C., Ferris, S. H., Ernesto, C., Grundman, M., ... & Thal, L. J. (1997). Development of cognitive instruments for use in clinical trials of antidementia drugs: additions to the Alzheimer's Disease Assessment Scale that broaden its scope. *Alzheimer disease and associated disorders*.

Mons, N., Enderlin, V., Jaffard, R., & Higueret, P. (2001). Selective age-related changes in the PKC-sensitive, calmodulin-binding protein, neurogranin, in the mouse brain. *Journal of neurochemistry*, *79*(4), 859-867.

Montine, T. J., Phelps, C. H., Beach, T. G., Bigio, E. H., Cairns, N. J., Dickson, D. W., ... & Nelson, P. T. (2012). National Institute on Aging–Alzheimer's Association guidelines for the neuropathologic assessment of Alzheimer's disease: a practical approach. *Acta neuropathologica*, *123*(1), 1-11.

Montoliu-Gaya, L., & Villegas, S. (2016). Aβ-Immunotherapeutic strategies: a wide range of approaches for Alzheimer's disease treatment. *Expert reviews in molecular medicine*, *18*.

Mortamais, M., Ash, J. A., Harrison, J., Kaye, J., Kramer, J., Randolph, C., ... & Ritchie, K. (2017). Detecting cognitive changes in preclinical Alzheimer's disease: A review of its feasibility. *Alzheimer's & dementia*, *13*(4), 468-492.
Mormino, E. C., Smiljic, A., Hayenga, A. O., H. Onami, S., Greicius, M. D., Rabinovici, G. D., ... & Miller, B. L. (2011). Relationships between beta-amyloid and functional connectivity in different components of the default mode network in aging. *Cerebral cortex*, *21*(10), 2399-2407.

Mormino, E. C., Betensky, R. A., Hedden, T., Schultz, A. P., Ward, A., Huijbers, W., ... & Sperling, R. A. (2014). Alzheimer's Disease Neuroimaging I, Australian Imaging B, Lifestyle Flagship Study of A, Harvard Aging Brain S. Amyloid and APOE epsilon4 interact to influence short-term decline in preclinical Alzheimer disease. *Neurology*, *82*, 1760-1767.

Morris, J., Kunka, J. M., & Rossini, E. D. (1997). Development of alternate paragraphs for the Logical Memory subtest of the Wechsler Memory Scale-Revised. *The Clinical Neuropsychologist*, *11*(4), 370-374.

Morris, J., Swier-Vosnos, A., Woodworth, C., Umfleet, L. G., Czipri, S., & Kopald, B. (2014). Development of alternate paragraphs for the Logical Memory subtest of the Wechsler Memory Scale-IV. *Applied Neuropsychology: Adult*, *21*(2), 143-147.

Morris, J. C. (1993). The Clinical Dementia Rating (CDR): current version and scoring rules. *Neurology*.

Morris, J. C. (1997). Clinical dementia rating: a reliable and valid diagnostic and staging measure for dementia of the Alzheimer type. *International psychogeriatrics*, *9*(S1), 173-176.

Morris, J. C., Heyman, A., Mohs, R. C., Hughes, J. P., Van Belle, G., Fillenbaum, G. D. M. E., ... & Clark, C. (1989). The consortium to establish a registry for Alzheimer's disease (CERAD): I. Clinical and neuropsychological assessment of Alzheimer's disease. *Neurology*.

Morris, J. C., Roe, C. M., Xiong, C., Fagan, A. M., Goate, A. M., Holtzman, D. M., & Mintun, M. A. (2010). APOE predicts amyloid-beta but not tau Alzheimer pathology in cognitively normal aging. *Annals of neurology*, *67*(1), 122-131.

Morris, E., Chalkidou, A., Hammers, A., Peacock, J., Summers, J., & Keevil, S. (2016). Diagnostic accuracy of 18 F amyloid PET tracers for the diagnosis of Alzheimer's disease: a systematic review and meta-analysis. *European journal of nuclear medicine and molecular imaging*, *43*(2), 374-385.

Mortamais, M., Ash, J. A., Harrison, J., Kaye, J., Kramer, J., Randolph, C., ... & Ritchie, K. (2017). Detecting cognitive changes in preclinical Alzheimer's disease: A review of its feasibility. *Alzheimer's & Dementia*, *13*(4), 468-492.

Mosconi, L., Tsui, W. H., Herholz, K., Pupi, A., Drzezga, A., Lucignani, G., ... & Diehl-Schmid, J. (2008). Multicenter standardized 18F-FDG PET diagnosis of mild cognitive impairment, Alzheimer's disease, and other dementias. *Journal of nuclear medicine*, *49*(3), 390-398.

Moyer, V. A. (2014). Screening for cognitive impairment in older adults: US Preventive Services Task Force recommendation statement. *Annals of internal medicine*, *160*(11), 791-797.

Mueller, S. G., Weiner, M. W., Thal, L. J., Petersen, R. C., Jack, C. R., Jagust, W., ... & Beckett, L. (2005). Ways toward an early diagnosis in Alzheimer's disease: the Alzheimer's Disease Neuroimaging Initiative (ADNI). *Alzheimer's & Dementia*, *1*(1), 55-66.

Mullard, A. (2018). BACE failures lower AD expectations, again. *Nature reviews. Drug discovery*, *17*(6), 385.

Murphy, J., Doherty, T., Gee, M., Ito, S., Kanekiyo, M. & Albala, B. (2019). ASIAN AND NON-ASIAN COUNTRIES SCREEN SUBJECTS WITH SIMILAR MMSE SCORES TO THE ELENBECESTAT MISSIONAD GLOBAL PHASE 3 STUDIES IN

EARLY ALZHEIMER’S DISEASE. The Journal of Prevention of Alzheimer’s Disease - JPAD© Volume 6, Supplement 1, P149

N

N'Diaye, E. N., Branda, C. S., Branda, S. S., Nevarez, L., Colonna, M., Lowell, C., ... & Seaman, W. E. (2009). TREM-2 (triggering receptor expressed on myeloid cells 2) is a phagocytic receptor for bacteria. *J Cell Biol*, *184*(2), 215-223.

Naj, A. C., Jun, G., Beecham, G. W., Wang, L. S., Vardarajan, B. N., Buros, J., ... & Larson, E. B. (2011). Common variants at MS4A4/MS4A6E, CD2AP, CD33 and EPHA1 are associated with late-onset Alzheimer's disease. *Nature genetics*, *43*(5), 436.

Nasreddine, Z. S., Phillips, N. A., Bédirian, V., Charbonneau, S., Whitehead, V., Collin, I., ... & Chertkow, H. (2005). The Montreal Cognitive Assessment, MoCA: a brief screening tool for mild cognitive impairment. *Journal of the American Geriatrics Society*, *53*(4), 695-699.

Nathan, P. J., Lim, Y. Y., Abbott, R., Galluzzi, S., Marizzoni, M., Babiloni, C., ... & PharmaCog Consortium. (2017). Association between CSF biomarkers, hippocampal volume and cognitive function in patients with amnestic mild cognitive impairment (MCI). *Neurobiology of Aging*, *53*, 1-10.

Neff, R. A., Wang, M., Vatansever, S., Guo, L., Ming, C., Wang, Q., ... & Zhang, B. (2021). Molecular subtyping of Alzheimer’s disease using RNA sequencing data reveals novel mechanisms and targets. *Science advances*, *7*(2).

Nelson, P. T., Alafuzoff, I., Bigio, E. H., Bouras, C., Braak, H., Cairns, N. J., ... & Duyckaerts, C. (2012). Correlation of Alzheimer disease neuropathologic changes with cognitive status: a review of the literature. *Journal of Neuropathology & Experimental Neurology*, *71*(5), 362-381.

Nelson, P. T., Dickson, D. W., Trojanowski, J. Q., Jack, C. R., Boyle, P. A., Arfanakis, K., ... & Schneider, J. A. (2019). Limbic-predominant age-related TDP-43 encephalopathy (LATE): consensus working group report. *Brain*, *142*(6), 1503-1527.

Neumann, H., & Takahashi, K. (2007). Essential role of the microglial triggering receptor expressed on myeloid cells-2 (TREM2) for central nervous tissue immune homeostasis. *Journal of neuroimmunology*, *184*(1-2), 92-99.

Newcomer JW, Craft S, Hershey T, Askins K, & Bardgett ME. (1994). Glucocorticoid-induced impairment in declarative memory performance in adult humans. The Journal of Neuroscience. 14(4), 2047–2053.

Ng, K. P., Pascoal, T. A., Mathotaarachchi, S., Therriault, J., Kang, M. S., Shin, M., ... & Massarweh, G. (2017). Monoamine oxidase B inhibitor, selegiline, reduces 18 F-THK5351 uptake in the human brain. *Alzheimer's research & therapy*, *9*(1), 25.

Nichols, H. K., & Basu, T. K. (1994). Thiamin status of the elderly: dietary intake and thiamin pyrophosphate response. *Journal of the American College of Nutrition*, *13*(1), 57-61.

Nickerson, R. S. (2000). Null hypothesis significance testing: a review of an old and continuing controversy. *Psychological methods*, *5*(2), 241.

Novak, M., Jakes, R., Edwards, P. C., Milstein, C., & Wischik, C. M. (1991). Difference between the tau protein of Alzheimer paired helical filament core and normal tau revealed by epitope analysis of monoclonal antibodies 423 and 7.51. *Proceedings of the National Academy of Sciences of the United States of America*, *88*(13), 5837.

