Validation of linguistic markers in biologically defined prodromal AD and testing of their validity in differential diagnosis

Ronan O'Malley

A thesis submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy

The University of Sheffield
Faculty of Neuroscience
School of Medicine and Population Health

Submission Date 17th of June 2025

Abstract

The early identification of prodromal Alzheimer's disease (AD) is a pressing clinical challenge, particularly as disease-modifying therapies emerge. This thesis investigates whether semantic and linguistic features derived from speech can contribute towards the diagnosis of mild cognitive impairment (MCI) due to AD. It also evaluates the feasibility of remote, automated cognitive assessment using a virtual clinical agent.

Two primary studies were conducted. The first involved the development and testing of CognoSpeak, a fully automated system that administers verbal fluency and interview-style tasks. CognoSpeak demonstrated high classification accuracy, distinguishing patients with neurodegenerative conditions from healthy controls and those with functional cognitive symptoms.

The second study explored advanced semantic features extracted from verbal fluency and free speech samples, comparing their diagnostic utility to that of traditional cognitive assessments and structural imaging biomarkers, including hippocampal volume and voxel-based morphometry.

Key findings indicate that specific semantic language features correlate significantly with hippocampal volume and contributed towards accurate prediction distinction of prodromal AD from controls. When combined with cognitive scores and imaging data, language features enhanced classification performance. The research supports the hypothesis that semantic markers derived from speech can serve as valid indicators of early cognitive decline and, when embedded in automated systems, offer scalable, accessible tools for screening and monitoring.

These findings underscore the diagnostic value of language features and their potential integration into clinical pathways. The thesis concludes that language-based tools are not only feasible but also clinically meaningful adjuncts in the early detection of Alzheimer's disease.

Acknowledgements

I would like to express my sincere gratitude to my supervisors, Dr Blackburn and Professor Venneri, for their invaluable guidance, support, and patience throughout the course of this research. I would also like to thank my collaborators, colleagues in the CognoSpeak team, and the participants who generously contributed their time and insight.

Special thanks to my family and friends for their unwavering support and encouragement.

Contents

Abstract 1
Acknowledgements 3
List of Figures 6
List of Tables 7
List of Abbreviations 8
1. Introduction and Background 10
1.1 Hypotheses 10
1.2 Introduction 10
1.3 Background 12
1.3.5 Therapies 21
1.3.6 Language and AD 23
1.4 Summary 29
2 General Methodology 31
2.1 Introduction to General Methodology 31
2.2 Ethical Approval 31
2.3 Participants 31
2.4 Recruitment Procedure 31
2.5 Inclusion and Exclusion Criteria 32
2.5 Study Assessment Process 33
3.1 Introduction to Chapter: A fully automated cognitive screening tool based on assessment of speech and language. 36
Chapter 3.2 Published Research Article 38
A fully automated cognitive screening tool based on assessment of speech and language.. 38
3.2.1 Abstract 39
3.2.2 Introduction 40
3.2.3 Methods 40
3.2.4 Results 42
3.2.5 Discussion 45
3.2.6 Conclusion 48
4. Advanced Semantic Analysis 50
4.1. Introduction 50
4.2 Advanced Feature Analysis for MCI Detection: Methods 51
4.2.1 Longitudinal Review of the MCI Cohort 51
4.2.2 Structural MRI Image Acquisition 53

4.2.4 Language Based Features 56
4.2.5 Note on Statistical Analysis 59
4.3 Advanced Semantic Feature Analysis for MCI Detection: Results 60
4.3.1 Introduction to Results 60
4.3.2 Relationship Between Semantic and Imaging Features 61
4.3.3 Comparison of Features between groups 73
4.3.4 Summary of between feature and between group comparisons 87
4.3.5 Exploring the predictive value of individual features. 89
4.3.6 Exploring the predictive value of combined feature sets. 97
4.3.7 Exploring the potential benefits of machine learning models. 104
4.3.8. Results Summary 110
4.4 Advanced Semantic Feature Analysis for MCI Detection: Discussion 113
4.4.1 Introduction to Discussion 113
4.4.2 Methodological Considerations surrounding the composition of the MCI group. 113
4.4.3 Voxel Based Morphometry 115
4.4.4 Hippocampal Volumes 117
4.4.5 Language Based Features 118
4.4.6 Machine Learning Classifiers 120
4.4.7 Considerations for further work 121
5. Conclusion 122
5.1 Summary of Key Findings 122
5.2 Interpretation of Key Findings 122
5.3 Future Work 123
5.4 Final Reflections 124
Appendix A 125
CONSENT FORM 1 128
Patient Information Sheet 1 130
Appendix B 134
Bibliography 142

List of Figures

Figure 1 Confusion Matrix showing Two-Way Classification Accuracy (HC+FMD vs AD+MCI) 43
Figure 2 Confusion Matrix showing Three-Way Classification Accuracy (HC+FCS vs MCI vs AD) 44
Figure 3 Confusion Matrix showing Four-Way Classification Accuracy (HC vs FCD vs MCI vs AD) 45
Figure 4 Scatter plots demonstrating the results of manual and automated Troyer cluster analysis 58
Figure 5 Example Serial Recall Values for Control Subject 59
Figure 6 Image- Voxel Based Morphometry comparing Controls with MCI 62
Figure 7 Statistical Parametric Map (VBM) Controls compared to ADMCI. 64
Figure 8 Heatmap of Partial Correlations between Category Fluency Features and Hippocampal Volumes 67
Figure 9 Heatmap Demonstrating Correlations Between Semantic Features (VFT) 69
Figure 10 Correlation network of category fluency-derived semantic features. 70
Figure 11 Scatterplots Comparing Matched Semantic Features Between Category Fluency and Whole Interview Tasks 72
Figure 12 Overlay Forest Plot: Established Features in Controls vs MCI/ADMCI 76
Figure 13 Forest Plot: Linear Model comparing Imaging and Semantic Features between Healthy Controls and MCI 79
Figure 14 Forest Plot: Linear Model comparing Imaging and Semantic Features between Healthy Controls and ADMCI 82
Figure 15 Marginal Means Plot: Comparing Serial Recall Order between control and MCI groups. 84
Figure 16 Overlay Forest Plot: Comparing Differences in Category Fluency vs Whole Interview between Control and MCI groups. 85
Figure 17 Marginal Means Plot: Comparing Word Count and Filler word proportion between control and MCI groups. 86
Figure 18 Overlay Receiver Operating Characteristic (ROC) Curves for Logistic Models Based on Established Cognitive assessment and basic fluency Features 91
Figure 19 Individual Receiver Operating Characteristic (ROC) Curves for Logistic Models Based on Established Cognitive assessment and basic fluency Features. 91
Figure 20 ROC Curves demonstrating predictive power of imaging features. 92
Figure 21 Receiver Operating Curves demonstrating predictive power of individual semantic features. 94
Figure 22 Confusion Matrix for Clinician-Informed Model 99
Figure 23 Confusion matrix for features-optimized model. 101
Figure 24 Confusion Matrix demonstrating Random Forest Classification Performance 105
Figure 25 ROC Comparative Random Forest Models 108

List of Tables

Table 1 CognoSpeak Questions/ Tasks 35
Table 2 Demographic, psychological and imaging data for four-way comparison 42
Table 3 Contents of Lexical Dictionary including reference source 56
Table 4 Voxel Based Morphometry Comparing Controls and MCI 61
Table 5 Voxel-Based Morphometry Comparing Controls with ADMCI 63
Table 6 Significant Partial Correlations Between Category Fluency Features and Hippocampal Volumes 66
Table 7 Comparison of Semantic Features from Category Fluency and Whole Interview Tasks 71
Table 9 Hippocampal Volumes in Control and MCI groups 74
Table 10 Hippocampal Volumes in Control and ADMCI Groups 74
Table 10 Comparison of established features between MCI and control groups and ADMCI and control groups 75
Table 11 Descriptive comparison of VFT semantic features between groups 77
Table 12 Full results of Linear Model comparing Imaging and Semantic Features Between healthy controls and MCI 80
Table 14 Full results of Linear Model comparing Imaging and Semantic Features Between healthy controls and ADMCI 82
Table 15 Full results of Linear Model comparing Semantic Serial Recall Features Between healthy controls and MCI. 83
Table 16 Logistic regression demonstrating predictive performance of cognitive assessment and basic verbal fluency scores. 90
Table 17 Logistic Regression demonstrating the predictive power of imaging and semantic features (derived from verbal fluency task) 93
Table 18 Features with highest value in predicting diagnosis. 95
Table 19 Predictive Model Features 98
Table 20 Summary of Logistic Regression Model Using Clinician-Informed Features 100
Table 21 Summary of Logistic Regression Model using Feature-Optimized Model 102
Table 22 Random Forest Variable Importance 106
Table 23 Comparing performance between logistic regression and random forest model. . 107
Table 24 Diagnosis at time of writing 113
Table 25 Expected and observed behavior of language-based features in MCI 119

List of Abbreviations

ACE- Addenbrooke's Cognitive Examination

AChEI- Acetylcholinesterase Inhibitor

AD- Alzheimer's Disease

ADMCI- Alzheimer's Disease Mild Cognitive Impairment

ADRDA- Alzheimer's Disease and Related Disorders Association

AI- Artificial Intelligence

APOE- Apolipoprotein E

ARIA- Amyloid-Related Imaging Abnormalities

ARIA-E- Amyloid-Related Imaging Abnormalities- Edema

ARIA-H- Amyloid-Related Imaging Abnormalities- Hemorrhage

ASR- Automated Speech Recognition

ATN- Amyloid, Tau and Neurodegeneration

AUC- Area under the curve

CA- Conversation Analysis

CI- Confidence Interval

COVID- Coronavirus Disease 2019

CSF- Cerebrospinal fluid

CT- Computed Tomography

DICOM- Digital Imaging and Communications in Medicine

FCD- Functional Cognitive Disorder

FDG PET- Fluorodeoxyglucose Positron Emission Tomography

FDR- False discovery rate

FMD- Functional Memory Disorder

FND- Functional Neurological Disorder

FWE- Family-wise Error

FWHM- Full Width at Half Maximum

GDPR- General Data Protection Regulation

GFAP- Glial Fibrillary Acidic Protein

HC- Healthy Controls

IMPAX PACS- IMPAX Picture Archiving and Communication System

MCI- Mild Cognitive Impairment

MHRA- Medicines and Healthcare products Regulatory Agency

MMSE- Mini-Mental State Examination

MNI TEMPLATE- Montreal Neurological Institute

MOCA- Montreal Cognitive Assessment

MRI- Magnetic Resonance Imaging

NFL- Neurofilament Light

NFT- Neurofibrillary tangles

NIAA-AA- National Institute on Aging–Alzheimer's Association

NIFTI- Neuroimaging Informatics Technology Initiative

NINCDS- National Institute of Neurological and Communicative Disorders and Stroke

NMDA- N-methyl-D-aspartate

OR- Odds ratio

PET- Positron Emission Tomography

RF- Random forest

ROC- Receiver operating characteristic

TIV- Total Intracranial Volume

VBM- Voxel-Based Morphometry

VFT- Verbal Fluency Tasks

XNAT- Extensible Neuroimaging Archive Toolkit

1. Introduction and Background

1.1 Hypotheses

1. Unsupervised and remote interaction with a virtual clinical agent will permit the administration of verbal fluency tasks and provide data suitable for further analysis,

2. Semantic features extracted from speech will contribute towards the identification of prodromal Alzheimer's disease.

1.2 Introduction

This project aims to characterize the unique linguistic characteristics of patients with prodromal Alzheimer's disease (AD). I will explore how these linguistic features relate to traditional markers of disease and consider their applicability to an automated, minimally invasive and scalable assessment tool to aid in the early diagnosis of this condition.

AD is neurodegenerative process, defined by specific histopathological features, which leads to dementia. There is a prolonged preclinical stage of AD which progresses to a prodromal mild cognitive impairment (MCI) stage and then to AD-Dementia (Sperling et al., 2011).

Recent years have seen an increasing emphasis on the early identification of AD, driven by access to clinical trials and the imminent prospect of a disease modifying treatment. The move towards a biological, rather than syndromic definition of AD is finding significant traction. This is attributable, in part, to the role of imaging, CSF and serum biomarkers in establishing eligibility for, and demonstrating target engagement of, the emergent therapies (Jack et al., 2018). However, the interpretation of these biomarkers requires nuance and is likely to benefit from a robust and considered clinical correlate.

Longitudinal studies have demonstrated subtle semantic impairments in patients with preclinical and prodromal AD (Papp et al., 2016; Snowdon et al., 1996; Vonk et al., 2019) . These semantic impairments manifest in specific changes to the patient's use of language which can be assessed non-invasively and can be used to identify patients accurately (Blackburn et al., 2014). Specialist assessment of language can be arduous and expensive. An automated approach to language assessment has been developed by our team based on conversation analysis showing successful distinction of patients with normal memory and those with a neurodegenerative condition (Mirheidari et al., 2017, 2019). This now forms part of the foundational work

towards the development of a clinically useful automated diagnostic classification system.

Over the ensuing chapters, I plan to summarize the progress made towards an automated cognitive assessment tool. Based on the observations surrounding semantic impairment in prodromal AD, as well as abundance of semantic information contained in free speech, I will focus on language-based features that may inform diagnostic classification. I will assess the linguistic profiles of a well-described cohort of patients with prodromal AD. These markers would include results from manually administered tasks as well as from the automated system. I will explore the discriminative power of these features using established indicator of disease such as structural MRI to provide internal validation. I will explore whether machine learning techniques will contribute towards the development of a robust and transparent system.

Throughout the project I will maintain a clear focus on clinical utility, accessibility, and interpretability.

I posit that a distinct linguistic profile will be apparent in patients with prodromal AD and that automated assessment of language will have diagnostic value, with potential utility in clinical practice.

To address the hypotheses above, I will pursue the following questions:

1. Is an unsupervised, remotely administered cognitive assessment system capable of obtaining interpretable data suitable for further analysis?

2. Do candidate linguistic features' relationship to one another and selected regions of interest on structural imaging support their use as measures of semantic memory and processes?

3. Does the analysis of semantic function applied to participants' responses to standardized interview questions (administered by an automated system) resemble the analysis of the verbal fluency tasks?

4. Does the semantic content of a participant's speech, as collected using automated verbal fluency tests, enable accurate classification between Healthy Controls and patients with MCI?

5. Is the discriminant value of these semantic features comparable to established structural imaging features?

6. Do the language features, cognitive test scores and imaging features combined contribute to greater diagnostic accuracy than the information typically available after an initial appointment?

Throughout the following chapter I will discuss the pathophysiology of AD and its relationship to the observed clinical syndrome. I will evaluate the competing diagnostic criteria and summarize the myriad biomarkers which are rapidly assuming a prominent position in the minds of patients and clinicians alike. To put this fervor into context, I will discuss the disease modifying therapies currently available.