Nuttall, R., Pasquini, L., Scherr, M., Sorg, C., & Alzheimer's Disease Neuroimaging Initiative. (2016). Degradation in intrinsic connectivity networks across the Alzheimer's disease spectrum. *Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring*, *5*, 35-42.

Nygaard, H. B., Wagner, A. F., Bowen, G. S., Good, S. P., MacAvoy, M. G., Strittmatter, K. A., ... & Chen, K. (2015). A phase Ib multiple ascending dose study of the safety, tolerability, and central nervous system availability of AZD0530 (saracatinib) in Alzheimer’s disease. *Alzheimer's research & therapy*, *7*(1), 35.

Nyunt, M. S. Z., Chong, M. S., Lim, W. S., Lee, T. S., Yap, P., & Ng, T. P. (2013). Reliability and validity of the Clinical Dementia Rating for community-living elderly subjects without an informant. *Dementia and geriatric cognitive disorders extra*, *3*(1), 407-416.

O

O'Carroll, R. E., & Gilleard, C. J. (1986). Estimation of premorbid intelligence in dementia. *British Journal of Clinical Psychology*, *25*(2), 157-158.

O'Donoghue, M. C., Murphy, S. E., Zamboni, G., Nobre, A. C., & Mackay, C. E. (2018). APOE genotype and cognition in healthy individuals at risk of Alzheimer's disease: A review. *Cortex*, *104*, 103-123.

O’Shea, D. M., Fieo, R., Woods, A., Williamson, J., Porges, E., & Cohen, R. (2018). Discrepancies between crystallized and fluid ability are associated

with frequency of social and physical engagement in community dwelling older adults. *Journal of Clinical and Experimental Neuropsychology*, *40*(10), 963-970.

Okamura, N., Harada, R., Ishiki, A., Kikuchi, A., Nakamura, T., & Kudo, Y. (2018). The development and validation of tau PET tracers: current status and future directions. *Clinical and translational imaging*, *6*(4), 305-316.

Oltra-Cucarella, J., Ferrer-Cascales, R., Alegret, M., Gasparini, R., Díaz-Ortiz, L. M., Rios, R., ... & Sanchez-SanSegundo, M. (2018). Risk of progression to Alzheimer's disease for different neuropsychological Mild Cognitive Impairment subtypes: A hierarchical meta-analysis of longitudinal studies. *Psychology and aging*, *33*(7), 1007.

Oriol, J. D. V., Vallejo, E. E., Estrada, K., Peña, J. G. T., & Alzheimer's Disease Neuroimaging Initiative. (2019). Benchmarking machine learning models for late-onset alzheimer's disease prediction from genomic data. *BMC bioinformatics*, *20*.

Orru, G., Pettersson-Yeo, W., Marquand, A. F., Sartori, G., & Mechelli, A. (2012). Using support vector machine to identify imaging biomarkers of neurological and psychiatric disease: a critical review. *Neuroscience & Biobehavioral Reviews*, *36*(4), 1140-1152.

Ossenkoppele, R., Smith, R., Ohlsson, T., Strandberg, O., Mattsson, N., Insel, P. S., ... & Hansson, O. (2019). Associations between tau, Aβ, and cortical thickness with cognition in Alzheimer disease. *Neurology*, *92*(6), e601-e612.

Ou-Yang, M. H., Kurz, J. E., Nomura, T., Popovic, J., Rajapaksha, T. W., Dong, H., ... & Vassar, R. (2018). Axonal organization defects in the hippocampus of adult conditional BACE1 knockout mice. *Science translational medicine*, *10*(459), eaao5620.

P

Palmqvist, S., Schöll, M., Strandberg, O., Mattsson, N., Stomrud, E., Zetterberg, H., ... & Hansson, O. (2017). Earliest accumulation of β-amyloid occurs within the default-mode network and concurrently affects brain connectivity. *Nature communications*, *8*(1), 1214.

Palmqvist, S., Insel, P. S., Stomrud, E., Janelidze, S., Zetterberg, H., Brix, B., ... & Hansson, O. (2019). Cerebrospinal fluid and plasma biomarker trajectories with increasing amyloid deposition in Alzheimer's disease. *EMBO molecular medicine*, *11*(12), e11170.

Panza, F., Lozupone, M., Solfrizzi, V., Sardone, R., Piccininni, C., Dibello, V., ... & Daniele, A. (2018). BACE inhibitors in clinical development for the treatment of Alzheimer's disease. *Expert review of neurotherapeutics*, *18*(11), 847-857.

Papp, K. V., Rentz, D. M., Orlovsky, I., Sperling, R. A., & Mormino, E. C. (2017). Optimizing the preclinical Alzheimer's cognitive composite with semantic processing: The PACC5. *Alzheimer's & Dementia: Translational Research & Clinical Interventions*, *3*(4), 668-677.

Pariente, A., Helmer, C., Merliere, Y., Moore, N., Fourrier-Réglat, A., & Dartigues, J. F. (2008). Prevalence of cholinesterase inhibitors in subjects with dementia in Europe. *Pharmacoepidemiology and drug safety*, *17*(7), 655-660.

Park, D. C., Lautenschlager, G., Hedden, T., Davidson, N. S., Smith, A. D., & Smith, P. K. (2002). Models of visuospatial and verbal memory across the adult life span. *Psychology and aging*, *17*(2), 299.

Park, L. Q., Gross, A. L., McLaren, D. G., Pa, J., Johnson, J. K., Mitchell, M., & Manly, J. J. (2012). Confirmatory factor analysis of the ADNI neuropsychological battery. *Brain imaging and behavior*, *6*(4), 528-539.

Patton, D. E., Duff, K., Schoenberg, M. R., Mold, J., Scott, J. G., & Adams, R. L. (2005). Base rates of longitudinal RBANS discrepancies at one-and two-year intervals in community-dwelling older adults. *The Clinical Neuropsychologist*, *19*(1), 27-44.

Pauwels, K., Williams, T. L., Morris, K. L., Jonckheere, W., Vandersteen, A., Kelly, G., ... & Broersen, K. (2012). Structural basis for increased toxicity of pathological aβ42: aβ40 ratios in Alzheimer disease. *Journal of biological chemistry*, *287*(8), 5650-5660.

Pedersen, J. T., & Sigurdsson, E. M. (2015). Tau immunotherapy for Alzheimer's disease. *Trends in molecular medicine*, *21*(6), 394-402.

Perneczky, R., Wagenpfeil, S., Komossa, K., Grimmer, T., Diehl, J., & Kurz, A. (2006). Mapping scores onto stages: mini-mental state examination and clinical dementia rating. *The American journal of geriatric psychiatry*, *14*(2), 139-144.

Perrotin, A., Mormino, E. C., Madison, C. M., Hayenga, A. O., & Jagust, W. J. (2012). Subjective cognition and amyloid deposition imaging: a Pittsburgh Compound B positron emission tomography study in normal elderly individuals. *Archives of neurology*, *69*(2), 223-229.

Petersen, R. C., Smith, G. E., Waring, S. C., Ivnik, R. J., Tangalos, E. G., & Kokmen, E. (1999). Mild cognitive impairment: clinical characterization and outcome. *Archives of neurology*, *56*(3), 303-308.

Petersen, R. C., Aisen, P. S., Beckett, L. A., Donohue, M. C., Gamst, A. C., Harvey, D. J., ... & Weiner, M. W. (2010). Alzheimer's disease neuroimaging initiative (ADNI): clinical characterization. *Neurology*, *74*(3), 201-209.

Petkova, A. T., Ishii, Y., Balbach, J. J., Antzutkin, O. N., Leapman, R. D., Delaglio, F., & Tycko, R. (2002). A structural model for Alzheimer's β-amyloid fibrils based on experimental constraints from solid state NMR. *Proceedings of the National Academy of Sciences*, *99*(26), 16742-16747.

Pettigrew, C., & Soldan, A. (2019). Defining Cognitive Reserve and Implications for Cognitive Aging. *Current neurology and neuroscience reports*, *19*(1), 1.

Piccio, L., Deming, Y., Del-Águila, J. L., Ghezzi, L., Holtzman, D. M., Fagan, A. M., ... & Cruchaga, C. (2016). Cerebrospinal fluid soluble TREM2 is higher in Alzheimer disease and associated with mutation status. *Acta neuropathologica*, *131*(6), 925-933.

Pietrzak, R. H., Lim, Y. Y., Neumeister, A., Ames, D., Ellis, K. A., Harrington, K., ... & Villemagne, V. L. (2015). Amyloid-β, anxiety, and cognitive decline in preclinical Alzheimer disease: a multicenter, prospective cohort study. *JAMA psychiatry*, *72*(3), 284-291.

Podhorna, J., Krahnke, T., Shear, M., & Harrison, J. E. (2016). Alzheimer's Disease Assessment Scale–Cognitive subscale variants in mild cognitive impairment and mild Alzheimer's disease: change over time and the effect of enrichment strategies. *Alzheimer's research & therapy*, *8*(1), 8.

Portelius, E., Zetterberg, H., Skillbäck, T., Törnqvist, U., Andreasson, U., Trojanowski, J. Q., ... & Blennow, K. (2015). Cerebrospinal fluid neurogranin: relation to cognition and neurodegeneration in Alzheimer's disease. *Brain*, *138*(11), 3373-3385.

Prince M, Wimo A, Guerchet M, Ali GYW, Prina M. World Alzheimer Report 2015. The global impact of dementia. An analysis of Prevalence, Incidence, Cost and Trends. London, 2015.

Prins, N. D., & Scheltens, P. (2013). Treating Alzheimer's disease with monoclonal antibodies: current status and outlook for the future. *Alzheimer's research & therapy*, *5*(6), 56.

Prokop, S., Miller, K. R., Drost, N., Handrick, S., Mathur, V., Luo, J., ... & Heppner, F. L. (2015). Impact of peripheral myeloid cells on amyloid-β pathology in Alzheimer's disease–like mice. *Journal of Experimental Medicine*, *212*(11), 1811-1818.

Proust-Lima, C., Amieva, H., Letenneur, L., Orgogozo, J. M., Jacqmin-Gadda, H., & Dartigues, J. F. (2008). Gender and education impact on brain aging: A general cognitive factor approach. *Psychology and aging*, *23*(3), 608.

Puzo, C., Labriola, C., Sugarman, M. A., Tripodis, Y., Martin, B., Palmisano, J. N., ... & Alosco, M. L. (2019). Independent effects of white matter hyperintensities on cognitive, neuropsychiatric, and functional decline: a longitudinal investigation using the National Alzheimer's Coordinating Center Uniform Data Set. *Alzheimer's Research & Therapy*, *11*(1), 1-13.

Pyo, G., Elble, R. J., Ala, T., & Markwell, S. J. (2006). The characteristics of patients with uncertain/mild cognitive impairment on the Alzheimer disease assessment scale-cognitive subscale. *Alzheimer Disease & Associated Disorders*, *20*(1), 16-22.

Q

Qi-Takahara, Y., Morishima-Kawashima, M., Tanimura, Y., Dolios, G., Hirotani, N., Horikoshi, Y., ... & Ihara, Y. (2005). Longer forms of amyloid β protein: implications for the mechanism of intramembrane cleavage by γ-secretase. *Journal of Neuroscience*, *25*(2), 436-445.

R

Rabbitt, P. (1993). Crystal quest: A search for the basis of maintenance of practised skills into old age.

Raber, J., Wong, D., Yu, G. Q., Buttini, M., Mahley, R. W., Pitas, R. E., & Mucke, L. (2000). Alzheimer's disease: Apolipoprotein E and cognitive performance. *Nature*, *404*(6776), 352.

Rabin, L. A., Saykin, A. J., West, J. D., Borgos, M. J., Wishart, H. A., Nutter-Upham, K. E., ... & Santulli, R. B. (2009). Judgment in older adults with normal cognition, cognitive complaints, MCI, and mild AD: Relation to regional frontal gray matter. *Brain imaging and behavior*, *3*(2), 212-219.

Rabin, J. S., Schultz, A. P., Hedden, T., Viswanathan, A., Marshall, G. A., Kilpatrick, E., ... & Rao, V. (2018). Interactive associations of vascular risk and β-amyloid burden with cognitive decline in clinically normal elderly individuals: findings from the Harvard Aging Brain Study. *JAMA neurology*, *75*(9), 1124-1131.

Raghavan, N., Samtani, M. N., Farnum, M., Yang, E., Novak, G., Grundman, M., ... & Alzheimer’s Disease Neuroimaging Initiative. (2013). The ADAS-Cog revisited: novel composite scales based on ADAS-Cog to improve efficiency in MCI and early AD trials. *Alzheimer's & Dementia*, *9*(1), S21-S31.

Rauchmann, B. S., Schneider-Axmann, T., Alexopoulos, P., Perneczky, R., & Alzheimer's Disease Neuroimaging Initiative. (2019). CSF soluble TREM2 as a measure of immune response along the Alzheimer's disease continuum. *Neurobiology of aging*, *74*, 182-190.

Rawle, M. J., Davis, D., Bendayan, R., Wong, A., Kuh, D., & Richards, M. (2018). Apolipoprotein-E (Apoe) ε4 and cognitive decline over the adult life course. *Translational psychiatry*, *8*(1), 1-8.

Reddy, P. H., Mani, G., Park, B. S., Jacques, J., Murdoch, G., Whetsell Jr, W., ... & Manczak, M. (2005). Differential loss of synaptic proteins in Alzheimer's disease: implications for synaptic dysfunction. *Journal of Alzheimer's Disease*, *7*(2), 103-117.

Reinert, G. (1970). Comparative factor analytic studies of intelligence throughout the human life-span. In *Life-span developmental psychology* (pp. 467-484). Academic Press.

Rentz, D. M., Locascio, J. J., Becker, J. A., Moran, E. K., Eng, E., Buckner, R. L., ... & Johnson, K. A. (2010). Cognition, reserve, and amyloid deposition in normal aging. *Annals of neurology*, *67*(3), 353-364.

Rentz, D. M., Mormino, E. C., Papp, K. V., Betensky, R. A., Sperling, R. A., & Johnson, K. A. (2017). Cognitive resilience in clinical and preclinical Alzheimer’s disease: the Association of Amyloid and Tau Burden on cognitive performance. *Brain imaging and behavior*, *11*(2), 383-390.

Rentz, D. M., & Weintraub, S. (2000). Neuropsychological detection of early probable Alzheimer’s disease. In *Early diagnosis of Alzheimer’s disease* (pp. 169-189). Humana Press, Totowa, NJ.

Rey, A. (1941). L’examen psychologique dans les cas d’enc´ephalopathie traumatique. Archives de Psychologie, 28, 21

Riedel, G., Platt, B., & Micheau, J. (2003). Glutamate receptor function in learning and memory. *Behavioural brain research*, *140*(1-2), 1-47.

Rinne, J. O., Wong, D. F., Wolk, D. A., Leinonen, V., Arnold, S. E., Buckley, C., ... & Kailajärvi, M. (2012). [18 F] Flutemetamol PET imaging and cortical biopsy histopathology for fibrillar amyloid β detection in living subjects with

normal pressure hydrocephalus: pooled analysis of four studies. *Acta neuropathologica*, *124*(6), 833-845.

Ritchie, K., Carrière, I., Berr, C., Amieva, H., Dartigues, J. F., Ancelin, M. L., & Ritchie, C. W. (2016). The clinical picture of Alzheimer's disease in the decade before diagnosis: clinical and biomarker trajectories. *The Journal of clinical psychiatry*, *77*(3), e305-11.

Robbins, T. W., James, M., Owen, A. M., Sahakian, B. J., McInnes, L., & Rabbitt, P. (1994). Cambridge Neuropsychological Test Automated Battery (CANTAB): a factor analytic study of a large sample of normal elderly volunteers. *Dementia and geriatric cognitive disorders*, *5*(5), 266-281.

Roberts, C., Kaplow, J., Giroux, M., Krause, S., & Kanekiyo, M. (2021). Amyloid and APOE status of screened subjects in the elenbecestat MissionAD phase 3 program. *The Journal of Prevention of Alzheimer's Disease*, *8*(2), 218-223.

Rockwood, K., Strang, D., MacKnight, C., Downer, R., & Morris, J. C. (2000). Interrater reliability of the Clinical Dementia Rating in a multicenter trial. *Journal of the American Geriatrics Society*, *48*(5), 558-559.

Roe, C. M., Mintun, M. A., D'Angelo, G., Xiong, C., Grant, E. A., & Morris, J. C. (2008). Alzheimer disease and cognitive reserve: variation of education effect with carbon 11–labeled Pittsburgh Compound B uptake. *Archives of neurology*, *65*(11), 1467-1471.

Roe, C. M., Fagan, A. M., Grant, E. A., Hassenstab, J., Moulder, K. L., Dreyfus, D. M., ... & Morris, J. C. (2013). Amyloid imaging and CSF biomarkers in predicting cognitive impairment up to 7.5 years later. *Neurology*, *80*(19), 1784-1791.

Rogers, S. L., Farlow, M. R., Doody, R. S., Mohs, R., & Friedhoff, L. T. (1998). A 24-week, double-blind, placebo-controlled trial of donepezil in patients with Alzheimer's disease. *Neurology*, *50*(1), 136-145.

Rosen, W. G., Mohs, R. C., & Davis, K. L. (1984). A new rating scale for Alzheimer's disease. *The American journal of psychiatry*.

Roses, A. D. (1996). Apolipoprotein E and Alzheimer's disease. A rapidly expanding field with medical and epidemiological consequences. *Annals of the New York Academy of Sciences*, *802*, 50-57.

Ross, T. P., Calhoun, E., Cox, T., Wenner, C., Kono, W., & Pleasant, M. (2007). The reliability and validity of qualitative scores for the Controlled Oral Word Association Test. *Archives of Clinical Neuropsychology*, *22*(4), 475-488.

Rowe, C. C., Ackerman, U., Browne, W., Mulligan, R., Pike, K. L., O'Keefe, G., ... & Dickinson-Rowe, K. L. (2008). Imaging of amyloid β in Alzheimer's disease with 18F-BAY94-9172, a novel PET tracer: proof of mechanism. *The Lancet Neurology*, *7*(2), 129-135.

Ryan, J. J., & Paolo, A. M. (1992). A screening procedure for estimating premorbid intelligence in the elderly. *The Clinical Neuropsychologist*, *6*(1), 53-62.

Ryan, C. L., & Siebens, J. (2012). Educational Attainment in the United States: 2009. Population Characteristics. Current Population Reports. P20-566. *US Census Bureau*.

S

Sabri, O., Sabbagh, M. N., Seibyl, J., Barthel, H., Akatsu, H., Ouchi, Y., ... & Beach, T. G. (2015). Florbetaben PET imaging to detect amyloid beta plaques in Alzheimer's disease: phase 3 study. *Alzheimer's & Dementia*, *11*(8), 964-974.

Saito, T., Matsuba, Y., Mihira, N., Takano, J., Nilsson, P., Itohara, S., ... & Saido, T. C. (2014). Single App knock-in mouse models of Alzheimer's disease. *Nature neuroscience*, *17*(5), 661.