I will then focus on providing a detailed description of the linguistic profile observed in AD, exploring the processes underpinning this and the tools that have been employed in their characterization to date. This will position the reader optimally to proceed through the questions above.

1.3 Background

1.3.1Alzheimer's Disease

Originally described in 1906 by Alois Alzheimer (in the brain of Auguste Deter), (Cipriani et al., 2011) Alzheimer's disease is the most common cause of dementia, a syndrome characterized by impaired cognition sufficiently severe to interfere with daily activities and social functioning. AD is estimated to account for between 50-75% of dementia cases (Prince et al., 2015). Across Europe the age-standardized prevalence of AD dementia is 4.4%. This figure increases exponentially with age, reaching 28.5% by the age of 90 (Lobo et al., 2000). It is mostly sporadic in the elderly and is responsible for about 11.2% of years living with a disability in the over 60s (Walt, 2004). AD imposes a substantial and increasing economic burden to society with annual per-patient costs increasing form $18,517-$47,000 in the 1980's to $171,283 in recent years (Tay et al., 2024). Costs include direct factors such as medical care and social support as well as indirect factors such as informal caregiving and lost productivity (Rice et al., 1993). Alongside these broader societal considerations, AD is associated with profound personal consequences for patients, caregivers and relatives (Coon & Edgerly, 1999).

Familial AD is rare, accounting for less than 0.5% of cases (Masters et al., 2015). It is typically autosomal dominant, with mutations in genes pre-senelin1 pre-senelin2 and APP most frequently implicated (Wu et al., 2012). The average age of onset for these patients is comparatively lower than for sporadic AD, frequently reported in patients younger than 65 (and often in their 40's) (Karve et al., 2012) compared to the sporadic group 65-75 respectively (Jack et al., 2018; Karve et al., 2012).

There is an increasing recognition of genetic factors contributing towards supposed "sporadic disease" with up to 70% of AD risk being attributable to genetic factors. The apolipoprotein (APOE) e4 is well recognized to confer an increased risk of developing AD with homozygotes facing an up to 10 times higher risk of developing AD compared to e3 homozygotes (Slooter et al., 1998). Testing for APOE e4 status has not formed part of routine clinical practice. However, this is likely to change given that eligibility for the amyloid-based disease modifying treatment is conditional

on APOE e4 status due to the risk of complications (Zimmer et al., 2025). Other risk factors for AD include hypertension, hypercholesterolemia, smoking (Meng et al., 2014) and diabetes (Xu et al., 2015).

AD is associated with the accumulation of extracellular amyloid-beta (Aβ) and intra-cellular tau proteins (Selkoe & Hardy, 2016). Aggregates of Aβ, derived from the amyloid precursor protein (APP), form toxic oligomers, initially as fibrils and later as plaques, that are harmful to the surrounding neurons (Haass & Selkoe, 2007). The toxic effect of Aβ are widespread, with evidence of disrupted mitochondrial function, synaptic function and inflammation and ultimately neuronal death (Palop & Mucke, 2010; Wang et al., 2007). Aβ typically begins in the frontal brain regions and spreads to allocortical and brainstem areas (Raj et al., 2016). The pattern of spread appears to be determined by proximity (reflecting the extracellular nature of Aβ) rather than neuronal connectivity (Kim et al., 2019). The APOE e4 allele appears to exacerbate AD pathology by impacting on aggregation, clearance and inflammatory response (Tzioras et al., 2019).

Tau forms part of the internal scaffolding of the neurons, contributing to structural stability as it binds to microtubules (Esmaeili et al., 2023). In AD abnormal hyperphosphorylation of Tau leads to misfolding, aggregation and accumulation, forming intracellular neurofibrillary tangles (NFT) (Thal & Tomé, 2022). These NFT's can disrupt axonal transport leading to mitochondrial dysfunction and synaptic degradation (Mandelkow, 2008). Hyperphosphorylated Tau is sufficiently harmful to neuronal function to result in dementia syndromes in the absence of AD amyloid pathology (Rosenmann et al., 2012). The distribution and extent of abnormal Tau is closely correlated to the severity and phenotype of the clinical syndrome (Robinson et al., 2020). Typically Tau is observed earliest in the entorhinal cortex before becoming more widely distributed (Thal & Tomé, 2022). Abnormal hyperphosphorylated Tau aggregates propagate through connected neurons and can induce a conformation shift in the other Tau isoforms, resembling the behavior of a prion (Diamond, 2024).

Braak et al divide the pathological spread of AD (defined by the presence of NFT's) into 6 stages (which can be more broadly considered 1-2, 3-4, 5-6). The earliest site of involvement is the transentorhinal cortical region. From there the disease spreads into the transentorhinal region. Stages 3 and 4 are defined by the involvement of the limbic regions and hippocampus. In stages 5 and 6 the lesions spread, fan like throughout the neocortex (Braak et al., 2006).

The amyloid cascade theory posits that Aβ deposition is the inciting event in AD, leading to hyperphosphorylation of Tau and the combined downstream toxic effects of both. This is supported by the observation that the genes implicated in familiar AD all relate to amyloid, with pathogenic mutations resulting in Aβ deposition (Reitz, 2012).

However, questions as to the validity of the amyloid hypothesis have been raised. Studies using transgenic mouse models, over-expressing both A□ amyloid plaques and A□ oligomers, failed to demonstrate either the formation of NFT's or neuronal cell loss (Kim et al., 2013). Furthermore, studies with amyloid binding ligands PET scans have reported significant amyloid deposition in cognitively normal people, as well as patients with an AD clinical phenotype but relatively little amyloid (Fagan et al., 2009; Li et al., 2008). These concerns may have been allayed somewhat by the (admittedly modest) success of the anti- Aβ antibody treatments donanemab (Sims et al., 2023) and Lecanemab (Van Dyck et al., 2023).

1.3.2 Mild Cognitive Impairment

MCI is characterized by the impairment of cognition in excess of what is to be expected in normal ageing whilst retaining the ability to function in everyday life (McKhann et al., 2011; Petersen, 2004). MCI can be classified in two main subcategories, amnestic and non-amnestic, with further subgrouping depending on whether one or multiple cognitive domains are involved (Petersen et al., 2014).

Comprehensive neuropsychological testing is required to classify MCI into specific subtypes. Whilst no definite "cut-offs" are specified, research guidelines suggest a performance of 1.5 standard deviations below age matched norms to reflect an objective impairment in cognition (Petersen, 2004; Petersen et al., 2014).

It is felt that the subtype of MCI may relate to the underlying pathophysiological process, and therefore enable prediction on how the condition may progress (Petersen, 2004). Longitudinal data from the Sydney Memory and Ageing Study supports this suggestion, demonstrating that participants with amnestic MCI had a significantly higher rate of progression to AD compared with other subtypes (Aerts et al., 2017). The prevalence of MCI in adults aged > 65 is between 10-20% (Langa & Levine, 2014; Prince et al., 2015). Conversion rates from MCI to dementia are approximately 15% (range <5-20% per year) (Farias et al., 2009; Plassman et al., 2008).

Other risk factors for progression to AD include older age, fewer years of education, comorbid stroke and diabetes, having the amnestic subtype of MCI (a-MCI) (Campbell et al., 2013) and carrying the Apolipoprotein epsilon 4 (*APOE* □4) allele (Elias-Sonnenschein et al., 2011; Mitchell & Shiri-Feshki, 2009). At present there are no neuro-imaging or laboratory tests recommended for predicting MCI progression to dementia in clinical practice (Jack et al., 2018).

MCI due to AD may be considered an intermediate point between preclinical AD and Alzheimer's dementia (Jack et al., 2018; Sperling et al., 2011) and thus provides an opportunity to investigate the early pathogenesis of sporadic AD as well as the efficacy of potential treatments.

1.3.3 Diagnostic Criteria

Traditionally, a definite diagnosis of Alzheimer's disease could only be made on the basis of post-mortem histopathological assessment. Varying levels of diagnostic certainty could be reached based on clinical features and adjunct testing (Jack et al., 2011; McKhann et al., 2011). The development of biomarkers has led to a more nuanced understanding of the pathogenesis behind AD that is now thought to be best conceptualized as a biological and clinical continuum, ranging from preclinical to clinical (Aisen et al., 2017). Efforts have been made, therefore, to construct a diagnostic/ research framework, using widely agreed terms to refer to the various stages and how they relate to the biomarker changes and, thus, underlying pathophysiological changes at play.

Various attempts at creating a unified framework have been made (Dubois et al., 2014; Jack et al., 2011). Prior to 2011 the recommendations provided by the Alzheimer's Association workgroup and National Institutes of Health Workgroup (NINCDS-ADRDA) were considered the practice standards. These criteria assumed close correlation between the clinical syndrome and underlying pathophysiology. Acknowledging the technological limitations of the time (namely the absence of any one confirmatory pre-mortem test) the NINCDS-ADRDA criteria relied on clinical and demographic features to permit a diagnosis of "Probable Alzheimer Disease". To fulfil the criteria an individual had to demonstrate evidence of an objective, progressive cognitive impairment over at least two domains in the absence of a more likely explanation (McKhann et al., 2011).

By 2009 the tide of opinion was beginning to shift. Significant discrepancies between phenotype and pathology, as well as more reliable diagnostic technologies led to an erosion of the pre-eminence previously assigned exclusively to the clinical presentation. MCI was reframed as prodromal AD (Jack et al., 2011)

A major shift was seen on release of the 2018 NIAA-AA diagnostic criteria. This aimed to provide a shared lexicon to describe both the biological changes observed in AD as well as the clinical manifestations (Jack et al., 2018). These recommendations superseded previous guidance related to the biological classification of MCI (Jack et al., 2011). The group proposed the "ATN" system, with each letter reflecting the presence (using a +) or absence (using a -) of abnormal amyloid (A), tau (T), and neurodegeneration (N). The presence of each feature is informed by related biomarkers and formed the "biological" description of the disease process. A numeric clinical staging system to describe the severity of the symptoms and provide greater descriptive resolution than the previous MCI and dementia labels, This staging system ranges from 1 to 6, with stage 1 characterized by a complete absence of disease signs of symptoms and stage 6 being characterized by a severe cognitive impairment (equivalent to severe dementia) (Jack et al., 2018).

This position has been fortified with the revised criteria for diagnosis and staging of Alzheimer's disease circulated in 2024. The Alzheimer's association advocate for a

complete conceptual decoupling of phenotype and etiology. This statement introduces three additional letters to the coding taxonomy. Alongside A, T and N; I, V and S have been added. These stand for Inflammatory/ Immune, vascular disease and alpha-synucleinopathy (with the latter two included to encompass the likelihood of co-pathology in older populations) (Jack Jr. et al., 2024).

Whilst the 2018 framework was positioned for use primarily in the context of observational and interventional research studies, the Alzheimer's Association 2024 statement suggests a broader application towards real world diagnosis. This taxonomy has been widely adopted in research settings, with interventional trials mandating the biological evidence of AD. This serves to ensure a homogeneous cohort and also provides a metric to assess "target engagement" of the trial therapy. It also permits enrolment into studies at an earlier stage. The efficacy of treatments directed at the underlying pathogenic mechanisms is presumed to benefit from early delivery, before significant neurodegeneration (and thus cognitive decline) has occurred. Thus, the move towards a predominantly biological classification of AD has led to the ubiquitous use of biomarkers in research protocols (Klein et al., 2019; Mignon et al., 2018). With the prospect of disease modifying therapies and the broader application of the most recent criteria, biomarkers (and ATN style diagnoses) are becoming increasingly standard in clinical practice (VandeVrede & Schindler, 2024).

In 2024 Dubois et al (international working group) published: Alzheimer disease is a clinical-biological construct: An IWG recommendation. They conceptualize the biomarkers as AD risk factors rather than being sufficient to make the diagnosis. In clinical use they suggest that these markers be used to support or refute a clinically suspected diagnosis.

Alzheimer neuropathologic changes (as observed) are necessary but not sufficient to diagnose AD. These changes are frequently observed in on post-mortem in patients with no symptoms of the disease. Life-time risk of Alzheimer's Dementia in a 65-year-old, amyloid positive man is 21.9 % only 1.7 times higher than if he did not have amyloid. These findings have been replicated in ADNI (lack of significant progression over 8 years).

They suggest a relabelling from preclinical AD to "at risk for AD" is more accurate and avoids the personal and societal impacts of a false positive diagnosis (Dubois *et al*., 2024).

1.3.4 AD Biomarkers

Biomarkers are measurable indicators of biological processes, pathological changes or responses to intervention (Rong & Wang, 2012). Recent years have seen significant progress in the development of accurate AD biomarkers (Blennow &

Zetterberg, 2018). These advances have informed the reconfiguration of the AD diagnostic and staging criteria (Jack Jr. et al., 2024).

Jack et al have proposed a stratification of biomarkers based on their relationship to underlying pathology and their presence with relation to time. Core AD biomarkers are those which directly relate to the biological changes observed in AD (the accumulation and spread of abnormal amyloid and tau). These include CSF and serum measures of AB42/20 and phosphorylated tau as well as amyloid and tau PET imaging. Biomarkers of downstream neurodegeneration, inflammation or co-pathology with vascular or alpha-synuclein drive disease such as neurofilament light, GFAP and structural imaging features of vascular disease or atrophy are considered non-core.

The core biomarkers are further stratified into Core 1 and Core 2 depending on when they are observed to be abnormal in the course of the disease. Concordant with the amyloid hypothesis, amyloid based CSF and imaging features are the earliest observable markers of AD (Blennow & Zetterberg, 2018).

Longitudinal studies indicate that CSF AB42 changes are observed first (around an average age of 48), followed by amyloid PET changes (around the age of 54) followed by tau PET changes (around the age of 60) (Luo et al., 2020).

Early changing “Core 1” biomarkers include amyloid PET, CSF biomarkers of abnormal amyloid and tau handling and accurate blood based biomarkers (with evidence suggesting that P-Tau217 proposed as the most useful (Janelidze et al., 2020). At the time of writing, only CSF and amyloid PET are approved for use. The presence of any one Core 1 biomarker is sufficient for the diagnosis of AD. (Jack Jr. et al., 2024)

Tau related Core biomarkers can be divided into T_1 or T_2 . T_1 markers are detectably abnormal from around the same time as amyloid markers and include blood and CSF analytes of soluble tau fragments (e.g. P-Tau217). T_2 markers become abnormal later and include tau PET imaging. T_1 biomarkers are included in the “Core 1” category. T_2 biomarkers are labelled “Core 2”. presence or absence of Core 2 biomarkers, can be used for prognostication and titrating confidence that the clinical picture is due to AD. The Core 2 biomarkers can be used to inform prognosis (with more advanced biological disease staging being associated with a worse prognosis) and confidence that the clinical picture observed relates to AD (for example, a patient with a severe degree of impairment but absent Core 2 and N biomarkers may suggest against AD as the most likely cause) (Jack Jr. et al., 2024).