Salas-Gonzalez, D., Górriz, J. M., Ramírez, J., López, M., Alvarez, I., Segovia, F., ... & Puntonet, C. G. (2010). Computer-aided diagnosis of Alzheimer's disease using support vector machines and classification trees. *Physics in Medicine & Biology*, *55*(10), 2807.

Salloway, S., Sperling, R., Fox, N. C., Blennow, K., Klunk, W., Raskind, M., ... & Reichert, M. (2014). Two phase 3 trials of bapineuzumab in mild-to-moderate Alzheimer's disease. *New England Journal of Medicine*, *370*(4), 322-333.

Salloway, S., Gamez, J. E., Singh, U., Sadowsky, C. H., Villena, T., Sabbagh, M. N., ... & Walker, Z. (2017). Performance of [18F] flutemetamol amyloid imaging against the neuritic plaque component of CERAD and the current (2012) NIA-AA recommendations for the neuropathologic diagnosis of Alzheimer's disease. *Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring*, *9*, 25-34.

Salmon, D. P., Ferris, S. H., Thomas, R. G., Sano, M., Cummings, J. L., Sperling, R. A., ... & Aisen, P. S. (2013). Age and apolipoprotein E genotype influence rate of cognitive decline in nondemented elderly. *Neuropsychology*, *27*(4), 391.

Salthouse, T. A. (2010). Influence of age on practice effects in longitudinal neurocognitive change. *Neuropsychology*, *24*(5), 563.

Sämgård, K., Zetterberg, H., Blennow, K., Hansson, O., Minthon, L., & Londos, E. (2010). Cerebrospinal fluid total tau as a marker of Alzheimer's disease intensity. *International Journal of Geriatric Psychiatry: A journal of the psychiatry of late life and allied sciences*, *25*(4), 403-410.

Sampson, D. L., Parker, T. J., Upton, Z., & Hurst, C. P. (2011). A comparison of methods for classifying clinical samples based on proteomics data: a case study for statistical and machine learning approaches. *PloS one*, *6*(9), e24973.

Samtani, M. N., Raghavan, N., Novak, G., Nandy, P., & Narayan, V. A. (2014). Disease progression model for clinical dementia rating–sum of boxes in mild cognitive impairment and Alzheimer’s subjects from the Alzheimer’s disease Neuroimaging initiative. *Neuropsychiatric disease and treatment*, *10*, 929.

Sancesario, G. M., Esposito, Z., Nuccetelli, M., Bernardini, S., Sorge, R., Martorana, A., ... & Sancesario, G. (2010). Aβ1–42 Detection in CSF of Alzheimer's disease is influenced by temperature: Indication of reversible Aβ1–42 aggregation?. *Experimental neurology*, *223*(2), 371-376.

Sankaranarayanan, S., Barten, D. M., Vana, L., Devidze, N., Yang, L., Cadelina, G., ... & Cao, Y. (2015). Passive immunization with phospho-tau antibodies reduces tau pathology and functional deficits in two distinct mouse tauopathy models. *PLoS One*, *10*(5), e0125614.

Saunders, A. M., Strittmatter, W. J., Schmechel, D., George-Hyslop, P. S., Pericak-Vance, M. A., Joo, S. H., ... & Hulette, C. (1993). Association of apolipoprotein E allele ϵ4 with late-onset familial and sporadic Alzheimer's disease. *Neurology*, *43*(8), 1467-1467.

Scarmeas, N., & Stern, Y. (2003). Cognitive reserve and lifestyle. *Journal of clinical and experimental neuropsychology*, *25*(5), 625-633.

Scarmeas, N., Luchsinger, J. A., Schupf, N., Brickman, A. M., Cosentino, S., Tang, M. X., & Stern, Y. (2009). Physical activity, diet, and risk of Alzheimer disease. *Jama*, *302*(6), 627-637.

Scheuner D, Eckman C, Jensen M, Song X, Citron M, Suzuki N, Bird TD, Hardy J, Hutton M, Kukull W *et al* (1996) Secreted amyloid b-protein similar to that in the senile plaques of Alzheimer's disease is increased in vivo by the presenilin 1 and 2 and APP mutations linked to familial Alzheimer's disease. Nature Med 2: 864–870

Scheff, S. W., Price, D. A., Schmitt, F. A., DeKosky, S. T., & Mufson, E. J. (2007). Synaptic alterations in CA1 in mild Alzheimer disease and mild cognitive impairment. *Neurology*, *68*(18), 1501-1508.

Schlepckow, K., Kleinberger, G., Fukumori, A., Feederle, R., Lichtenthaler, S. F., Steiner, H., & Haass, C. (2017). An Alzheimer-associated TREM2 variant occurs at the ADAM cleavage site and affects shedding and phagocytic function. *EMBO molecular medicine*, *9*(10), 1356-1365.

Schmid, N. S., Taylor, K. I., Foldi, N. S., Berres, M., & Monsch, A. U. (2013). Neuropsychological signs of Alzheimer's disease 8 years prior to diagnosis. *Journal of Alzheimer's Disease*, *34*(2), 537-546.

Schmitt, F. A., Ashford, W., Ernesto, C., Saxton, J., Schneider, L. S., Clark, C. M., ... & Thal, L. J. (1997). The severe impairment battery: concurrent validity and the assessment of longitudinal change in Alzheimer's disease. The Alzheimer's disease cooperative study. *Alzheimer disease and associated disorders*, *11*, S51-6.

Schmidt, M., Sachse, C., Richter, W., Xu, C., Fändrich, M., & Grigorieff, N. (2009). Comparison of Alzheimer Aβ (1–40) and Aβ (1–42) amyloid fibrils reveals similar protofilament structures. *Proceedings of the National Academy of Sciences*, *106*(47), 19813-19818.

Schmidt, M. E., Chiao, P., Klein, G., Matthews, D., Thurfjell, L., Cole, P. E., ... & De Santi, S. (2015). The influence of biological and technical factors on quantitative analysis of amyloid PET: points to consider and recommendations for controlling variability in longitudinal data. *Alzheimer's & Dementia*, *11*(9), 1050-1068.

Schnabel, R. (2012). Overcoming the challenge of re-assessing logical memory. *The Clinical Neuropsychologist*, *26*(1), 102-115.

Schneider, L. S., & Farlow, M. R. (1995). Predicting response to cholinesterase inhibitors in Alzheimer’s disease. *CNS drugs*, *4*(2), 114-124.

Schöll, M., Lockhart, S. N., Schonhaut, D. R., O’Neil, J. P., Janabi, M., Ossenkoppele, R., ... & Rabinovici, G. D. (2016). PET imaging of tau deposition in the aging human brain. *Neuron*, *89*(5), 971-982.

Schwarz, A. J., Shcherbinin, S., Slieker, L. J., Risacher, S. L., Charil, A., Irizarry, M. C., ... & Miller, B. B. (2018). Topographic staging of tau positron emission tomography images. *Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring*, *10*, 221-231.

Seab, J. P., Jagust, W. J., Wong, S. T. S., Roos, M. S., Reed, B. R., & Budinger, T. F. (1988). Quantitative NMR measurements of hippocampal atrophy in Alzheimer's disease. *Magnetic Resonance in Medicine*, *8*(2), 200-208.

Seidenberg, M., Kay, C. D., Woodard, J. L., Nielson, K. A., Smith, J. C., Kandah, C., ... & Matthews, M. (2013). Recognition of famous names predicts cognitive decline in healthy elders. *Neuropsychology*, *27*(3), 333.

Selkoe DJ (1991) The molecular pathology of Alzheimer's disease. Neuron 6: 487–498

Selkoe, D. J., & Hardy, J. (2016). The amyloid hypothesis of Alzheimer's disease at 25 years. *EMBO molecular medicine*, *8*(6), 595-608.

Seo, E. H., Kim, S. H., Park, S. H., Kang, S. H., Choo, I. L., & Alzheimer's Disease Neuroimaging Initiative. (2016). Independent and interactive influences of the APOE genotype and beta-amyloid burden on cognitive function in mild cognitive impairment. *Journal of Korean medical science*, *31*(2), 286-295.

Sereno, L., Coma, M., Rodriguez, M., Sanchez-Ferrer, P., Sanchez, M. B., Gich, I., ... & Clarimon, J. (2009). A novel GSK-3β inhibitor reduces Alzheimer's pathology and rescues neuronal loss in vivo. *Neurobiology of disease*, *35*(3), 359-367.

Sevigny, J., Chiao, P., Bussière, T., Weinreb, P. H., Williams, L., Maier, M., ... & O'Gorman, J. (2016). The antibody aducanumab reduces Aβ plaques in Alzheimer's disease. *Nature*, *537*(7618), 50.

Siemers, E. R., Sundell, K. L., Carlson, C., Case, M., Sethuraman, G., Liu-Seifert, H., ... & Demattos, R. (2016). Phase 3 solanezumab trials: secondary outcomes in mild Alzheimer's disease patients. *Alzheimer's & Dementia*, *12*(2), 110-120.

Sigurdsson, E. M. (2018). Tau immunotherapies for Alzheimer's disease and related tauopathies: progress and potential pitfalls. *Journal of Alzheimer's Disease*, (Preprint), 1-11.

Sims, J. R., Selzler, K. J., Downing, A. M., Willis, B. A., Aluise, C. D., Zimmer, J., ... & Eads, J. (2017). Development Review of the BACE1 Inhibitor Lanabecestat (AZD3293/LY3314814). *The journal of prevention of Alzheimer's disease*, *4*(4), 247-254.