1.3.4.1 Imaging Biomarkers

Structural MRI serves two purposes: to exclude another cause for impaired cognition and to provide positive diagnostic support by detecting regional atrophy. Regional

atrophy typically first manifests in the medial temporal lobe (Johnson et al., 2012) and has a sensitivity and specificity of 50-70% in predicting MCI's progression to AD (DeCarli et al., 2007). Hippocampal volumetry is likely to be more accurate than visual assessment but is not widely available (Scheltens et al., 2016). Automated hippocampal segmentation algorithms can be implemented as a research tool but their clinical utility is limited by availability (Boccardi et al., 2015). Changes on structural imaging serve as proxies for neurodegeneration (the N of the ATN classification system) and correlate with AD pathology (Vemuri & Jack, 2010).

FDG PET reflects synaptic activity (Rocher et al., 2003) and, in the context of AD, detects hypometabolism that correlates closely with those areas typically affected by the disease. These areas include the medial temporal lobes and the parietal lobes. FDG PET changes in the entorhinal cortex predict conversion from pre-clinical AD to MCI and symptomatic AD in longitudinal studies (de Leon et al., 2001). However, hypometabolism is not specific for AD and may also be observed in the context of alternative, or additional pathologies such as cerebrovascular disease (Johnson et al., 2013). FDG PET is expensive and fails to reflect molecular changes specific to AD.

Amyloid PET detects ligand binding in areas that reflect the typical distribution of A□ in AD on autopsy (Clark et al., 2012) and is considered a Core 1 biomarker (Jack Jr. et al., 2024). A review of 15 research studies of amyloid PET on clinically diagnosed AD patients found that 96% of these patients were positive for amyloid (Johnson et al., 2012). Five longitudinal studies of MCI (155 patients overall) over 1-3 years showed that, of those patients who progressed to AD, 93% were amyloid PET positive at baseline (Johnson et al., 2012). Though some concerns have been raised regarding the clinical meaningfulness of reduced amyloid on PET when interpreting the efficacy of anti-amyloid treatments (Høilund-Carlsen et al., 2023) recent data suggests that it correlates with slowed disease progression, supporting its role as a surrogate marker in clinical trials (Willis et al., 2024). The main limitations of amyloid PET include the high cost and limited availability (Petersen & Mayo Alzheimer's Disease Research Center, 2018).

A number of Tau PET ligands have been developed over the last few years (Femminella et al., 2018) with recent compounds demonstrating improved selectivity and binding (Kroth et al., 2019; Okamura et al., 2018) . Though areas of tracer uptake of Tau PET demonstrate close negative correlation with FDG-PET (Chiotis et al., 2018), there are concerns regarding reliability (Femminella et al., 2018). Changes on Tau PET are observed later on in the disease (Luo et al., 2020). It is considered a Core 2 biomarker and can be used to stage AD (contributing to prognostication and interpretation of the clinical picture) but not diagnosis (Jack Jr. et al., 2024). The regional distribution and the amount of abnormal Tau on PET closely correlate to clinical presentation (Ossenkoppele et al., 2016; Tetzloff et al., 2018). Tau distribution in atypical AD variants such as posterior cortical atrophy correspond to the clinical phenotype closely (Okamura et al., 2018). The Alzheimer's Association

suggests a standardized nomenclature to describe the presence, strength of signal and distribution of Tau PET abnormalities (Jack Jr. et al., 2024).

1.3.4.2 CSF Biomarkers

CSF A□42 is reduced in AD patients (Blennow & Zetterberg, 2018), reflecting increased plaque deposition (Fagan et al., 2009). This is most useful when interpreted as a ratio of AB42/40 (Parnetti & Eusebi, 2018). High concordance has been demonstrated between CSF A□42 levels and amyloid burden on PET scanning in both cross-sectional and longitudinal studies (Blennow & Zetterberg, 2018). Exceptions to this have been observed however, with some elderly, non-AD individuals having low CSF A□42 and normal PET. These individuals accumulate amyloid on longitudinal PET scanning at a rate three times higher than individuals with normal CSF A□42 levels. This observation suggests that CSF A□42 may be an early indicator of incipient AD pathology (Palmqvist et al., 2016). CSF Amyloid biomarkers belong in the Core 1 category and are sufficient, if abnormal, to diagnose AD (Jack Jr. et al., 2024).

A significant increase in CSF total tau (t-tau) is found in AD patients (Olsson et al., 2016) but this is non-specific and has also been observed in people with head injury and Creutzfeldt Jacob disease. CSF phosphorylated tau (p-tau) has been found to be more specific, enabling discrimination between AD and other causes of neuronal damage (Blennow & Zetterberg, 2018). The predictive value of these markers has been confirmed in longitudinal studies of MCI. For example, abnormal levels of CSF Amyloid and Tau at baseline predicted conversion to AD with a sensitivity of 95% and a specificity of 87% over a period up to 6.8 years (Hansson et al., 2006). CSF tau changes have been shown to occur ~15 years before the onset of clinical AD (Bateman et al., 2012; Fagan et al., 2009), and decline in CSF A□42 is extrapolated within subject longitudinal analysis up to 20 years before symptom onset.

Phosphorylated mid-region tau variants (such as P-Tau217) become abnormal around the same time as amyloid PET, whilst other Tau fragments become abnormal later. The category of Core Biomarker is therefore dependent on the tau fragment being considered.

Neurofilament light (Nfl) is a non-Core biomarker that serves as a proxy for neurodegeneration (Mattsson et al., 2017). It is significantly elevated in patients with AD and MCI when compared to healthy controls (Fan et al., 2023) and is used to inform the N label of the expanded ATN biological classification (Jack Jr. et al., 2024).

Other potential CSF biomarkers include A□ oligomers, proposed to be the pathogenic form of A□, and neurogranin, a marker of synaptic integrity. These biomarkers are promising, but do not have wide clinical or research application yet (Scheltens et al., 2016).

1.3.4.3 Serum Biomarkers

Blood based biomarkers have been of increasing interest over the last few years, representing an attractive alternative to CSF and imaging. They are less invasive and have greater scaling potential (Chohan et al., 2022a).

Blood based biomarkers can be broadly categorized by pathological process they represent. Core biomarkers include Amyloid markers (AB42/40 oligomer ratio) and Tau Markers (including the phosphorylated midrange fragments like P-Tau217

Non-core biomarkers include glial fibrillary acidic protein (GFAP), a non-specific marker of CNS inflammation seen to become elevated in AD (Ozcelikay-Akyildiz et al., 2024) and neurofilament light (Nfl), a marker of neurodegeneration (Mattsson et al., 2017).

Grande et all conducted a longitudinal analysis investigating the predictive power of a variety of blood-based biomarkers. The demonstrated that elevated baseline levels of p-tau181, p-tau217, nFL and GFAP demonstrated strong predictive performance for 10 year all cause and AD dementia (with AUC's ranging between 0.71 and 0.82) (Grande et al., 2025).

On an individual level, the accuracy of the various blood based core biomarkers varies significantly, with their ability to predict amyloid PET status spanning AUC's between 0.6 and 0.9 (Janelidze et al., 2021, 2023). Overall P-Tau appears to perform the best , with excellent discriminant ability between AD and other neurodegenerative pathologies, with performance comparable to CSF and PET measures (Palmqvist et al., 2020).

At the time of writing none of the blood-based biomarkers have received regulatory approval for use in clinical practice. Work continues to demonstrate the real-world application of these novel measures, with two large trials ongoing in the UK at present as part of the Blood Biomarker Challenge (a multi-million-pound initiative led by the Alzheimer's society). The READ-OUT (REAL world Dementia OUTcomes) trial will multiple existing blood-based biomarkers over a number of cognitive conditions whilst the ADAPT (Alzheimer's disease Diagnosis and Plasma pTau217) study will focus specifically on the value of P-Tau217.

1.3.5 Therapies

1.3.5.1 Symptomatic

Symptomatic therapies for AD aim to provide a boost to the patients cognition. As such they are often referred to as cognitive enhancer. They include the acetylcholinesterase inhibitors (AChEIs) donepezil, galantamine, and rivastigmine (reflecting the cholinergic deficit consistently observed in AD (Palmer, 2003) and the NMDA receptor antagonist memantine (which may have a protective effect against Aβ (Lipton, 2005).

Donepezil is the most frequently used cognitive enhancer in the UK and the US (DiBello et al., 2023). In the UK approximately 65% of newly diagnosed AD patients are commenced on a cognitive enhancer. Donepezil is associated with a statistically significant improvement in measures of cognition (such as MMSE), with the effect most marked in the moderate stages (Cummings et al., 2010), and is generally safe and well tolerated (Rogers et al., 2000). There is evidence that combination therapy is most effective. Combination therapy with AChEIs and memantine is considered optimal (Cummings et al., 2019).

1.3.5.2 Disease Modifying

The pursuit of effective disease modifying therapies has been a long-standing goal in AD research. Repeated failures of apparently promising lines of enquiry called into question our understanding of the basic mechanisms underpinning the disease (such as the validity of the amyloid hypothesis (Mullane & Williams, 2020).

Recent years have seen what many people in the field consider to be a breakthrough. Landmark trials of novel agents directed towards Aβ have demonstrated both target engagement and clinical effect (Sims et al., 2023; Van Dyck et al., 2023).

Aducanumab is a monoclonal antibody targeting Aβ. The EMERGE and ENGAGE trials were 18-month, phase 3 studies evaluating the effect of aducanumab on the rate of progression of cognitive impairment in biologically proven AD. They both failed to demonstrate any significant benefit, though on further analysis and with further follow-up, positive results were reported. This resulted in the FDA approving the medication for the treatment of AD in the US. Concerns surrounding methodological approach and the decision to approve the medication were raised (Kuller & Lopez, 2021) and Biogen have since been withdrawn the medication.

Donanemab is monoclonal antibody directed against insoluble (pathogenic) Aβ. The results of 17-month, multicenter, double-blinded, placebo-controlled phase 3 trial (TRAILBLAZER-ALZ 2) were first published in July 2023. This trial demonstrated a statistically significant difference in the rate of change of the Alzheimer Disease

Rating Scale score between groups, suggesting a modest slowing of deterioration in the treatment group (Sims et al., 2023).

Lecanemab is another humanized monoclonal antibody targeting soluble and insoluble Aβ (Niidome et al., 2024). The results of an 18-month, multicenter, double-blinded, placebo-controlled phase 3 trial (CLARITY-AD) involving around 1,800 patients were first published in November 2022. These demonstrated that patient treated with Lecanemab experienced moderately less decline on measures of cognition (including the Clinical Dementia scale, ADCS-MCI-ADL, ADAS-Cog and ADCOMS measured) over the trial period when compared to the placebo group. A sub-study of 698 patients demonstrated a reduction in amyloid burden on PET when compared to the placebo group (Van Dyck et al., 2023).

Both of these medications have demonstrated modest benefit in slowing the rate of progression in biologically proven AD, with Donanemab outperforming Lecanemab with regards to cognitive outcomes, and Lecanemab outperforming Donanemab with regards to reducing amyloid burden (Jeremic et al., 2024). Both have been approved for use by the FDA. The MHRA in the UK have provided a license but neither are approved for use by NICE due to concerns regarding cost effectiveness (‘NICE Rejects Second Alzheimer’s Disease Treatment for NHS Use’, 2024)

Aside from the expense, there are also valid safety concerns surrounding these medications. All of the anti-amyloid therapies described above (as well as others omitted for the sake of brevity) are associated with a risk of amyloid-related imaging abnormalities (ARIA) (Erichsen, 2024). ARIA can be divided into two categories, ARIA-E, which is characterized by vasogenic oedema, and ARIA-H, characterized by hemorrhages and/or hemosiderosis (Roytman et al., 2023). CAUSE. Risk factors for ARIA include the APOE-e4 allele, cerebral amyloid angiopathy, higher drug doses and concurrent antithrombotic therapy. ARIA may be asymptomatic, though headache, confusion and death have been reported (Doran & Sawyer, 2024). ARIA is likely to be due to increased vascular permeability due to an inflammatory response to the drug, direct antibody binding to vascular amyloid, or a combination of the two (Greenberg, 2024).

A meta-analysis of 8 phase 3 RCT’s (looking at 9 cohorts) was conducted to evaluate the safety of the anti-amyloid agents. This reported an adjusted, pooled incidence of symptomatic ARIA-E of 6.7% (and severe ARIA-E of 3.5%). The pooled incidence of ARIA-H was 17.8%. Homozygotes for APOE-e4 were significantly more likely to develop ARIA-E than non-carriers (OR 5.6). This increased risk was observed in APOE-e4 heterozygotes also, to a letter extent (OR 1.9) (Jeong et al., 2025).

Despite not receiving approval for use in the UK, the existence of effective disease modifying agents to AD has altered the landscape significantly. The eligibility criteria for the treatments above mandates a biologically proven diagnosis. This will almost

certainly change the practice of most memory clinics in the UK. The prospect of an effective treatment becoming available in the foreseeable future may be anticipated to motivate patients to seek earlier diagnoses, with increasing numbers of younger patients seeking assessment. This is likely to test the limits of services unused to dealing with such high numbers of patients.

1.3.6 Language and AD

1.3.6.1 Semantic Memory

Semantic memory is one of two components of declarative memory, the other component being episodic. Declarative memory can be contrasted with implicit or non-declarative memory such as conditioned responses. It is a system that permits the conscious recollection of facts and events (Squire, 1992)

The distinction between episodic and semantic memory was first drawn by Tulving, who posited the existence of two, discreet memory systems, one applied to autobiographical, context riche information (episodic) and the other applied to the general knowledge of the world (semantic) (Tulving, 1972).

The distinction between these systems (and by implication their neuroanatomical substrates) appeared to be supported by observations in patients with severe amnestic syndromes. Some were demonstrated to retain the ability to acquire and retain semantic information far out of proportion to their episodic abilities (Vargha-Khadem et al., 1997). Others were found to perform less well, with the semantic domain still preserved relative to the episodic (Glisky et al., 1986) whilst others still demonstrated severe impairment in both domains (Gabrieli et al., 1988).

The issue can be examined from the other perspective. Patients with semantic dementia, a rare neurodegenerative condition associated with the degeneration of semantic function, have been observed to retain episodic memory function until reasonably advanced stages of the disease (despite atrophy of the hippocampus) (Irish et al., 2016). Again however, the picture is neither simple nor uniform. Patients with semantic dementia are reported to experience impaired conception of the future (an episodic function) (Irish et al., 2012) and have significantly better retention for recent events compared to remote (Hodges & Graham, 2001).

Reviewing the multitude of conflicting and contradictory reports, it may be safest to consider the components of declarative memory to be distinct, but overlapping with some shared neuroanatomical correlates (including the hippocampus) (Duff et al., 2020).