Sjögren, M., Blomberg, M., Jonsson, M., Wahlund, L. O., Edman, Å., Lind, K., ... & Wallin, A. (2001). Neurofilament protein in cerebrospinal fluid: a marker of white matter changes. *Journal of neuroscience research*, *66*(3), 510-516.

Shankar, G. M., Li, S., Mehta, T. H., Garcia-Munoz, A., Shepardson, N. E., Smith, I., ... & Regan, C. M. (2008). Amyloid-β protein dimers isolated directly from Alzheimer's brains impair synaptic plasticity and memory. *Nature medicine*, *14*(8), 837.

Shi, F., Liu, B., Zhou, Y., Yu, C., & Jiang, T. (2009). Hippocampal volume and asymmetry in mild cognitive impairment and Alzheimer's disease: Meta-analyses of MRI studies. *Hippocampus*, *19*(11), 1055-1064.

Shim, Y. S., Yang, D. W., Kim, H. J., Park, Y. H., & Kim, S. (2017). Characteristic differences in the mini-mental state examination used in Asian countries. *BMC neurology*, *17*(1), 141.

Shiroky, J. S., Schipper, H. M., Bergman, H., & Chertkow, H. (2007). Can you have dementia with an MMSE score of 30?. *American Journal of Alzheimer's Disease & Other Dementias*, *22*(5), 406-415.

Sikorski, P., Atkins, E. D., & Serpell, L. C. (2003). Structure and texture of fibrous crystals formed by Alzheimer's Aβ (11–25) peptide fragment. *Structure*, *11*(8), 915-926.

Skrobot, O. A., Attems, J., Esiri, M., Hortobágyi, T., Ironside, J. W., Kalaria, R. N., ... & Love, S. (2016). Vascular cognitive impairment neuropathology guidelines (VCING): the contribution of cerebrovascular pathology to cognitive impairment. *Brain*, *139*(11), 2957-2969.

Slezak M, Pfrieger FW. New roles for astrocytes: regulation of CNS synaptogenesis. Trends Neurosci. 2003;26:531–535.

Small, S. A., & Duff, K. (2008). Linking Aβ and tau in late-onset Alzheimer's disease: a dual pathway hypothesis. *Neuron*, *60*(4), 534-542.

Small, B. J., Herlitz, A., Fratiglioni, L., Almkvist, O., & Bäckman, L. (1997). Cognitive predictors of incident Alzheimer's disease: a prospective longitudinal study. *Neuropsychology*, *11*(3), 413.

Smith, E. E., & Beaudin, A. E. (2018). New insights into cerebral small vessel disease and vascular cognitive impairment from MRI. *Current opinion in neurology*, *31*(1), 36-43.

Smith, E. E., Schneider, J. A., Wardlaw, J. M., & Greenberg, S. M. (2012). Cerebral microinfarcts: the invisible lesions. *The Lancet Neurology*, *11*(3), 272-282.

Smith, R., Schöll, M., Londos, E., Ohlsson, T., & Hansson, O. (2018). 18 F-AV-1451 in Parkinson's Disease with and without dementia and in Dementia with Lewy Bodies. *Scientific reports*, *8*(1), 4717.

Snyder, S. W., Ladror, U. S., Wade, W. S., Wang, G. T., Barrett, L. W., Matayoshi, E. D., ... & Holzman, T. F. (1994). Amyloid-beta aggregation: selective inhibition of aggregation in mixtures of amyloid with different chain lengths. *Biophysical Journal*, *67*(3), 1216-1228.

Soldan, A., Pettigrew, C., Cai, Q., Wang, J., Wang, M. C., Moghekar, A., ... & BIOCARD Research Team. (2017). Cognitive reserve and long-term change in cognition in aging and preclinical Alzheimer's disease. *Neurobiology of aging*, *60*, 164-172.

Song, Z., Insel, P. S., Buckley, S., Yohannes, S., Mezher, A., Simonson, A., ... & Miller, B. L. (2015). Brain amyloid-β burden is associated with disruption of intrinsic functional connectivity within the medial temporal lobe in cognitively normal elderly. *Journal of Neuroscience*, *35*(7), 3240-3247.

Song, W. M., Joshita, S., Zhou, Y., Ulland, T. K., Gilfillan, S., & Colonna, M. (2018). Humanized TREM2 mice reveal microglia-intrinsic and-extrinsic effects of R47H polymorphism. *Journal of Experimental Medicine*, *215*(3), 745-760.

Soobiah, C., Tadrous, M., Knowles, S., Blondal, E., Ashoor, H. M., Ghassemi, M., ... & Straus, S. E. (2019). Variability in the validity and reliability of outcome measures identified in a systematic review to assess treatment efficacy of cognitive enhancers for Alzheimer's Dementia. *PloS one*, *14*(4), e0215225.

Soontornniyomkij, V., Lynch, M. D., Mermash, S., Pomakian, J., Badkoobehi, H., Clare, R., & Vinters, H. V. (2010). Cerebral microinfarcts associated with severe cerebral β-amyloid angiopathy. *Brain pathology*, *20*(2), 459-467.

Spampinato, M. V., Langdon, B. R., Patrick, K. E., Parker, R. O., Collins, H., Pravata, E., & Alzheimer's Disease Neuroimaging Initiative. (2016). Gender, apolipoprotein E genotype, and mesial temporal atrophy: 2-year follow-up in patients with stable mild cognitive impairment and with progression from mild cognitive impairment to Alzheimer's disease. *Neuroradiology*, *58*(11), 1143-1151.

Spencer, R. J., Wendell, C. R., Giggey, P. P., Katzel, L. I., Lefkowitz, D. M., Siegel, E. L., & Waldstein, S. R. (2013). Psychometric limitations of the mini-mental state examination among nondemented older adults: an evaluation of neurocognitive and magnetic resonance imaging correlates. *Experimental aging research*, *39*(4), 382-397.

Sperling, R. A., Jack Jr, C. R., & Aisen, P. S. (2011). Testing the right target and right drug at the right stage. *Science translational medicine*, *3*(111), 111-113.

Sperling, R. A., Jack Jr, C. R., Black, S. E., Frosch, M. P., Greenberg, S. M., Hyman, B. T., ... & Black, R. S. (2011). Amyloid-related imaging abnormalities in amyloid-modifying therapeutic trials: recommendations from the Alzheimer's Association Research Roundtable Workgroup. *Alzheimer's & Dementia*, *7*(4), 367-385.

Sperling, R., Salloway, S., Brooks, D. J., Tampieri, D., Barakos, J., Fox, N. C., ... & Lieberburg, I. (2012). Amyloid-related imaging abnormalities in patients with Alzheimer's disease treated with bapineuzumab: a retrospective analysis. *The Lancet Neurology*, *11*(3), 241-249.

Sperling, R. A., Karlawish, J., & Johnson, K. A. (2013a). Preclinical Alzheimer disease—the challenges ahead. *Nature Reviews Neurology*, *9*(1), 54.

Sperling, R. A., Johnson, K. A., Doraiswamy, P. M., Reiman, E. M., Fleisher, A. S., Sabbagh, M. N., ... & Flitter, M. (2013b). Amyloid deposition detected with florbetapir F 18 (18F-AV-45) is related to lower episodic memory performance in clinically normal older individuals. *Neurobiology of aging*, *34*(3), 822-831.

Sperling, R., Mormino, E., & Johnson, K. (2014). The evolution of preclinical Alzheimer's disease: implications for prevention trials. *Neuron*, *84*(3), 608-622.

Spreng, R. N., & Turner, G. R. (2019). The shifting architecture of cognition and brain function in older adulthood.

Stefani, A., Martorana, A., Bernardini, S., Panella, M., Mercati, F., Orlacchio, A., & Pierantozzi, M. (2006). CSF markers in Alzheimer disease patients are not related to the different degree of cognitive impairment. *Journal of the neurological sciences*, *251*(1-2), 124-128.

Stephens, A., Kroth, H., Berndt, M., Capotosti, F., & Mueller, A. (2017). Characterization of novel PET tracers for the assessment of tau pathology in Alzheimer's disease and other tauopathies. *Neurodegener. Dis*, *17*.

Stern, Y., Gurland, B., Tatemichi, T. K., Tang, M. X., Wilder, D., & Mayeux, R. (1994). Influence of education and occupation on the incidence of Alzheimer's disease. *Jama*, *271*(13), 1004-1010.

Stern, Y. (2006). Cognitive reserve and Alzheimer disease. *Alzheimer Disease & Associated Disorders*, *20*, S69-S74.

Stern, Y. (2009). Cognitive reserve. *Neuropsychologia*, *47*(10), 2015-2028.

Stern, Y. (2012). Cognitive reserve in ageing and Alzheimer's disease. *The Lancet Neurology*, *11*(11), 1006-1012.

Stonnington, C. M., Chen, K., Lee, W., Locke, D. E., Dueck, A. C., Liu, X., ... & Reiman, E. M. (2014). Fibrillar amyloid correlates of preclinical cognitive decline. *Alzheimer's & Dementia*, *10*(1), e1-e8.

Stricker, N. H., Lundt, E. S., Alden, E. C., Albertson, S. M., Machulda, M. M., Kremers, W. K., ... & Mielke, M. M. (2020). Longitudinal comparison of in clinic and at home administration of the cogstate brief battery and demonstrated practice effects in the Mayo Clinic Study of Aging. *The journal of prevention of Alzheimer's disease*, *7*(1), 21-28.

Strite D, Massman PJ, Cooke N, & Doody RS (1997). Neuropsychological asymmetry in Alzheimer’s disease: Verbal versus visuoconstructional deficits across stages of dementia. Journal of the International Neuropsychological Society, 3(5), 420–427.