Semantic memory is more difficult to conceptualize than episodic (ironically). Episodic memory can be thought of as an awareness of subjective experiences over time (Endel Tulving, 2002), an internal monitor on which to review the videos of your life. Semantic memory meanwhile is rather more abstract. Semantic memory

describes our explicit knowledge of everything that we encounter from objects to more abstract notions like remote historical situations and their lasting impact on the political landscape. It describes concepts and their representations in the mind.

Two related models of semantic memory compete for dominance within the literature.

Binder proposes the “embodied abstraction” model. This model suggests that semantic representations are formed in modal specific domains distributed throughout the cortex. One stimulus may activate numerous regions (Damasio, 1989). This is referred to as embodied cognition and is supported by neuroimaging and behavioral studies (Thompson-Schill et al., 1997). Selective impairment for the comprehension of action verbs has been demonstrated in patients with motor neuron disease (Grossman et al., 2008).

Above these modal specific domains are “convergence zones”. There areas include the inferior parietal cortex and the lateral and ventral temporal cortexes (Binder & Desai, 2011). Their involvement is supported by the non-selective deficits in comprehension observed in patients with selective atrophy of the temporal lobe (as seen in semantic dementia) as well as functional imaging studies which demonstrate consistent activation of these areas when engaged in a variety of semantic tasks (Binder & Desai, 2011).

Binder suggests a system that processes semantic representations different levels, ranging from the modal specific (such as the sensory connotations of the word) to the more abstract (handled by the convergence zones) as most consistent with the functional imaging and behavioral studies. These modal and convergence zones represent the sematic stores whilst prefrontal regions control top-down activation (through attention) and selection of content (Binder & Desai, 2011).

Another is the “hub and spoke” model proposed by Pobric et al in 2010. This model shares many of the elements of Binder’s but asserts the role of a solitary hub in the left anterior temporal lobe towards which all of the modal “spokes” feed (Pobric et al., 2010).

1.3.6.2 Semantic Memory in AD

Historically AD has been considered to be a disorder, predominantly, affecting episodic memory at least in the earlier stages (Graham & Hodges, 1997). In recent years there has been increasing interest in impairments of language, specifically those aspects of language relying on the integrity of semantic memory, with some proposing that the initial pathological changes to the anterior medial temporal lobe may be responsible for impaired semantic memory years before episodic memory becomes affected (Didic et al., 2011). Braak stage 1 is defined by NFT’s in the transentorhinal region, an area recognized to be closely associated with semantic memory (Davies et al., 2004). It is intuitive therefore, that semantic memory would represent an area of particular and early vulnerability.

There is a growing body of evidence that supports this hypothesis including one retrospective, longitudinal study that compared linguistic features in autobiographical diaries (written decades before death) and neuropathological findings at autopsy. This demonstrated a significant relationship between semantic linguistic characteristics decades before death, and the presence of AD related histopathological changes (Snowdon et al., 1996).

A similar approach was taken by Garrard et al, who assessed the linguistic content of the works of the author Iris Murdoch at various stages throughout her career, prior to her diagnosis of AD (which was histopathologically confirmed after death). This approach was useful due to the idiosyncratic habits of Murdoch as an author; she declined all editorial input. Whilst there was no observable change in the structure or grammatical arrangements in her work over time, a marked decline in lexical complexity was observed between her earliest and later work. This lent further support to the suggestion that semantic involvement is an early feature of AD pathology (Garrard et al., 2005) which was supported further by a second analysis demonstrating declining semantic complexity from her late 40's (Le et al., 2011).

With a view to assessing semantic function, Vogel et al administered tailored psychological tests to 102 patients with mild AD (MMSE exceeding 19), 22 patients with MCI and 58 healthy controls. They demonstrated significant between group differences. These were most apparent in the category fluency test and naming of famous faces (Vogel et al., 2005).

Prospective studies of language change support this finding, with subtle changes being apparent years, and in some cases, decades before the patient or family becomes aware of any cognitive impairment (Ahmed et al., 2013; Forbes-McKay et al., 2005). This has led to increased interest in identifying features of semantic memory that can be used as biomarkers in the prodromal stage of the illness.

Joubert et al demonstrated semantic impairment in patients with amnestic MCI and AD when compared to healthy controls. Voxel based morphometry revealed significant cortical atrophy in the disease groups affecting the anterior temporal lobe and the inferior prefrontal cortex. (Joubert et al., 2010).

The power of semantic features to predict progression from MCI to AD was explored by Marra et al. Over a 5-year period cohorts of 251 patients with MCI (178 classified as mild and 114 classified as moderate) 114 patients with moderate AD and 262 healthy controls were followed. The participants underwent testing sensitive to semantic and episodic memory. A meta-feature, the difference between the category fluency and phonemic fluency score (and referred to as the semantic-phonemic delta (SPD) was also calculated. A general linear model demonstrated a significant effect of the SPD. The SPD and the MMSE were the only features that contributed towards the prediction of which MCI patients would progress to AD (Marra et al., 2021).

Support for the predictive powers of features related to semantic memory were further bolstered by the observation that scores of lexical-semantic performance in individuals with preclinical AD due to Presenilin mutations compared with matched controls were significantly lower (Arango-Lasprilla et al., 2007). A similar effect was observed when comparing amnestic MCI APOE-e4 carriers with non-APOE-e4 carrying healthy controls (Biundo et al., 2011).

1.3.6.3 Assessing Semantic Function

As demonstrated, semantic memory assessment offers valuable information in the assessment of cognitive complaints. It has been observed to predict the development of AD by periods of up to decades and can distinguish preclinical AD from healthy controls.

Various tools have been developed to evaluate different aspects of semantic memory. These include batteries assessing multiple semantic functions (Ohman et al., 2022) tests of naming, category fluency, and semantic recognition (Tchakouté et al., 2017), and assessments of autobiographical and personal semantic memory (Kopelman et al., 1989; Schmidtke & Vollmer-Schmolck, 1999). By far the most common measure of semantic function in use is the verbal fluency task. I will discuss a selection of features which have demonstrated discriminant value between AD and controls. This is not intended to be comprehensive but serves as an explanation for their inclusion in the later experiment.

Verbal Fluency tasks

Verbal fluency tasks involve generating as many words as possible within a time limit, conforming to certain restrictions. In category fluency tasks the participant is limited to producing words belonging only to a certain category. In phonemic fluency tasks the participant is tasked with producing as many words as they can, beginning with a given letter and excluding proper nouns. The correct number of responses are recorded as well as the number of errors made (Monsch et al., 1997a).

Category fluency predominantly tests semantic function and maps to temporal cortex performance. Phonemic fluency is more directed towards the assessment of executive function and is primarily mediated by the frontal cortex (Baldo et al., 2006; Henry & Crawford, 2004).

Category fluency has been shown to be particularly sensitive in the detection of AD (Monsch et al., 1997a) presumably due to its ability to assess semantic performance.

Clustering and Switching

Clustering and switching describes the production of related words (clustering) before changing to another category (switching). Clustering is associated with temporal lobe functioning (representing the integrity of the semantic stores) whereas switching is associated with frontal lobe functioning (representing executive function). Troyer et al explored this by evaluating the performance of two patients groups, one with frontal lobe lesions and one with temporal lobe lesions (as well as a control group). They found that patients with temporal lobe pathology performed normally on phonemic fluency testing but produced smaller clusters than controls on category fluency testing. Patients with frontal lesions switched less frequently than controls but produced clusters of normal size (Troyer, Moscovitch, Winocur, Alexander, et al., 1998).

Gruenewald et al demonstrated that time between clusters tends to increase over the course of the task whilst time between words within clusters remains stable. The inter-cluster time is posited to represent a search process whilst the time within cluster demonstrates memory structure/ store (Gruenewald & Lockhead, 1980). The number of switches correlates with frontal lobe function (Hirshorn & Thompson-Schill, 2006).

Age of Acquisition

Age of Acquisition describes the average age that a certain word has been learnt. For example, the word "Dog" is acquired earlier (at an average age of 2.80 years) than the word "Orangutan" (at an average age of 7.83 years). Early acquired words tend to be more resilient in patients with semantic impairment (Brysbaert & Ellis, 2016). Patients with AD have been consistently observed to produce words with a lower average age of acquisition compared to controls (Cuetos et al., 2010). This effect can be observed when analysis is applied to fluency testing (see below) or spontaneous speech (Gayraud et al., 2012).

Forbes-McKay et al applied a number of variables, including age of acquisition, word length, typicality and frequency to the category fluency responses of 96 patients with AD and 40 controls. Discriminant function analysis showed that age of acquisition was the best predictor of group membership, accurately classifying 95% of the healthy controls and 88% of the AD patients (Forbes-McKay et al., 2005).

The discriminative value of age of acquisition was examined further by Sailor et al. Responses to verbal fluency tests performed by AD patients (n=34) and healthy controls (n=36) were assigned age of acquisition scores. The between group differences for age of acquisition values were compared between category fluency and phonemic fluency results. The mean age of acquisition was found to be

significantly lower between the AD and control groups when derived from the category fluency test (Sailor et al., 2011).

Typicality and Frequency

Typicality describes how closely a word is felt to relate to a given category. Words are ranked from 1-7 depending on how closely they are felt to relate to the category in question. This may be dependent on age so normative data scores are divided into age categories. For example, when administering the animal category fluency test, "Conger Eel" has lower typicality (1.39) than a "Cat" (6.7).

Frequency describes the frequency that a given word appears in a certain language. Each word is assigned a value on the Zipf-scale between 1-7. This is a logarithmic scale with higher values reflecting greater word frequency in the given language. For example, the word "Dog" (Zipf: 5.17) appears more frequently than the word "Ocelot" (Zipf: 2.52) in the UK, in English.

Patients with MCI and Alzheimer's dementia produce words with a higher typicality and frequency when compared to healthy controls (Forbes-McKay et al., 2005; Vita et al., 2014)

Affective Values

Words may be assigned values relating to their affective connotations by applying the VAD dimensions (Mohammad, 2018). Valence refers to how pleasant the word is (or at least the semantic concept represented, this is not a phonological quality), arousal refers to the intensity of the emotional response to a word and dominance refers to the sense of control associated with the word (Sutton et al., 2019).

Valence has been demonstrated to distinguish control groups and AD groups. Paek et al compared the performance of 17 patients with AD and 15 controls on a number of psychological assessments. Valence scores were calculated where possible. Patients with AD were found to produce words with a statistically significant valence value compared to the control groups. Interestingly, the valence score in the AD groups correlated negatively with performance in other domains, suggesting that as patients with AD deteriorate their language is characterized by increasingly pleasant output (Paek, 2021a).

Serial Recall Order

The serial recall order describes the correlation coefficient calculated between the order in which a word was produced and value of an assigned feature over a fluency

test. For example, age of acquisition for serial recall order would represent the trend observed in age of acquisition over the course of the task. A positive coefficient would suggest increasing age of acquisition over time (potentially reflecting integrity of semantic stores) whilst a negative coefficient would suggest declining age of acquisition over time. With the potential to explore semantic function, it is posited that this feature may have discriminant value between controls and AD (De Marco & Venneri, 2021).

1.4 Summary

In the sections above I have aimed to provide a context for this project. Alzheimer's disease is common and exerts a devastating toll on societal and personal levels. Advances in biomarkers enable an ever more accurate and early diagnosis. At present, however, these tests are inaccessible to most. With the prospect of disease modifying treatments in the foreseeable future, it is vital that we do not become complacent about the need for accurate clinical phenotyping. Scalable, accessible and minimally invasive screening tools will be essential to identify those patients most at risk of developing AD. Such tools will enable a refined cohort of high-risk patients to undergo more invasive, expensive and (admittedly) accurate tests, ensuring memory services direct their resources optimally.

Alzheimer's disease has a consistent and recognizable linguistic profile. This is characterized by early degradation of semantic performance, which reflects the involvement of the transentorhinal cortex at the earliest stage of the disease. There are a wealth of language-based features that have demonstrated powerful predictive value in identifying AD in the prodromal phase. I suggest that incorporating such features into a screening tool may improve diagnostic accuracy.

Throughout the following chapters I will describe my contributions toward the development of an automated, language-based cognitive assessment system and establish the feasibility of applying a more detailed analysis of language to the data collected. I will go on to present work assessing the predictive accuracy of a range of semantic features, extracted from data collected by an automated, virtual clinician. I will compare the properties of these features between a well described cohort of patients with MCI and a control group. The predictive value of these semantic feature will be evaluated and compared with established structural imaging features and cognitive tests. I will explore whether the addition of semantic features to a decision-making model yields improved accuracy, thus informing further work.

I will answer the following questions:

1. Is an unsupervised, remotely administered cognitive assessment system capable of obtaining interpretable data suitable for further analysis?

2. Do candidate linguistic features' relationship to one another and selected regions of interest on structural imaging support their use as measures of semantic memory and processes?

3. Does the analysis of semantic function applied to participants' responses to standardized interview questions (administered by an automated system) resemble the analysis of the verbal fluency tasks?

4. Does the semantic content of a participant's speech, as collected using automated verbal fluency tests, enable accurate classification between Healthy Controls and patients with MCI?

5. Is the discriminant value of these semantic features comparable to established structural imaging features?

6. Do the language features, cognitive test scores and imaging features combined contribute to greater diagnostic accuracy than the information typically available after an initial appointment?

by doing so I will endeavor to address the following hypotheses:

1. Unsupervised and remote interaction with a virtual clinical agent will permit the administration of verbal fluency tasks and provide data suitable for further analysis.

2. Semantic features extracted from speech will contribute towards the identification of prodromal Alzheimer's disease.

2 General Methodology

2.1 Introduction to General Methodology

The following section will provide a detailed description of the methodology common to both of the experiments presented in later sections. Those elements of the methods unique to each experiment will be covered in the relevant section.

2.2 Ethical Approval

The study was approved by the [NRES Committee Southwest Central Bristol] (REC reference: 16/LO/0737). Participants were screened for capacity to consent and, where appropriate, assent was obtained from a consultee under the Mental Capacity Act (2005).

2.3 Participants

The study design is a prospective, cross-sectional exploratory study directed towards the investigation of the diagnostic value of language and imaging features in patients with prodromal Alzheimer's disease.

All patients with mild cognitive impairment were recruited from the Sheffield Teaching Hospitals specialist memory service over a four-year period, between June 2018 and June 2022. Due to recruitment restrictions imposed by the COVID 19 pandemic an amendment was made to the ethical permission to allow for recruitment over the telephone as the majority of consultations were converted to remote for that period.

Participant demographics will be provided for the relevant experiments in later sections.

2.4 Recruitment Procedure

Eligible patient group members were approached to explore their interest in participating in research when they attended the Sheffield Teaching Hospitals Specialist Memory clinic for assessment or review. If an interest was expressed, they were provided with an information sheet and provided time to consider whether they were interested in participating. A follow-up telephone call was arranged to confirm whether they would like to proceed. Informed consent was sought from all participants on a follow-up meeting. As this study recruited patients in the prodromal/ mild cognitive impairment stage it was felt that capacity to consent was unlikely to be an issue.