Suárez-Calvet M, Kleinberger G, Araque Caballero MÁ, Brendel M, Rominger A, Alcolea D, Fortea J, Lleó A, Blesa R, Gispert JD *et al* (2016a) sTREM2 cerebrospinal fluid levels are a potential biomarker for microglia activity in early-stage Alzheimer's disease and associate with neuronal injury markers. EMBO Mol Med 2016

Suárez-Calvet, M., Caballero, M. Á. A., Kleinberger, G., Bateman, R. J., Fagan, A. M., Morris, J. C., ... & Dominantly Inherited Alzheimer Network. (2016b). Early changes in CSF sTREM2 in dominantly inherited Alzheimer’s disease occur after amyloid deposition and neuronal injury. *Science translational medicine*, *8*(369)

Sucher, N. J., Awobuluyi, M., Choi, Y. B., & Lipton, S. A. (1996). NMDA receptors: from genes to channels. *Trends in pharmacological sciences*, *17*(10), 348-355.

Sullivan, K. (2005). Alternate forms of prose passages for the assessment of auditory–verbal memory. *Archives of Clinical Neuropsychology*, *20*(6), 745-753.

Sunderland, T., Wolozin, B., Galasko, D., Levy, J., Dukoff, R., Bahro, M., ... & Seubert, P. (1999). Longitudinal stability of CSF tau levels in Alzheimer patients. *Biological psychiatry*, *46*(6), 750-755.

Suzuki, N., Cheung, T. T., Cai, X. D., Odaka, A., Otvos, L., Eckman, C., ... & Younkin, S. G. (1994). An increased percentage of long amyloid beta protein secreted by familial amyloid beta protein precursor (beta APP717) mutants. *Science*, *264*(5163), 1336-1340.

Swanson, C. J., Zhang, Y., Dhadda, S., Wang, J., Kaplow, J., Lai, R. Y., ... & Cummings, J. L. (2021). A randomized, double-blind, phase 2b proof-of-concept clinical trial in early Alzheimer's disease with lecanemab, an anti-Aβ protofibril antibody. *Alzheimer's research & therapy*, *13*(1), 1-14.

Szaruga, M., Munteanu, B., Lismont, S., Veugelen, S., Horré, K., Mercken, M., ... & Gallardo, R. (2017). Alzheimer's-causing mutations shift Aβ length by destabilizing γ-secretase-Aβn interactions. *Cell*, *170*(3), 443-456.

Sze, C. I., Troncoso, J. C., Kawas, C., Mouton, P., Price, D. L., & Martin, L. J. (1997). Loss of the presynaptic vesicle protein synaptophysin in hippocampus correlates with cognitive decline in Alzheimer disease. *Journal of Neuropathology & Experimental Neurology*, *56*(8), 933-944.

T

Takami, M., Nagashima, Y., Sano, Y., Ishihara, S., Morishima-Kawashima, M., Funamoto, S., & Ihara, Y. (2009). γ-Secretase: successive tripeptide and tetrapeptide release from the transmembrane domain of β-carboxyl terminal fragment. *Journal of Neuroscience*, *29*(41), 13042-13052.

Takahashi, K., Rochford, C. D., & Neumann, H. (2005). Clearance of apoptotic neurons without inflammation by microglial triggering receptor expressed on myeloid cells-2. *Journal of Experimental Medicine*, *201*(4), 647-657.

Takahashi, K., Prinz, M., Stagi, M., Chechneva, O., & Neumann, H. (2007). TREM2-transduced myeloid precursors mediate nervous tissue debris clearance and facilitate recovery in an animal model of multiple sclerosis. *PLoS medicine*, *4*(4), e124.

Takaiwa, A., Kuwayama, N., Akioka, N., Kashiwazaki, D., & Kuroda, S. (2018). Discrepancy analysis between crystallized and fluid intelligence tests: a novel method to detect mild cognitive impairment in patients with asymptomatic carotid artery stenosis. *European Journal of Neurology*, *25*(2), 313-319.

Taler, V., Davidson, P. S., Sheppard, C., & Gardiner, J. (2021). A discourse-theoretic approach to story recall in aging and mild cognitive impairment. *Aging, Neuropsychology, and Cognition*, *28*(5), 762-780.

Tanveer, M., Richhariya, B., Khan, R. U., Rashid, A. H., Khanna, P., Prasad, M., & Lin, C. T. (2020). Machine learning techniques for the diagnosis of Alzheimer's disease: A review. *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)*, *16*(1s), 1-35.

Tanzi, R. E. (2015). TREM2 and risk of Alzheimer's disease—Friend or foe?. *New England Journal of Medicine*, *372*(26), 2564-2565.

Tarawneh, R., D'angelo, G., Crimmins, D., Herries, E., Griest, T., Fagan, A. M., ... & Holtzman, D. M. (2016). Diagnostic and prognostic utility of the synaptic marker neurogranin in Alzheimer disease. *JAMA neurology*, *73*(5), 561-571.

Taylor, E. M. (1959). Psychological appraisal of children with cerebral defects. Cambridge, MA: Harvard University Press, pp. 423–428

Ten Kate, M., Redolfi, A., Peira, E., Bos, I., Vos, S. J., Vandenberghe, R., ... & Barkhof, F. (2018). MRI predictors of amyloid pathology: results from the EMIF-AD Multimodal Biomarker Discovery study. *Alzheimer's research & therapy*, *10*(1), 1-12.

Thal, D. R., Rüb, U., Orantes, M., & Braak, H. (2002). Phases of Aβ-deposition in the human brain and its relevance for the development of AD. *Neurology*, *58*(12), 1791-1800.

Thall, P.F. (2021) "Adaptive Enrichment Designs in Clinical Trials," *https://doi.org/10.1146/annurev-statistics-040720-032818*, 8, pp. 393–411.

Thomas, S. J., & Grossberg, G. T. (2009). Memantine: a review of studies into its safety and efficacy in treating Alzheimer's disease and other dementias. *Clinical interventions in aging*, *4*, 367.

Thompson, T. A., Wilson, P. H., Snyder, P. J., Pietrzak, R. H., Darby, D., Maruff, P., & Buschke, H. (2011). Sensitivity and test–retest reliability of the international shopping list test in assessing verbal learning and memory in mild Alzheimer's disease. Archives of clinical neuropsychology, 26(5), 412-424.

Thornton, P., Sevalle, J., Deery, M. J., Fraser, G., Zhou, Y., Ståhl, S., ... & Nicol, L. S. (2017). TREM2 shedding by cleavage at the H157-S158 bond is accelerated for the Alzheimer's disease-associated H157Y variant. *EMBO molecular medicine*, *9*(10), 1366-1378.

Tijms, B. M., Gobom, J., Reus, L., Jansen, I., Hong, S., Dobricic, V., ... & Visser, P. J. (2020). Pathophysiological subtypes of Alzheimer's disease based on cerebrospinal fluid proteomics. *Brain*, *143*(12), 3776-3792.

Tilvis, R. S., Kähönen-Väre, M. H., Jolkkonen, J., Valvanne, J., Pitkala, K. H., & Strandberg, T. E. (2004). Predictors of cognitive decline and mortality of aged people over a 10-year period. *The Journals of Gerontology Series A: Biological Sciences and Medical Sciences*, *59*(3), M268-M274.

Toledo, J. B., Bjerke, M., Chen, K., Rozycki, M., Jack, C. R., Weiner, M. W., ... & Trojanowski, J. Q. (2015). Memory, executive, and multidomain subtle cognitive impairment: clinical and biomarker findings. *Neurology*, *85*(2), 144-153.

Tombaugh, T. N., & McIntyre, N. J. (1992). The mini-mental state examination: a comprehensive review. *Journal of the American Geriatrics Society*, *40*(9), 922-935.

Trenkle, D. L., Shankle, W. R., & Azen, S. P. (2007). Detecting cognitive impairment in primary care: performance assessment of three screening instruments. *Journal of Alzheimer's Disease*, *11*(3), 323-335.

Trifilio, E., Tanner, J. J., Butterfield, L., Mangal, P., Maye, J. E., Marsiske, M., ... & Bowers, D. (2021). A tale of two stories: Validity of an alternative story memory test in a sample of older adults. *The Clinical Neuropsychologist*, *34*(1), 158-173.

Tryon, W. W. (2001). Evaluating statistical difference, equivalence, and indeterminacy using inferential confidence intervals: an integrated alternative method of conducting null hypothesis statistical tests. *Psychological methods*, *6*(4), 371.

Turnbull, I. R., Gilfillan, S., Cella, M., Aoshi, T., Miller, M., Piccio, L., ... & Colonna, M. (2006). Cutting edge: TREM-2 attenuates macrophage activation. *The Journal of Immunology*, *177*(6), 3520-3524.

U

Ueckert, S., Plan, E. L., Ito, K., Karlsson, M. O., Corrigan, B., Hooker, A. C., & Alzheimer’s Disease Neuroimaging Initiative. (2014). Improved utilization of ADAS-cog assessment data through item response theory based pharmacometric modeling. *Pharmaceutical research*, *31*(8), 2152-2165.

Ulrich, J. D., Finn, M. B., Wang, Y., Shen, A., Mahan, T. E., Jiang, H., ... & Holtzman, D. M. (2014). Altered microglial response to Aβ plaques in APPPS1-21 mice heterozygous for TREM2. *Molecular neurodegeneration*, *9*(1), 20.