Healthy controls were recruited either through the University of Third Age, affiliated with the University of Sheffield, or by approaching the spouse or relative of a patient participant.

Copies of the consent forms and information leaflets can be reviewed in Appendix A.

2.5 Inclusion and Exclusion Criteria

Inclusion criteria for the MCI group were:

- Clinical diagnosis of MCI/ Prodromal AD made by a consultant neurologist with a specialist interest in cognition, following a detailed clinical assessment, the administration of a cognitive assessment, standard blood tests and structural imaging. A proportion of the patients underwent 99m Tc-HMPAO single photon emission computer tomography, cerebrospinal fluid AD biomarker testing and/or detailed neuropsychological assessment. The diagnosis of MCI was made in accordance with the National Institute on Aging and Alzheimer's Association (NIA-AA) clinical criteria (ATN classification not considered practical owing to the limited access to biomarkers, particularly early on in the recruitment period). Thus, there was a mandatory requirement for objective cognitive impairment in the absence of meaningful functional impairment. Efforts were made to recruit patients with an amnestic presentation based on history and cognitive assessment.

- No history of any conditions that might impair cognition or language function. This included a history of stroke, epilepsy or significant mental-health concerns.

- Aged 55 or older

- Capable of providing informed consent.

- English as a first language

- Access to, and ability to operate (with assistance if needed) an internet-enabled computer to permit remote assessment with the virtual clinician.

Exclusion criteria for the MCI group included:

- A diagnosis other than mild cognitive impairment felt to be due to Alzheimer's disease.

- A history of any additional condition that resulting in a cognitive or communication impairment.

- Aged 54 years or younger

- Inability to provide informed consent.

- English not as first language

- No access to or ability to operate an internet-enabled computer.

Inclusion Criteria for the control group included:

- No diagnosis of a condition known to impair cognition or communication.

- Aged 55 or older

- English as a first language

- Access to an internet enabled computer to permit assessment with virtual clinician.

Exclusion Criteria for the control group included:

- A diagnosis of any condition known to impair cognition or communication.

- Aged 54 or younger

- English not as first language

- No access to an internet-enabled computer

2.5 Study Assessment Process

The design of the study was conceived to complement standard clinical care and reduce the need for extraneous travel and investigation. On the same visit , informed consent was collected as many of the study and clinical procedure were arranged as possible to minimize disruption.

2.5.1 Cognitive Assessment

If this had been performed within the last few weeks (during the standard clinical appointment) then ethical approval permitted the use of this test for research purposes. The majority of the MCI patients underwent testing with the Addenbrooke's cognitive examination III (Hsieh et al., 2013), administered by either the patient's usual doctor or myself at the time of recruitment. During the COVID 19 period a number of patients were recruited remotely (with necessary amendments made to the ethical permissions). These patients did not undergo testing with the ACE though all had recently been tested with a mini-mental state examination (MMSE) (Folstein et al., 1975).

2.5.2 CognoSpeak Procedure

Those participants that were assessed in person were directed to sit in front of a front-facing laptop camera. If accompanied by another person, this individual was encouraged to provide only prompts and support if necessary. The participants were instructed verbally to attempt to answer the questions posed and to use the "enter" key to proceed to the next question. Audio data were recorded using a Tascam DR-40 recorder as well as the laptop inbuilt microphone. Video recordings of the participant were captured using the in-built front-facing camera of the laptop used to present the CognoSpeak. In early iterations, a separate camera was positioned for a side profile. This served to capture interactions with an accompanying person.

The CognoSpeak assessment process varied very slightly over the course of recruitment. Only elements that were preserved throughout every iteration have been included in the following sections.

The questions posed by the system to the participant are outlined in Table 1.

In early iterations of the design, the video files were modified using the Windows Movie Maker edge detection feature to anonymize the interaction. This process was later automated. Video analysis was not undertaken in this study.

An Automatic Speech Recognition system was used to transcribe the audio into a string of words. A divarication tool was then applied to provide annotation of which words were said by which speaker. The combined output of the ASR and divarication tool provided information on the content of the speech and the duration of the contributions of different speakers. Manual transcripts were also prepared.

The automated analysis of the CognoSpeak data is described in more detail in the published article.

Table 1 CognoSpeak Questions/ Tasks

Order	*Question/ Task*
1	Where have you come in from today and what are you hoping to find out?
2	Tell me what problems you have noticed with your memory recently?
3	Who is most worried about your memory; you or somebody else?
4	What did you do over last weekend? Giving as much detail as you can.
5	What has been in the news recently?
6	Tell me about the school you went to and how old you were when you left?
7	Tell me what you did when you left school; what jobs did you do?
8	Tell me about your last job; give as much detail as you can.
9	Who manages your finances; you or somebody else? Has this changed recently?
10	Please name as many words beginning with the letter P as you can. It can be any word except for the names of people or places. (60 second timer)
11	Please name as many animals as you can. (60 second timer)
12	Please describe this picture. Give as much detail as you can. (Picture displaying the “Cookie Theft Scene”)

2.6 Data Storage and Management

All study data, including neuroimaging files, behavioral measures, and derived feature sets, were securely stored and managed in accordance with institutional and data protection guidelines. Raw and processed data were stored on encrypted, access-controlled servers hosted by Sheffield Teaching Hospitals Foundation Trust and The University of Sheffield, with regular automated backups. Personally identifiable information was stored separately from research data and linked only via anonymized participant codes. Imaging data were managed using the IMPAX PACS system and XNAT and exported in DICOM and NIfTI format for analysis. Derived datasets were processed and analyzed using R, with scripts and outputs version-controlled using Git. All data handling procedures complied with the General Data Protection Regulation (GDPR) and local ethical approvals to ensure confidentiality and integrity of participant information.

3.1 Introduction to Chapter: A fully automated cognitive screening tool based on assessment of speech and language.

This chapter presents a published original research study for which I was the first author. It aims to address the hypothesis:

1. Unsupervised and remote interaction with a virtual clinical agent will permit the administration of verbal fluency tasks and provide data suitable for further analysis.

To support this hypothesis, I will present my work on the development and efficacy of an automated language based diagnostic classification system.

This research work represents the product of collaboration between member of the computer science, neuroscience and human communication faculties at the University of Sheffield. The project was originally inspired by conversational analysis approaches to diagnostic classification in cognitive disorders. This demonstrated the emergence of distinct conversational profiles between the patients with, and without neurodegenerative pathology (Jones et al., 2016). Further work employed a bespoke, conversation analysis inspired, diagnostic scoring aid which was developed by linguists to exploit the distinguishing features of the aforementioned conversational profiles. After a training process was complete, the scoring aid demonstrated a sensitivity of 86.7% for neuro-degenerative causes of memory impairment and a specificity of 100% (Reuber et al., 2018).

Mirheidari et al published integrated these conversation-analysis inspired features into a computer based, automated classification tool. A range of acoustic, syntactic, semantic and visual features were extracted from a series of recorded interviews between a memory clinician and patients with cognitive symptoms. Using manual transcripts of the interviews as inputs, the system achieved a 95% classification accuracy. This dropped to 79% when the manual transcripts were substituted for an automated speech recognition output (Mirheidari et al., 2016).

Work began on developing a completely automated version of the system, capable of engaging the participant sufficiently to emulate patient doctor interviews and extract the salient features necessary for classification. An “AVATAR” was produced. Consisting of a virtual clinician, posing questions and recording responses from a laptop, this Avatar proved effective (and compared to conversation analysis performed by specialist linguists, cost effective. The classification between neurodegenerative pathology and functional memory disorder remained high at 90%. The classifier was applied to both memory clinician- patient interview and AVATAR-patient interviews with comparable results (Mirheidari et al., 2019).

The paper presented in this chapter represents an early effort to advance the precision of the classification from “neurodegenerative” pathology to specific pathological entities.

Chapter 3.2 Published Research Article

The following chapter is based on a peer-reviewed articles that was published in the *Journal of Neurology, Neurosurgery and Neuropsychiatry* in November 2019.

A fully automated cognitive screening tool based on assessment of speech and language.

Ronan O'Malley[1&2], Bahman Mirheidari[3], Kirsty Harkness[2], Markus Reuber [1&2], Annalena Venneri[1], Traci Walker[4], Heidi Christensen[3] and Daniel Blackburn[1&2]

1. Department of Neuroscience, University of Sheffield, Sheffield, United Kingdom

2. Sheffield Teaching Hospitals NHS Foundation Trust, Department of Neurology.

3. Dept of Computer Science, University of Sheffield, Sheffield, United Kingdom

4. Human Communication Sciences, University of Sheffield, Sheffield, UK

The article is reproduced here in accordance with the publisher's policies. Formatting has been adapted to maintain consistency throughout the thesis. The study was conceived and led by the candidate who contributed towards the study design, data collection, analysis, and manuscript preparation.

3.2.1 Abstract

Introduction

The dramatic recent rise in referrals to specialist memory clinics has been associated with an increased proportion of patients referred with Functional Memory Disorder (FMD), i.e. non-progressive cognitive complaints. These referrals have exerted time and financial pressures on secondary care services, impairing their ability to deliver high-quality care for patients with neurodegenerative cognitive disorders. We have developed a fully automated system, "CognoSpeak", which enables risk stratification at the primary-secondary care interface and ongoing monitoring of patients with memory concerns.

Methods

We recruited 15 participants to each of four groups: Alzheimer's disease (AD), Mild Cognitive Impairment (MCI), FMD and healthy controls. Participants responded to 12 questions posed by a computer-presented talking head. Automatic analysis of the audio and speech data involved speaker segmentation, automatic speech recognition and machine learning classification.

Results

CognoSpeak could distinguish between participants in the AD or MCI groups and those in the FMD or healthy control groups with a sensitivity of 86.7%. Patients with MCI were identified with a sensitivity of 80%.

Discussion

Our fully automated system achieved levels of accuracy comparable to currently available, manually administered assessments. Greater accuracy should be achievable through further system training with a greater number of users, the inclusion of verbal fluency tasks and mood assessments. The current data supports CognoSpeak's promise as a screening and monitoring tool for patients with MCI. Pending confirmation of these findings, it may allow clinicians to offer patients at low risk of dementia earlier reassurance and relieve pressures on specialist memory services.

3.2.2 Introduction

Memory clinics assess patients with a variety of cognitive complaints disorders, including those related to Alzheimer's disease (AD), possible prodromal states (mild cognitive impairment, MCI) and those with Functional Cognitive Disorder (FCD, disabling but non-progressive cognitive complaints associated with emotional or psychological factors).

Early referral to an appropriate care pathway yields benefits for patient wellbeing and efficient resource allocation. Specialist memory clinics are a limited resource and patients with FCD can be successfully managed in other settings. Accurate pre-clinic stratification tools are required to direct patients towards the most appropriate service.

AD is associated with subtle impairments in language that may precede deficits in episodic memory by decades (1). Previous work has shown that qualitative analysis of conversational profiles inspired by the methodology of Conversation Analysis (CA) can discriminate between patients with FCD and those with neuro-degenerative (ND) conditions (2). However, this approach depends on highly trained experts and is not easily scalable.

Whilst the use of automated speech analysis has been explored previously to identify cognitive impairment, most studies do not describe fully automated solutions. Instead, they rely on the automated analysis of data collected from human-human interaction (3) or used manually generated transcripts (4).

We have created a fully automated stratification tool. "CognoSpeak" consists of a virtual clinician, a computer screen-presented talking head, which asks questions and records the patients' spoken responses. The system uses Automatic Speech Recognition (ASR) and divarication (segmenting the recording into contributions from different speakers) to extract acoustic and linguistic measures that are used by machine learning classifiers to select the most likely diagnostic category (5).

3.2.3 Methods

We recruited 60 participants; 15 each from four different diagnostic groups: AD, MCI, FCD and healthy controls (HC). HCs were recruited via the "University of the Third Age" and patient participants from a specialist memory clinic in Sheffield, between May 2016 and January 2019. Patients could be recorded on their own or in the presence of an accompanying person. Ethical permission was granted by the NRES Committee South West-Central Bristol (Rec number 16/ LO/0737) in May 2016. Neurological diagnoses were made according to standard diagnostic criteria[1] and after multidisciplinary team review. Presence of significant mood disturbance

[1] For AD, McKhann et al Alzheimer's & Dementia, 2011; for MCI Petersen et al Archives of Neurology 1999, Petersen Archives of Neurology 2001; for FCD Schmidtke et al American Journal of Geriatric Psychiatry, 2008

(ascertained through clinical history and Patient Health Questionnaire-9 (PHQ-9 > 15) and significant cerebrovascular disease resulted in exclusion. All underwent cognitive assessment using Addenbrooke's Cognitive Examination-Revision (ACE-R) tool or detailed neuropsychological evaluation. Brain imaging (including CT, MRI and Tc99m HMPAO single-photon emission computed tomography) was performed based on clinical need. A proportion of the healthy control group had MRI as part of their involvement in a previous study (VPH-DARE@IT http://www.vph-dare.eu/).

The CognoSpeak assessment process has been comprehensively described in prior papers (5). In brief, Participants were directed verbally to respond to the questions posed by the virtual clinician and to use the "enter" key to proceed from one question to the next. Audio data were recorded using a Tascam DR-40 recorder.

An ASR system was used to transcribe the audio into a string of words. A divarication tool was then applied to provide annotation of which words were said by which speaker. The combined output of ASR and divarication provided information on the content of the speech and the duration of the contributions of different speakers. The automatically transcribed text as well as the recorded audio were used to extract the range of features that the machine learning based classifier uses to assign participants to a diagnostic class.

72 features were extracted from the speech of patients and accompanying persons, including 17 CA -inspired features, 24 acoustic-only, 24 lexical-only, and 7-word vector features.

3.2.4 Results

Please see Table 2 for demographic variables, psychological measures performed and the results and a description of the imaging.

Table 2 Demographic, psychological and imaging data for four-way comparison

Characteristics	HC	FCD	MCI	AD
Gender (% male)	40	40	66.7	66.7
Age (Mean)	69.5 (± 4.0)	54.9 (± 4.1)	63.4 (± 4.2)	67.8 (± 4.2) †
Years of Education (Mean) *	18.1 (± 1.0)	16.4 (±0.6)	17.3 (±1.1)	18 (±1.6) ‡
Neuropsych Performed (%)	100	93.3	100	100
ACE (%)	100	85.7	86.7	93.3
MMSE (%)		14.3	13.3	6.7
ACE Score (Mean)	95.3 (± 1.6)	88.7 (± 5.9)	81.3 (± 3.6)	69.5 (± 8.2) *
MMSE Score (Mean)		27.3 (± 1.9)	27.0 (± 0.8)	23 (± 3.3)
Structural Imaging Performed (%)	40	86.8	100	100
Abnormal (%)	0	20¥	20	66.7
Consistent with ND (%)	0	0	13.3	26.7
SPECT Performed (%)	0	0	46.7	66.7
Consistent with ND (%)	0	0	85.7	70

† Age: HC vs FCD, HC vs MCI, FCD vs MCI, FCD vs AD p <0.05. HC vs AD and MCI vs AD p >0.05

‡ Years of Education: HC vs FCD and FCD vs AD p <0.05. HC vs MCI, HC vs AD, FCD vs MCI and MCI vs AD p <0.05 * ACE Score: HC vs MCI, HC vs AD, FCD vs MCI, FCD vs AD and MCI vs AD p< 0.05. HC vs FCD p >0.05. ¥ 3 Abnormal MRI scans demonstrating a previous contusional injury, generalized atrophy and mild small vessel disease.

ND = Neurodegeneration

The mean duration of interactions across all participant groups and CognoSpeak was 11 minutes 24 seconds. HCs spent an average of 9 minutes 58 seconds; FCD patients 11 minutes 2 seconds; MCI patients 9 minutes 34 seconds and AD patients 15 minutes 2 seconds. Only the difference between the MCI vs AD groups was significant (U=59, P=0.026).

Two-way automatic classification (AD & MCI v FMD & HC)

An effective cognitive stratification tool must be able to separate those with potential neurodegeneration from those without. In the two-way classification CognoSpeak system had an accuracy of identifying participants with MCI or AD of 87% (see Figure 1) whilst the accuracy of correctly allocating participants as either HC or FMD was 77%.

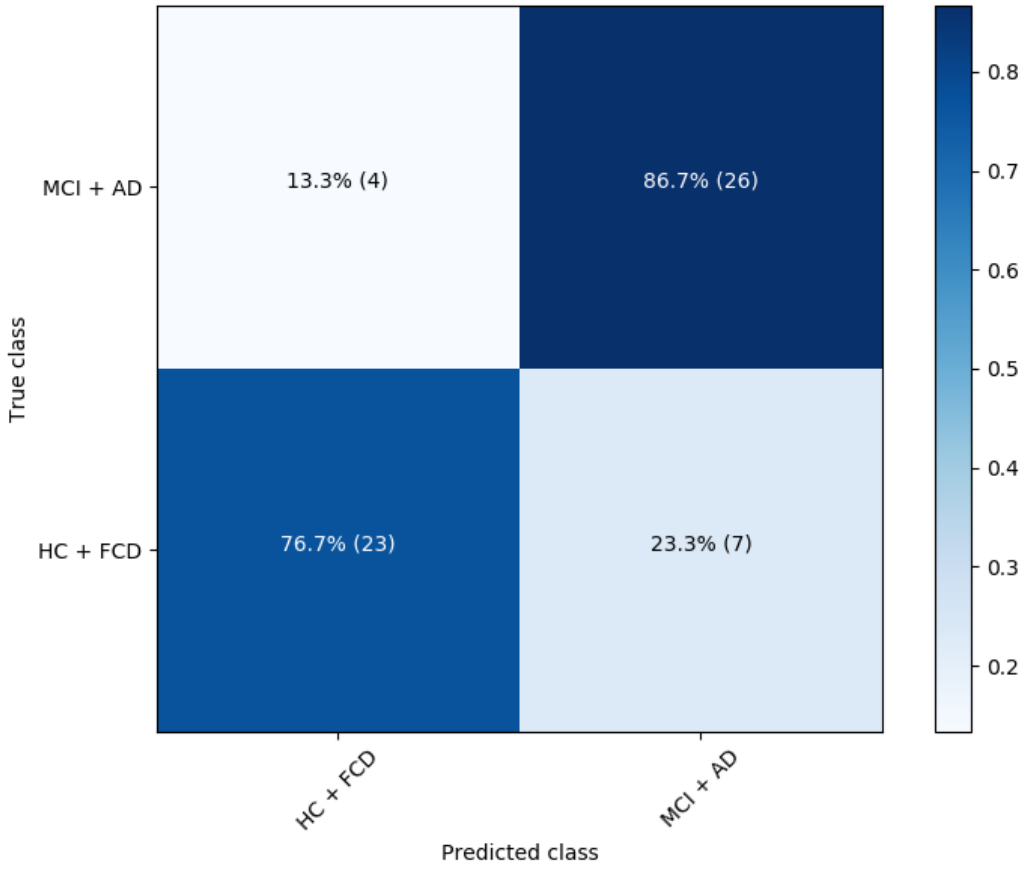


Figure 1 Confusion Matrix showing Two-Way Classification Accuracy (HC+FMD vs AD+MCI)

Three-way automatic classification (AD v MCI v FMD & HC)

Overall correct classification was achieved in 65% of cases. The accuracy of identifying participants with AD was 80% (see Figure 2). Two participants were incorrectly allocated as MCI and one as belonging to the FMD and HC group.

The identification accuracy of participants with MCI was 80%. One participant was incorrectly allocated as AD and two as FMD and HC. The accuracy of identifying the participants with either HC or FMD was 50%. Eight participants were incorrectly allocated as MCI and seven as AD.

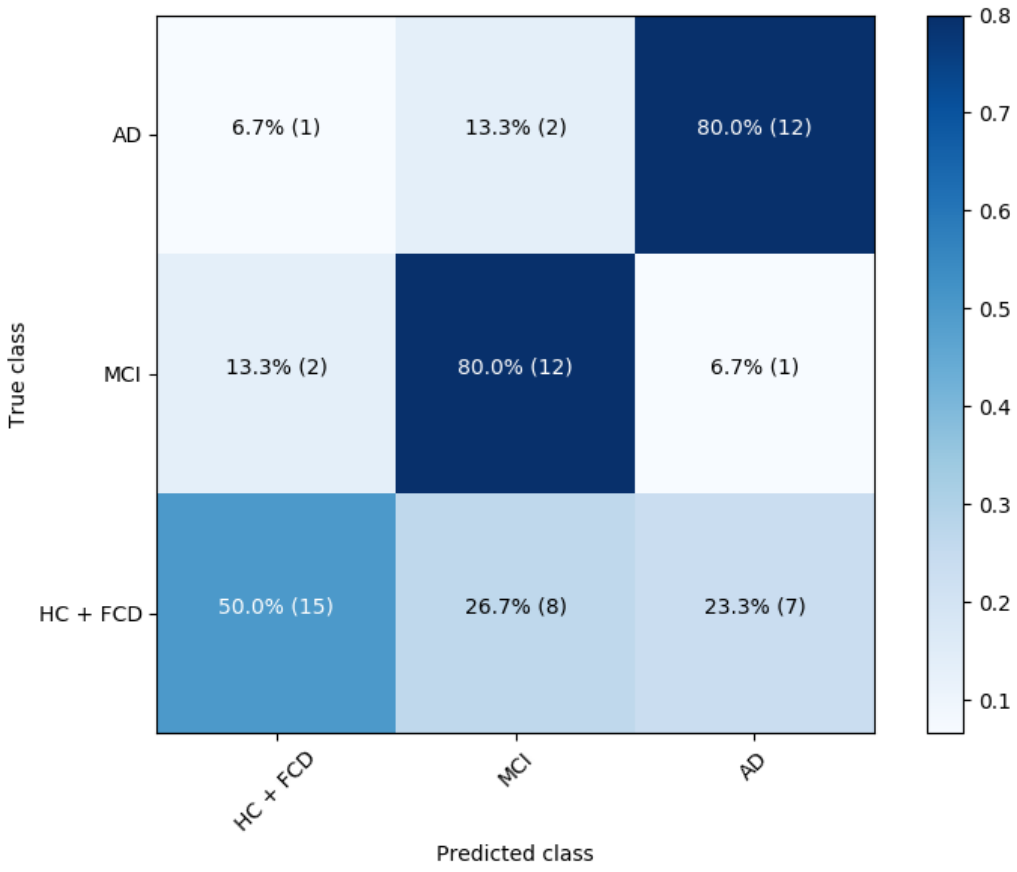


Figure 2 Confusion Matrix showing Three-Way Classification Accuracy (HC+FCS vs MCI vs AD)

Four-way automatic classification (AD v MCI v FMD v HC)

The more difficult task of identifying participants from all four groups revealed an overall classification accuracy of 60%. Accuracy in identifying AD participants was 80%; accuracy in identifying MCI participants was 60%; accuracy in identifying FCD participants was 47% and accuracy in identifying the HC participants was 53% (Figure 3), (of note the most frequent misclassifications occurred between the HC and FCD groups).

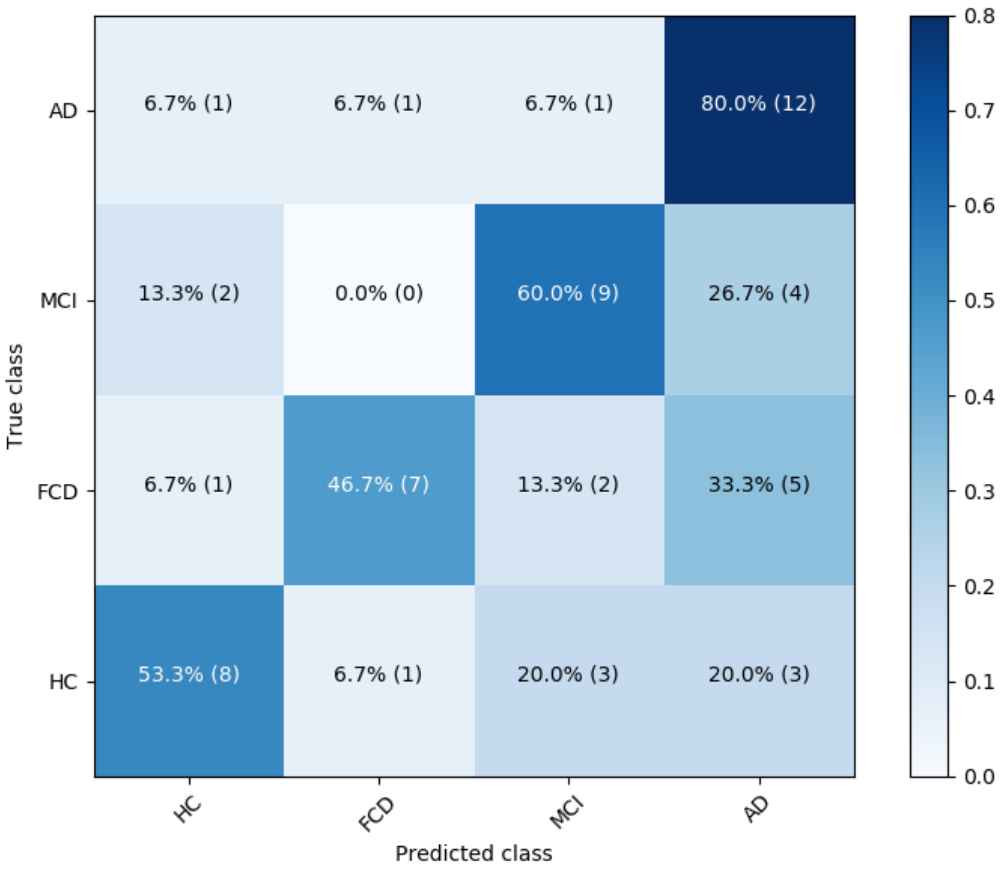


Figure 3 Confusion Matrix showing Four-Way Classification Accuracy (HC vs FCD vs MCI vs AD)

3.2.5 Discussion

The UK's National Dementia Strategy emphasizes the importance of early diagnosis and provision of support for patients with progressive cognitive disorders. Hence, our approach has prioritized sensitivity over specificity. The two-way classification system achieved a sensitivity of 86.7% for neurodegenerative memory disorders, comparing favorably with most commonly employed screening tools. The specificity was slightly lower at 76.7% but given the emphasis on early identification and the fact that "false positives" will be investigated by the specialist memory services, this

is acceptable. Using a three-way classification MCI, AD & FCD plus HC) CognoSpeak was able to identify participants with MCI with a sensitivity of 80.0% (95% CI: 51.9-95.7) and a specificity of 77.8% (95% CI: 62.9-88.8), exceeding sensitivity (66.34%) and specificity (72.94%) of the MMSE in discerning between MCI and HCs (6).

Accurate identification of patients with MCI may allow early intervention and facilitate participation in research studies. We are unaware of any other screening tool that has included patients with FCD in their development. However, the inclusion of this patient group is essential in validation studies of screening or stratification tools as those with FCD make up 24% of referrals to specialist memory services (7). The ability of our tool to distinguish between FCD and MCI / AD rather than only between HC and MCI/AD groups increases its ecological validity. The CognoSpeak 2-way classification system can distinguish between patients with MCI / AD and those without neurodegenerative pathology (HC & FCD) with a sensitivity of 86.7%.

Typical pre-clinic cognitive screening tools take between 5-10 minutes and require clinician time (8, 9). CognoSpeak compares favorably to this. A large 2016 systematic review described mean primary care physician consultation lengths across 67 countries. The group reported that 18 countries, accounting for approximately 50% of the world's population, had mean GP consultation lengths of 5 minutes or less (10). This suggests that even the most concise, traditional cognitive screening tool may not be applicable to a significant proportion of the world's population. CognoSpeak has the advantage of being automated, capable of remote administration via a tablet in a primary care office or at home using a secure website, without the necessity for direct clinician oversight. It has been estimated that in 2019, 58.8% of the global population have access to the internet (11). Results from at-home screening could be transferred and analyzed by local secondary care memory services, thus reducing primary care time. By identifying FCD this tool could be used to provide early reassurance and reduce the need for unnecessary and stressful memory clinic consultations. Furthermore, the results/reporting from CognoSpeak could triage those most at risk, plan scans before consultation and reduce the time needed for initial testing in specialist clinics. This may be especially important due to the accelerated use of telephone and virtual clinics during the Covid-19 pandemic. Furthermore, the capacity of CognoSpeak to distinguish between patients with AD and MCI raises the possibility that this automated tool could be used for longitudinal monitoring of patients with MCI.

We acknowledge a number of limitations to this study. The relatively small sample size was limited by the time taken to recruit well-described participants in each cohort. A larger cohort would provide more accurate information into the relative sensitivity and specificity of CognoSpeak. This analysis used the first 9 of 12 questions posed. The remaining three questions, verbal fluency tasks and a picture description tasks, are extensively used in clinical practice, with good accuracy in detecting MCI. These will likely contribute to increasing accuracy of CognoSpeak.

They have not been included in the current study as novel automated approaches to their analysis are still under development. Inherent to the iterative nature of the machine learning process, we anticipate attaining greater accuracy with access to larger sample sizes. This applies both to the accuracy of the automated speech recognition and the refinement of classifiers.

For this initial validation study, we have limited recruitment to patients belonging to the diagnostic groups most commonly represented in specialist memory clinics. Future studies will also include patient groups with non-AD dementias.

In conclusion, CognoSpeak is a fully automated cognitive screening system that can discern between normal cognition and neurodegenerative memory disorders with sensitivity comparable to traditional screening methods.

Funding

This work was supported in part by MRC CiC grant.