V

van der Flier, W. M., Pijnenburg, Y. A., Fox, N. C., & Scheltens, P. (2011). Early-onset versus late-onset Alzheimer's disease: the case of the missing APOE ε4 allele. *The Lancet Neurology*, *10*(3), 280-288.

van Dyck, C. H. (2018). Anti-amyloid-β monoclonal antibodies for Alzheimer's disease: pitfalls and promise. *Biological psychiatry*, *83*(4), 311-319.

Van Exel, E., Gussekloo, J., De Craen, A. J. M., Bootsma-Van Der Wiel, A., Houx, P., Knook, D. L., & Westendorp, R. G. J. (2001). Cognitive function in the oldest old: women perform better than men. *Journal of Neurology, Neurosurgery & Psychiatry*, *71*(1), 29-32.

Van Hooren, S. A. H., Valentijn, A. M., Bosma, H., Ponds, R. W. H. M., Van Boxtel, M. P. J., & Jolles, J. (2007). Cognitive functioning in healthy older adults aged 64–81: a cohort study into the effects of age, sex, and education. *Aging, Neuropsychology, and Cognition*, *14*(1), 40-54.

Van Veluw, S. J., Hilal, S., Kuijf, H. J., Ikram, M. K., Xin, X., Yeow, T. B., ... & Chen, C. (2015). a Cortical microinfarcts on 3T MRI: clinical correlates in memory-clinic patients. *Alzheimer's & dementia*, *11*(12), 1500-1509.
van Veluw, S. J., Biessels, G. J., Luijten, P. R., & Zwanenburg, J. J. (2015).b. Assessing cortical cerebral microinfarcts on high resolution MR images. *JoVE (Journal of Visualized Experiments)*, (105), e53125.

Varvel, N. H., Grathwohl, S. A., Degenhardt, K., Resch, C., Bosch, A., Jucker, M., & Neher, J. J. (2015). Replacement of brain-resident myeloid cells does not alter cerebral amyloid-β deposition in mouse models of Alzheimer's disease. *Journal of Experimental Medicine*, *212*(11), 1803-1809.

Vassar, R. (2019). Implications for BACE1 Inhibitor Clinical Trials: Adult Conditional BACE1 Knockout Mice Exhibit Axonal Organization Defects in the Hippocampus.

Vassar, R., Bennett, B. D., Babu-Khan, S., Kahn, S., Mendiaz, E. A., Denis, P., ... & Luo, Y. (1999). β-Secretase cleavage of Alzheimer's amyloid precursor protein by the transmembrane aspartic protease BACE. *science*, *286*(5440), 735-741.

Veitch, D. P., Weiner, M. W., Aisen, P. S., Beckett, L. A., Cairns, N. J., Green, R. C., ... & Petersen, R. C. (2018). Understanding disease progression and improving Alzheimer's disease clinical trials: Recent highlights from the Alzheimer's Disease Neuroimaging Initiative. *Alzheimer's & Dementia*.

Vellas, B., Andrieu, S., Sampaio, C., Coley, N., & Wilcock, G. (2008). Endpoints for trials in Alzheimer's disease: a European task force consensus. *The Lancet Neurology*, *7*(5), 436-450.

Vermeiren, C., Motte, P., Viot, D., Mairet-Coello, G., Courade, J. P., Citron, M., ... & Gillard, M. (2018). The tau positron-emission tomography tracer AV-1451 binds with similar affinities to tau fibrils and monoamine oxidases. *Movement Disorders*, *33*(2), 273-281.

Veugelen, S., Dewilde, M., De Strooper, B., & Chávez-Gutiérrez, L. (2017). Screening and characterization strategies for nanobodies targeting membrane proteins. In *Methods in enzymology* (Vol. 584, pp. 59-97). Academic Press.

Villain, N., Chételat, G., Grassiot, B., Bourgeat, P., Jones, G., Ellis, K. A., ... & Masters, C. L. (2012). Regional dynamics of amyloid-β deposition in healthy elderly, mild cognitive impairment and Alzheimer's disease: a voxelwise PiB–PET longitudinal study. *Brain*, *135*(7), 2126-2139.

Villeneuve, S., Rabinovici, G. D., Cohn-Sheehy, B. I., Madison, C., Ayakta, N., Ghosh, P. M., ... & Lehmann, M. (2015). Existing Pittsburgh Compound-B positron emission tomography thresholds are too high: statistical and pathological evaluation. *Brain*, *138*(7), 2020-2033.

Villemagne, V. L., Burnham, S., Bourgeat, P., Brown, B., Ellis, K. A., Salvado, O., ... & Ames, D. (2013). Amyloid β deposition, neurodegeneration, and cognitive decline in sporadic Alzheimer's disease: a prospective cohort study. *The Lancet Neurology*, *12*(4), 357-367.

Villemagne, V. L., Doré, V., Burnham, S. C., Masters, C. L., & Rowe, C. C. (2018). Imaging tau and amyloid-β proteinopathies in Alzheimer disease and other conditions. *Nature Reviews Neurology*, *14*(4), 225.

Visser, P. J., & Knopman, D. S. (2009). Amyloid imaging in the prediction of Alzheimer-type dementia in subjects with amnestic MCI. *Neurology*, *73*(10), 744-745.

Vogel A, Stokholm J, Gade A,et al. 2004. Awareness of deficits inmild cognitive impairment and Alzheimer's disease: do MCIpatients have impaired insight?Dement Geriatr Cogn Disord17: 181–187.

Vogel, J. W., Young, A. L., Oxtoby, N. P., Smith, R., Ossenkoppele, R., Strandberg, O. T., ... & Hansson, O. (2021). Four distinct trajectories of tau deposition identified in Alzheimer's disease. *Nature medicine*, *27*(5), 871-881.

Voss, T., Zhang, Y., Sur, C., Kost, J., Mo, Y., Suhy, J., ... & Mozley, L. H. (2016). DOES REGIONAL AMYLOID BURDEN IMPACT DOMAIN-SPECIFIC COGNITIVE

IMPAIRMENT?. *Alzheimer's & Dementia: The Journal of the Alzheimer's Association*, *12*(7), P1030.

Voytyuk, I., Mueller, S. A., Herber, J., Snellinx, A., Moechars, D., van Loo, G., ... & De Strooper, B. (2018). BACE2 distribution in major brain cell types and identification of novel substrates. *Life science alliance*, *1*(1), e201800026.

W

Wallin, Å. K., Blennow, K., Andreasen, N., & Minthon, L. (2006). CSF biomarkers for Alzheimer’s Disease: levels of β-amyloid, tau, phosphorylated tau relate to clinical symptoms and survival. *Dementia and geriatric cognitive disorders*, *21*(3), 131-138.

Walji, A. M., Hostetler, E. D., Selnick, H., Zeng, Z., Miller, P., Bennacef, I., ... & O’Malley, S. (2016). Discovery of 6-(Fluoro-18 F)-3-(1 H-pyrrolo [2, 3-c] pyridin-1-yl) isoquinolin-5-amine ([18F]-MK-6240): A Positron Emission Tomography (PET) Imaging Agent for Quantification of Neurofibrillary Tangles (NFTs). *Journal of medicinal chemistry*, *59*(10), 4778-4789.

Wang, J. Z., Gong, C. X., Zaidi, T., Grundke-Iqbal, I., & Iqbal, K. (1995). Dephosphorylation of Alzheimer paired helical filaments by protein phosphatase-2A and− 2B. *Journal of Biological Chemistry*, *270*(9), 4854-4860.

Wang, J. Z., Grundke-Iqbal, I., & Iqbal, K. (1996). Glycosylation of microtubule–associated protein tau: An abnormal posttranslational modification in Alzheimer's disease. *Nature medicine*, *2*(8), 871.

Wang, D. S., Dickson, D. W., & Malter, J. S. (2006). β-Amyloid degradation and Alzheimer's disease. *BioMed Research International*, *2006*.

Wang, X., Kota, U., He, K., Blackburn, K., Li, J., Goshe, M. B., ... & Clouse, S. D. (2008). Sequential transphosphorylation of the BRI1/BAK1 receptor kinase complex impacts early events in brassinosteroid signaling. *Developmental cell*, *15*(2), 220-235.

Wang Y, Cella M, Mallinson KJ, Ulrich JD, Young KL, Robinette ML, Gilfillan S, Krishnan GM, Sudhakar S, Zinselmeyer BH *et al* (2015) TREM2 lipid sensing sustains the microglial response in an Alzheimer's disease model. Cell 160: 1061–1071

Wang, J., Logovinsky, V., Hendrix, S. B., Stanworth, S. H., Perdomo, C., Xu, L., ... & Cummings, J. (2016). ADCOMS: a composite clinical outcome for prodromal Alzheimer's disease trials. *J Neurol Neurosurg Psychiatry*, *87*(9), 993-999.

Wechsler, D. (1944). The measurement of adult intelligence.

Wechsler, D. (1974). Wechsler intelligence scale for children—revised. New York: Psychological Corporation.

Wechsler, David, 1896-1981. (1981). WAIS-R : Wechsler adult intelligence scale-revised. New York, N.Y. :Psychological Corporation

Weschler, D. (1997). Weschler adult intelligence scale.

Wechsler, D. (2008). *Wechsler Adult Intelligence Scale* (4th ed.). San Antonio, TX: Pearson Assessment.

Wechsler, D. (2009). *Wechsler memory scale*(4th ed.). (WMS-IV). San Antonio, TX: Pearson Assessment.

Wechsler, D., & Stone, C. P. (1973). *Wechsler memory scale manual*. Psychological Corporation.

Weiner, M. W., Veitch, D. P., Aisen, P. S., Beckett, L. A., Cairns, N. J., Green, R. C., ... & Petersen, R. C. (2017). Recent publications from the Alzheimer's Disease Neuroimaging Initiative: Reviewing progress toward improved AD clinical trials. *Alzheimer's & Dementia*, *13*(4), e1-e85.