Acknowledgments

This is a summary of independent research carried out at the NIHR Sheffield Biomedical Research Centre (Translational Neuroscience). The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health.

We would like to thank all participants and care partners who took part in this study. We would like to thank all participating clinicians working in the memory clinic (Aijaz Khan). We would also like to thank students from the clinical neurology master's course (Imke Meyer, Casey Rutten, Thomas Swainston, Eric Brook and Lee-Anne Morris) and medical students (Isobel Sonksen, Alice Lewis) from the University of Sheffield for their contributions to data collection.

References

1. Amieva H, Le Goff M, Millet X, Orgogozo JM, Peres K, Barberger-Gateau P, et al. Prodromal Alzheimer's disease: successive emergence of the clinical symptoms. Ann Neurol. 2008;64(5):492-8.
2. Reuber M, Blackburn DJ, Elsey C, Wakefield S, Ardern KA, Harkness K, et al. An Interactional Profile to Assist the Differential Diagnosis of Neurodegenerative and Functional Memory Disorders. Alzheimer Dis Assoc Disord. 2018.
3. Konig A, Satt A, Sorin A, Hoory R, Toledo-Ronan O, Derreumaux A, et al. Automatic speech analysis for the assessment of patients with predementia and

Alzheimer's disease. Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring. 2015;1:112-24.
4. Tanaka H, Adachi H, Ukita N, Ikeda M, Kazui H, Kudo T, et al. Detecting Dementia Through Interactive Computer Avatars. IEEE J Transl Eng Health Med. 2017;5:2200111.
5. Al-Hameed S, Benaissa M, Christensen H, Mirheidari B, Blackburn D, Reuber M. A new diagnostic approach for the identification of patients with neurodegenerative cognitive complaints. PLOS ONE. 2019;14(5):e0217388.
6. Ciesielska N, Sokolowski R, Mazur E, Podhorecka M, Polak-Szabela A, Kedziora-Kornatowska K. Is the Montreal Cognitive Assessment (MoCA) test better suited than the Mini-Mental State Examination (MMSE) in mild cognitive impairment (MCI) detection among people aged over 60? Meta-analysis. Psychiatr Pol. 2016;50(5):1039-52.
7. McWhirter L, Ritchie C, Stone J, Carson A. Functional cognitive disorders: a systematic review. Lancet Psych. 2020;7(2): 191-207.
8. Lin JS, O'Connor E, Rossom RC, Perdue LA, Eckstrom E. Screening for cognitive impairment in older adults: A systematic review for the U.S. Preventive Services Task Force. Annals of internal medicine. 2013;159(9):601-12.

9. Ballard C B BA, Corbett A, Rasmussen J. Helping you to assess cognition: A practical toolkit for clinicians. In: Health Do, editor. London2015.
10. Irving G, Neves AL, Dambha-Miller H, Oishi A, Tagashira H, Verho A, et al. International variations in primary care physician consultation time: a systematic review of 67 countries. BMJ Open. 2017;7(10):e017902.
11. Stats IW. Internet World Stats 2019 [Available from: https://www.internetworldstats.com/stats.htm.

3.2.6 Conclusion

CognoSpeak is a non-invasive, automated assessment system entirely capable of remote administration. It can be administered over a short period of time and discriminated between patients with and without neurodegenerative disease with reasonable accuracy.

The reported accuracy of the most recent iteration of the CognoSpeak automatic speech recognition system is 76.8% (Mirheidari et al., 2024). The classification system performs well despite this, potentially due to the influence of acoustic and temporal features that are insensitive to high word error rates.

Whilst those features may be robust to low ASR accuracy, it is likely that this would impact negatively on the predictive value of semantic features.

Pakhomov et al achieved verbal fluency score results through completely automated methods thar were comparable to the scores of manually administered tasks. This was done through a combination of constraining the language model to a predefined lexicon (animals for animal category testing), confidence scoring and speaker adaptation (Pakhomov et al., 2015).

The current leaders in automatic speech recognition performance are OpenAI team with Open AI Whisper large-V3. This is reported to have achieved word error rates as low as 2.7% (which actually outperforms human transcribers) (Radford et al., n.d.). The performance would be unlikely to remain as high in real world situations.

CognoSpeak employs a number of features in its classification system. Owing to the nature of machine learning approach used for classification, the factors that inform classification are not easily retrievable. This runs the risk of reducing confidence in the output. Additionally, in cases where the classification is found to be incorrect it is not possible to review the output to identify a either the source of the error or an alternative explanation for the features observed.

With regards to my hypothesis:

1. Unsupervised and remote interaction with a virtual clinical agent will permit the administration of verbal fluency tasks and provide data suitable for further analysis.

Yes, patient find the system easy to engage with and it is capable of recording patients responses to a series of interview style questions. Verbal fluency tasks were administered and processed in a fashion that permitted further assessment. Though the high word error rate may preclude a completely automated approach at present, adjustments such as those applied by Pakhomov could be explored.

4. Advanced Semantic Analysis

4.1. Introduction

Over the next section I will present the methodology and results for the main set of experiments in this thesis.

This section is aimed towards addressing the hypothesis:

1. Semantic features extracted from speech will contribute towards the identification of prodromal Alzheimer's disease.

To do so I will describe in detail the methods used and the demonstrate the results comprehensively. These will be considered as they are presented and discussed further in a later section.

All of the language data used in the following section was gathered using CognoSpeak though manual transcriptions were used rather than the ASR.

4.2 Advanced Feature Analysis for MCI Detection: Methods

This chapter will present detailed methodology for the advanced feature analysis. In the interests of brevity, aspects of the methodology general to both experiments such as inclusion and exclusion criteria will not be repeated.

4.2.1 Longitudinal Review of the MCI Cohort

Table 3 presents the demographic information and cognitive assessment scores for the study participants. Healthy controls (HC; n = 50) and patients with mild cognitive impairment (MCI; n = 50) differed significantly in age and sex distribution. The MCI group was significantly younger than controls (median 66.5 vs 72 years, $p < 0.05$) and had a lower proportion of males (42% vs 66%, $p < 0.05$). Years of education did not differ significantly between groups. Compared with healthy controls, the MCI group demonstrated significantly poorer global cognitive performance, with lower scores on both the ACE total and MMSE (both $p < 0.05$). Examination of ACE subdomains showed significantly reduced performance in attention, memory, fluency, and language in the MCI group relative to controls (all $p < 0.05$), while visuospatial performance did not differ significantly between groups. ACE total scores were available for all healthy controls and 34 individuals with MCI, while ACE sub-scores were available for all healthy controls and 26 individuals with MCI.

Table 3 Demographic Information and Cognitive Assessment scores of Controls, MCI and ADMCI groups

	Healthy Controls (n = 50)	MCI (n = 50)	p(HC vs MCI)	ADMCI (n = 31	p(HC vs ADMCI)
Age	72 [67.25-76]	66.5 [61-72.75]	p = <0.05	65 [57.5-70.5]	p = < 0.05
Sex (%male)	66	42	p= < 0.05	58	p = > 0.05
Years of Education	18 [16-19.75]	16 [16-18]	p = >0.05	16 [16-19.5]	p = > 0.05
ACE Total	96 [93.25-98]	82.5 [75.25-85]	p = <0.05	83 [77.5-87]	p = < 0.05
ACE Attention	18	17 [16.25-18]	p = <0.05	17 [16.5-18]	p = < 0.05

ACE Memory	24 [23-25]	15.5 [13-18]	p = <0.05	15.5 [13.75-17.25]	p = < 0.05
ACE Fluency	12 [13-14]	9 [8-11]	p = <0.05	10 [8.75-12]	p = < 0.05
ACE Language	25.5 [25-26]	24.5 [23-25]	p = <0.05	24.5 [23-25.25]	p = < 0.05
ACE Visuospatial	16 [15-16]	15 [15-16]	p = >0.05	16 [15-16]	p = > 0.05
MMSE	29.5 [28.9-29.9]	27 [25.4-27.3]	p = <0.05	27 [26-28.1]	p = < 0.05

Between group comparisons: Sex- Chi-Squared, rest of variables Mann-Whitney U test. ACE total available for all HC and 34 MCI. Sub-scores available for all HC and 26 MCI. ACE total available for 23 ADMCI, sub scores available for 16

Due to unforeseen circumstances my PhD project was attenuated by a significant leave of absence. Whilst this presented significant challenges, it also permitted me to review the progress of the MCI cohort. The purpose of this was to assess whether any of the participants had gone on to have AD biomarker testing, progressed to dementia or developed a phenotype observably distinct from AD.

I undertook a comprehensive review of all of the patients clinical notes, results and investigations (with the requisite ethical approvals). Though the results of this review are more comprehensively described in the discussion section of this chapter, interpretation of the results will rely on some explanation. Of the 50 patients with MCI presumed to be due to AD (as defined by NIA-AA clinical criteria following an expert clinical assessment, cognitive testing and imaging), 31 had had the diagnosis confirmed. Twelve of these patients had undertaken CSF biomarker analysis, which confirmed AD in every case.

Of the remaining 19 patients, 10 were still described as MCI. They had not progressed to meet the clinical criteria, undergone testing to confirm/ refute the diagnosis on biological ground or developed a clinical presentation no longer felt to be consistent with AD. These patients retained the label MCI, presumed to be due to AD.

Of the other 9 patients, 3 were later diagnosed with Lewy Body Disease, 2 with Frontotemporal Dementia, 1 with corticobasal degeneration syndrome, two with progressive supranuclear palsy and one with Parkinson's Disease Dementia (the providence of this diagnosis is uncertain given that pre-existing Parkinson's disease is an obligatory feature (Poewe et al., 2008) and there is neither mention of this in the clinical notes or evidence of motor features in anonymized video or audio recordings).

To explore whether the ADMCI group behaved differently from the MCI group, a nested cohort was created. These were matched with 31 healthy controls based on age, sex and years of education using the pairmatch command of the OptMatch package in R. All of the statistical tests presented in the next section were applied to both groups. The details of these groups can be reviewed in Appendix B. The impact of this is considered in the discussion section.

Table 3 demonstrates how the ADMCI and healthy control groups compare. Healthy controls (HC; $n = 50$) and ADMCI patients (ADMCI; $n = 31$) differed significantly in age but not sex distribution. The ADMCI group was significantly younger than controls (median 65 vs 72 years, $p < 0.05$), while the proportion of males did not differ significantly between groups. Years of education were comparable between healthy controls and the ADMCI group. As expected, individuals with ADMCI showed significantly lower global cognitive scores than controls, with reduced performance on the ACE total and MMSE (both $p < 0.05$). Across ACE subdomains, the ADMCI group demonstrated significantly poorer performance in attention, memory, fluency, and language compared with healthy controls (all $p < 0.05$), with no significant between-group difference observed for visuospatial scores. ACE total scores were available for all healthy controls and 23 individuals with ADMCI, while ACE sub-scores were available for all healthy controls and 16 individuals with ADMCI.

4.2.2 Structural MRI Image Acquisition

A “Memory Protocol MRI” was arranged for each patient with MCI. This is a standard protocol employed in Sheffield Teaching Hospitals for the assessment of patients presenting with memory problems. It consists of anatomical acquisitions obtained with a Philips Ingenia 3T scanner. Only one of the acquired MRI sequences was used in this study: the three-dimensional T1-weighted image (voxel size: 0.94 mm × 0.94 mm × 1.00 mm; TR: 8.2 s; TE: 3.8 s; FOV: 256 mm; matrix size: 256 × 256 × 170).

The healthy controls that were recruited to this study did not undergo detailed structural imaging as we did not acquire necessary ethical permissions. It was deemed reasonable to assemble a group of healthy controls from the VPHDare Platform (Westman et al., 2012) matched for age, sex and years of education. The details of the matched healthy participants are available in the Appendix C. Matching was carried out using the “pairmatch” function (in an r environment using the “optmatch” package). This uses a propensity score approach.

4.2.2.1 Image Transfer

The raw imaging was transferred from the Sheffield Teaching Hospitals IMPAX (Agfa HealthCare, NV) system using XNAT (eXtensible Neuroimaging Archive Toolkit) (Marcus et al., 2007). This enabled me to export and anonymize the DICOM images

before transferring them to a secure, password protected server in the university. The images were labelled with the appropriate study code at the time of transfer.

4.2.2.2 Image Pre-processing

The processing and modelling pipeline was run using Matlab R2014a (Mathworks Inc., UK) and Statistical Parametric Mapping (SPM) 12 software (Wellcome Trust Centre for Neuroimaging, London, UK). Using the graphical user interface, the T1-weighted images underwent segmentation to compute native space maps of grey matter, white matter, and cerebrospinal fluid.

Volumetric quantifications of each tissue class were then carried out using the "get_totals" script (http://www0.cs.ucl.ac.uk/staff/g.ridgway/vbm/get_totals.m).

Total intracranial volume was calculated as the sum of the three tissue classes. Grey matter and white matter fractions were then computed as a ratio between each tissue volume and the total intracranial volume.

The native space grey matter maps were then spatially normalized to the Montreal Neurological Institute (MNI) template. To render the maps amenable to parametric testing they then undergo spatial smoothing, applying a standard 8 mm full width at half maximum (FWHM) Gaussian kernel.

4.2.2.3 Voxel Based Morphometry

Once the images are pre-processed, the parametric tests are put together in SPM. Contrasts were constructed to compare controls > MCI. Covariates included age, sex, years of education and total intracranial volume. In this study I ran a series of two sample t-tests. A voxel wide threshold of < 0.001 was applied and cluster-level significance was assessed using family-wise error (FWE) correction, with a threshold of $p < 0.05$ (FWE-corrected) to control for multiple comparisons across the whole brain. In some instances, a cluster level-threshold of $p<0.001$ (FWE-corrected) was applied to explore which differences survived the more stringent significance threshold. Minimum cluster extent was set at $k >20$ voxels (Whitwell, 2009).

For publication-quality visualization, thresholded t-maps were exported and rendered using MRIcroGL (Li et al., 2016) an open-source tool for neuroimaging display. Maximum intensity projections ("glass brains") were used to display significant clusters, and slice overlays were adjusted to best illustrate the spatial distribution of effects. Images were formatted for publication and figure composition, with color maps representing voxel-wise t-statistics and anatomical templates in standard MNI space.

Tables demonstrating coordinated in MNI space were exported from Matlab. The format was cleaned in R. The MNI template coordinates were manually extracted

and entered into the BioImage Suite Web application to retrieve anatomical labels. Tables for publication where then exported from R.

4.2.2.4 Automated Hippocampal Segmentation

The anonymized raw-T1 images were uploaded to the University College London Centre for Medical Image Computing "NiftyWeb BRAIN-STEPS" platform. This is a free to use online region of interest extraction tool that employs the STEPS (Similarity and Truth Estimation for Propagated Segmentations) protocol, an automated, machine-learning segmentation tool-kit with excellent reported segmentation accuracy (mean Dice score 0.925) (Jorge Cardoso et al., 2013; Prados Carrasco et al., 2016).