Weingarten, M. D., Lockwood, A. H., Hwo, S. Y., & Kirschner, M. W. (1975). A protein factor essential for microtubule assembly. *Proceedings of the National Academy of Sciences*, *72*(5), 1858-1862.

Weintraub, S., Carrillo, M. C., Farias, S. T., Goldberg, T. E., Hendrix, J. A., Jaeger, J., ... & Randolph, C. (2018). Measuring cognition and function in the preclinical stage of Alzheimer's disease. *Alzheimer's & Dementia: Translational Research & Clinical Interventions*, *4*, 64-75.

Wessels, A. M., Edgar, C. J., Nathan, P. J., Siemers, E. R., Maruff, P., & Harrison, J. (2021). Cognitive Go/No-Go decision-making criteria in Alzheimer’s disease drug development. *Drug Discovery Today*, *26*(5), 1330-1336.

Wiederholt, W. C., Cahn, D., Butters, N. M., Salmon, D. P., Kritz-Silverstein, D., & Barrett-Connor, E. (1993). Effects of age, gender and education on selected neuropsychological tests in an elderly community cohort. *Journal of the American Geriatrics Society*, *41*(6), 639-647.

Wijsman, E. M., Pankratz, N. D., Choi, Y., Rothstein, J. H., Faber, K. M., Cheng, R., ... & Goate, A. M. (2011). Genome-wide association of familial late-onset

Alzheimer's disease replicates BIN1 and CLU and nominates CUGBP2 in interaction with APOE. *PLoS genetics*, *7*(2), e1001308.

Wilcock, G. K., Gauthier, S., Frisoni, G. B., Jia, J., Hardlund, J. H., Moebius, H. J., ... & Davis, C. S. (2018). Potential of low dose leuco-methylthioninium bis (hydromethanesulphonate)(LMTM) monotherapy for treatment of mild Alzheimer's disease: Cohort analysis as modified primary outcome in a phase III clinical trial. *Journal of Alzheimer's Disease*, *61*(1), 435-457.

Williams, M. M., Roe, C. M., & Morris, J. C. (2009). Stability of the clinical dementia rating, 1979-2007. *Archives of neurology*, *66*(6), 773-777.

Williams, M. M., Storandt, M., Roe, C. M., & Morris, J. C. (2013). Progression of Alzheimer's disease as measured by Clinical Dementia Rating Sum of Boxes scores. *Alzheimer's & Dementia*, *9*(1), S39-S44.

Wilson, R. S., Bennett, D. A., Bienias, J. L., De Leon, C. M., Morris, M. C., & Evans, D. A. (2003a). Cognitive activity and cognitive decline in a biracial community population. *Neurology*, *61*(6), 812-816.

Wilson, R. S., & Bennett, D. A. (2003b). Cognitive activity and risk of Alzheimer's disease. *Current Directions in Psychological Science*, *12*(3), 87-91.

Wilson, R. S., Krueger, K. R., Arnold, S. E., Schneider, J. A., Kelly, J. F., Barnes, L. L., ... & Bennett, D. A. (2007). Loneliness and risk of Alzheimer disease. *Archives of general psychiatry*, *64*(2), 234-240.

Wilson, R. S., Capuano, A. W., Yu, L., Yang, J., Kim, N., Leurgans, S. E., ... & Boyle, P. A. (2018). Neurodegenerative disease and cognitive retest learning. Neurobiology of aging, 66, 122-130.

Winblad, B., Gauthier, S., Scinto, L., Feldman, H., Wilcock, G. K., Truyen, L., ... & Nye, J. S. (2008). Safety and efficacy of galantamine in subjects with mild cognitive impairment. *Neurology*, *70*(22), 2024-2035.

Wirth, M., Villeneuve, S., Haase, C. M., Madison, C. M., Oh, H., Landau, S. M., ... & Jagust, W. J. (2013). Associations between Alzheimer disease biomarkers, neurodegeneration, and cognition in cognitively normal older people. *JAMA neurology*, *70*(12), 1512-1519.

Wolfe, M. S., Xia, W., Moore, C. L., Leatherwood, D. D., Ostaszewski, B., Rahmati, T., ... & Selkoe, D. J. (1999). Peptidomimetic probes and molecular modeling suggest that Alzheimer's γ-secretase is an intramembrane-cleaving aspartyl protease. *Biochemistry*, *38*(15), 4720-4727.

Wong, D. F., Rosenberg, P. B., Zhou, Y., Kumar, A., Raymont, V., Ravert, H. T., ... & Hilton, J. (2010). In vivo imaging of amyloid deposition in Alzheimer disease using the radioligand 18F-AV-45 (flobetapir F 18). *Journal of Nuclear Medicine*, *51*(6), 913-920.

Wong, D. F., Comley, R., Kuwabara, H., Rosenberg, P. B., Resnick, S. M., Ostrowitzki, S., ... & Honer, M. (2018). First in-human PET study of 3 novel tau radiopharmaceuticals:[11C] RO6924963,[11C] RO6931643, and [18F] RO6958948. *Journal of Nuclear Medicine*, jnumed-118.

Woodward, H. R. (1996). *Reliability of traditional neurological sensory and motor tests*. Ball State University.

Wu, J., Li, J., Huang, K. P., & Huang, F. L. (2002). Attenuation of protein kinase C and cAMP-dependent protein kinase signal transduction in the neurogranin knockout mouse. *Journal of Biological Chemistry*, *277*(22), 19498-19505.

X

Xia, C. F., Arteaga, J., Chen, G., Gangadharmath, U., Gomez, L. F., Kasi, D., ... & Mu, F. (2013). [18F] T807, a novel tau positron emission tomography imaging agent for Alzheimer's disease. *Alzheimer's & Dementia*, *9*(6), 666-676.

Y

Yan, Y., & Wang, C. (2007). Aβ40 protects non-toxic Aβ42 monomer from aggregation. *Journal of molecular biology*, *369*(4), 909-916.

Yanamandra, K., Kfoury, N., Jiang, H., Mahan, T. E., Ma, S., Maloney, S. E., ... & Holtzman, D. M. (2013). Anti-tau antibodies that block tau aggregate seeding in vitro markedly decrease pathology and improve cognition in vivo. *Neuron*, *80*(2), 402-414.

Yoon, C. W., Shin, J. S., Kim, H. J., Cho, H., Noh, Y., Kim, G. H., ... & Na, D. L. (2013). Cognitive deficits of pure subcortical vascular dementia vs Alzheimer disease: PiB-PET–based study. *Neurology*, *80*(6), 569-573.

Yoshiike, Y., Chui, D. H., Akagi, T., Tanaka, N., & Takashima, A. (2003). Specific compositions of amyloid-β peptides as the determinant of toxic β-aggregation. *Journal of Biological Chemistry*, *278*(26), 23648-23655.

Yotter, R. A., Doshi, J., Clark, V., Sojkova, J., Zhou, Y., Wong, D. F., ... & Davatzikos, C. (2013). Memory decline shows stronger associations with estimated spatial patterns of amyloid deposition progression than total amyloid burden. *Neurobiology of aging*, *34*(12), 2835-2842.

Z

Zachary, R. A., & Shipley, W. C. (1986). *Shipley institute of living scale: Revised manual*. Los Angeles, CA: Western Psychological Services.

Zalocusky, K. A., Nelson, M. R., & Huang, Y. (2019). An Alzheimer's-disease-protective APOE mutation. *Nature Medicine*, *25*(11), 1648-1649.

Zanotta, D., Puricelli, S., & Bonoldi, G. (2014). Cognitive effects of a dietary supplement made from extract of Bacopa monnieri, astaxanthin, phosphatidylserine, and vitamin E in subjects with mild cognitive impairment: a noncomparative, exploratory clinical study. *Neuropsychiatric disease and treatment*, *10*, 225.

Zetterberg, H., Pedersen, M., Lind, K., Svensson, M., Rolstad, S., Eckerström, C., ... & Nordlund, A. (2007). Intra-individual stability of CSF biomarkers for Alzheimer's disease over two years. *Journal of Alzheimer's Disease*, *12*(3), 255-260.

Zetterberg, H., Skillbäck, T., Mattsson, N., Trojanowski, J. Q., Portelius, E., Shaw, L. M., ... & Blennow, K. (2016). Association of cerebrospinal fluid neurofilament light concentration with Alzheimer disease progression. *JAMA neurology*, *73*(1), 60-67.

Zhang, Y. D., Wang, S., & Dong, Z. (2014). Classification of Alzheimer disease based on structural magnetic resonance imaging by kernel support vector machine decision tree. *Progress In Electromagnetics Research*, *144*, 171-184.

Zhong, L., Chen, X. F., Wang, T., Wang, Z., Liao, C., Wang, Z., ... & Jia, L. (2017). Soluble TREM2 induces inflammatory responses and enhances microglial survival. *Journal of Experimental Medicine*, *214*(3), 597-607.

Zhu, K., Xiang, X., Filser, S., Marinković, P., Dorostkar, M. M., Crux, S., ... & Lichtenthaler, S. F. (2018). Beta-site amyloid precursor protein cleaving enzyme 1 inhibition impairs synaptic plasticity via seizure protein 6. *Biological psychiatry*, *83*(5), 428-437.

Zokaei, N., Čepukaitytė, G., Board, A. G., Mackay, C. E., Husain, M., & Nobre, A. C. (2019). Dissociable effects of the apolipoprotein-E (APOE) gene on short-and long-term memories. *Neurobiology of Aging*, *73*, 115-122.