The segmented image are returned via email. The hippocampal volumes were calculated using the "get_totals" script. To standardize these values, they were generally expressed as a ratio of total intracranial volume.

4.2.3 Extraction of Participant Language Data

For this study the only output from the CognoSpeak assessment used was the manual transcript of the interview. This was anonymized and converted into a text file, removing additional formatting.

4.2.3.1 Verbal Fluency Tasks

The participant responses to questions 10 and 11 were extracted from the transcript. The responses were cleaned, removing any interjections from other parties. The cleaned text was then imported to R and tokenized to create a long data frame for each participant. The tasks were then scored according to convention (Hughes, 1970) with errors and repeats noted in parenthesis. To generate verbal fluency scores a count was performed of the errors and repeats and correct responses. These values were then summarized per participant using means.

For the rest of the analysis the error classification was ignored as it was felt that even the erroneous responses may be informative about underpinning linguistic processes. The repeats were removed from the analysis.

4.2.3.2 Whole Interview

The entire interview, including question 10 and 11 underwent cleaning in R. This involved removing all of the interviewer and accompanying persons input and substituting the questions for markers. The corpus of text was then segmented into questions using that label. The corpus was tokenized using the tidytext package and

common stopwords were removed using the tm package (with variables constructed to reflect total words/ question and total filler words/ question).

4.2.4 Language Based Features

4.2.4.1 Lexical Dictionary

A “lexical dictionary” was compiled using reference data. The contents of the dictionary are summarized in Table 4.

Table 4 Contents of Lexical Dictionary including reference source

Feature	*Reference Data Source*
Age of acquisition	(Brysbaert & Biemiller, 2017)
Typicality	(Biundo, 2010)
Frequency	(van Heuven et al., 2014)
Prevalence	(Brysbaert et al., 2019)
Recognition Time	(Mandera et al., 2020)
Dominance	(Warriner et al., 2013)
Valence	(Warriner et al., 2013)
Arousal	(Warriner et al., 2013)
Body-Object Interaction	(Pexman et al., 2019)
Concreteness	(Muraki et al., 2023)

These reference data frames were joined using whole_join in R, building a comprehensive though sparse lexical dictionary.

This was then cross-referenced with both the manually processed verbal fluency task as well as the automatically processed whole interview. Summary statistics were generated using means values.

4.2.4.2 Cluster Features

The cluster features were applied only to the category fluency task. for animal fluency were informed by the landmark work by Troyer et al.(Troyer, Moscovitch, Winocur, Leach, et al., 1998a)

Several techniques were trialed, inspired by various groups methods in automating the analysis (Hähnel et al., 2023; Kim et al., 2019; Pakhomov & Hemmy, 2014). Pakhomov et al employed an automated approach relying on Latent Semantic

Analysis (LSA) for determining the strength of semantic relatedness between word pairs. Applying these techniques, they confirmed that an increase in mean cluster size was significantly associated with a reduction in the risk of developing dementia.

I explored using a GloVe (Global Vectors for Word representation) and BERT (Bidirectional Encoder Representations from Transformers) powered models. These are both language models with some reported success in automated clustering (Alacam et al., 2022; Bushnell et al., 2022).

I also trialed an automated method for animal category fluency using the Troyer categories (available in the 1998 publication). I prepared a clean, readable version and imported it in an R environment. Code was written to assign each word in the participant data one or more Troyer inspired labels (for example: Cat – Pet – Feline). The resulting data was then grouped by subject ID and cluster metrics were calculated by sequentially assessing the assigned Troyer category. If a word shared an assigned term with the preceding category, they received that label and were considered to belong to the same cluster. If the following word did not share a Troyer category, then a new cluster was started.

Of all the automated/ semi-automated methods trialed this was by far the most effective. I suspect that this represents the limits of my coding abilities rather than the theoretical limits of the techniques (as indeed, other groups have demonstrated). The ease and apparent efficacy of this system comes at the expense of generalizability. Systems based on powerful word models do not require the kind of explicit instruction that I have provided above and can be applied to any version of the category fluency test.

See Figure 4 for a comparison of the results of the automated and manual methods, reflecting very strong correlation. To ensure the highest standards of accuracy for the ensuing exploratory work, the manual features were used.

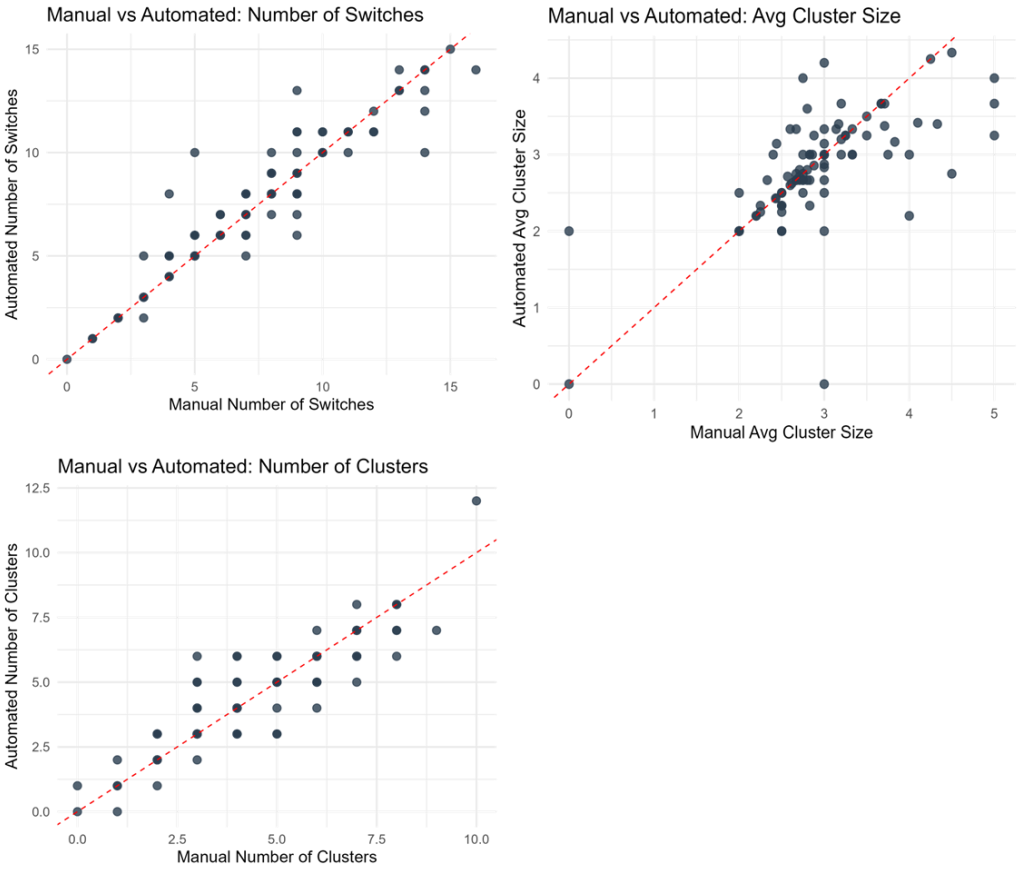


Figure 4 Scatter plots demonstrating the results of manual and automated Troyer cluster analysis

4.2.4.3 Serial Recall Order

Inspired by the work of DeMarco et al serial recall order was calculated for each feature in the lexical dictionary. The participant responses to both category and phonemic fluency tests were assigned an ordinal value. Spearman's Rho correlation coefficients were calculated between word order and lexical value. The direction of the coefficient expresses the trend of the lexical values over the course of the task. For examples, a positive correlation coefficient in age of acquisition serial word order would suggest that the participants responses became increasingly complex as time went on. Combining lexical semantic properties like age of acquisition (semantic processing measures) with recall order (memory retrieval) is thought to provide a more clear insight into the integrity of semantic memory (De Marco et al., 2021).

Figure 5 Shows example category fluency serial recall coefficients for a participant of the control group.

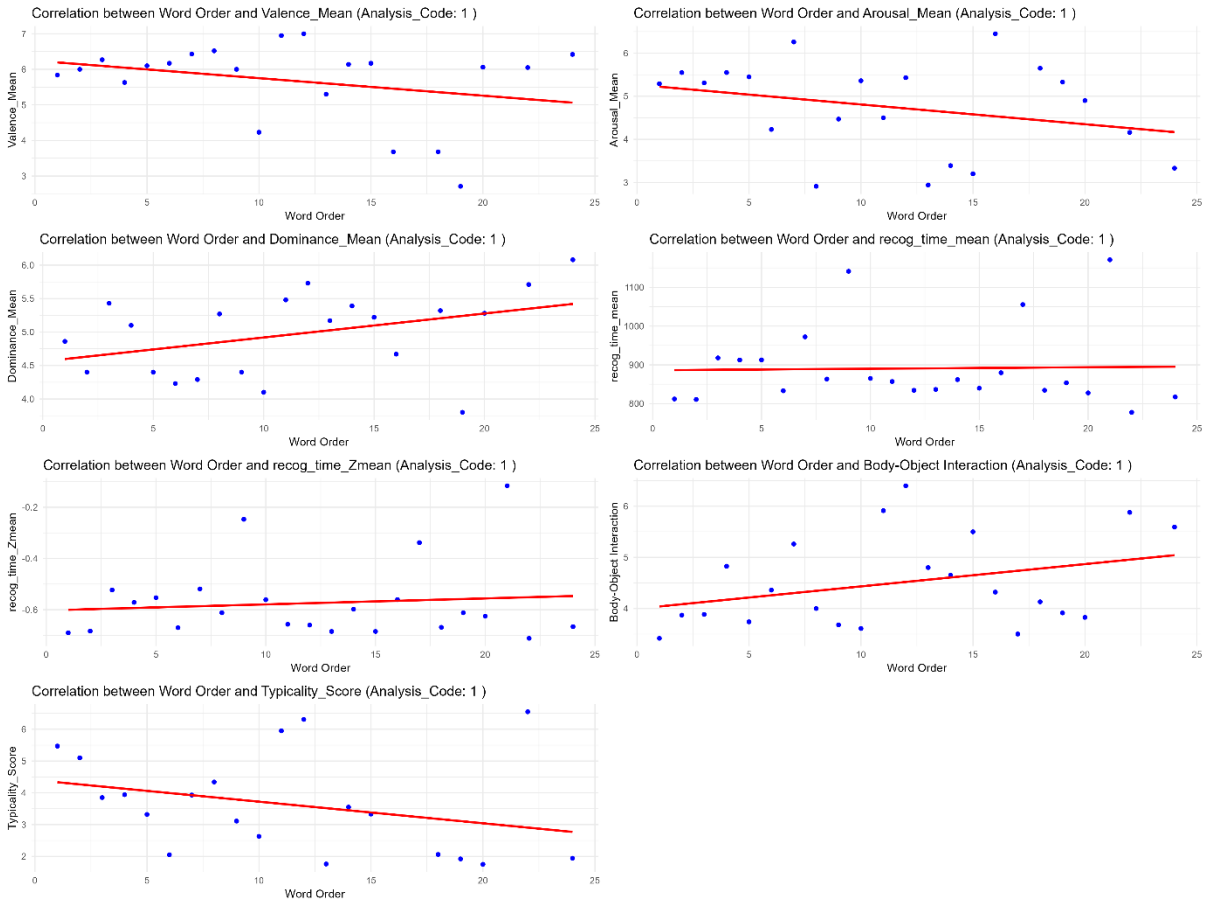


Figure 5 Example Serial Recall Values for Control Subject

Over the course of the fluency task values for typicality appear to decline, potentially reflecting the increasing difficulty in generating suitable words. Superior semantic performance would be expected to result in a relatively shallower gradient.

4.2.5 Note on Statistical Analysis

All statistical analyses was performed in R. Comparisons between features used normalized values (Z scores) unless provided purely for descriptive purposes. The relevant statistical test is explained in each section to aid interpretation.

4.3 Advanced Semantic Feature Analysis for MCI Detection: Results

4.3.1 Introduction to Results

Throughout this section we will undertake a detailed assessment of a series of semantic and imaging features, how they relate to one another and how they may contribute towards the identification of prodromal Alzheimer's disease. This is directed primarily towards the second of this project's hypotheses:

- Semantic features extracted from speech will meaningfully contribute towards the identification of prodromal Alzheimer's disease.

To address this hypothesis in a comprehensive and robust manner, the following section will explore the questions below:

1. Do the linguistic features' relationship to one another and selected regions of interest on structural imaging support their use as measures of semantic memory and processes?
2. Does the analysis of semantic function applied to participants' responses to standardized interview questions (administered using the same automated system) resemble the analysis of the verbal fluency tasks?
3. Does the semantic content of a participant's speech, as collected using automated verbal fluency tests, enable accurate classification between Healthy Controls and patients with MCI?
4. Is the discriminant value of these semantic features comparable to established structural imaging features?
5. Do the language features, cognitive test scores and imaging features combined contribute to greater diagnostic accuracy than the information typically available after an initial appointment?

4.3.2 Relationship Between Semantic and Imaging Features

This section explores the relationship between the structural imaging features and semantic features extracted from language. Internal relationships between the language-based features are assessed as well as whether the method of assessment (through the whole interview or just the verbal fluency tasks) has an impact. This aims to address questions one and two.

4.3.2.1 Voxel-Based Morphometry Comparisons Between Groups

To begin exploring the relationship between structural imaging and semantic features I first undertook a voxel-based morphometric study to compare the characteristics of the control and MCI. Identifying areas of atrophy in the MCI group compared to controls that are consistent with previous literature would support the premise that the MCI cohort is typical, at least in terms of imaging features, for prodromal Alzheimer's disease. This would support the use of structural imaging features in the later diagnostic classification analyses.

Given that the earliest structural changes in prodromal Alzheimer's disease often occur in brain regions associated with semantic processing (Braak et al., 2006; Didic et al., 2011), confirming the presence of such changes—alongside measurable semantic deficits in language—would strengthen the internal validity of our findings. Demonstrating a correspondence between structural and semantic features would provide confidence that the observed language impairments are genuinely related to underlying neurodegenerative processes. This alignment would enhance interpretability and support the use of these regions as targets for individual-level segmentation and volumetric analysis in further work.

Table 5 Voxel Based Morphometry Comparing Controls and MCI

Region	***x***	***y***	***z***	***T value***	***Z score***	***Cluster size***	***Cluster_ pFWE***	***Peak pFWE***
Right Putamen	30	-4	11	6.1	5.59	174	0	0.001

At a significance threshold of $p < 0.05$, a minimum cluster threshold of 20 voxels and following correction for multiple comparisons the only statistically significant difference between the control group and the entire MCI cohort was found in the right putamen (peak voxel MNI coordinate x = 30, y = -4, z = 11; cluster size 174, pFWE = 0.001).

Figure 6 shows a statistical parametric map highlighting the region of significantly reduced grey matter volume in the MCI group. For ease of interpretation this is overlaid on a standard anatomical template.

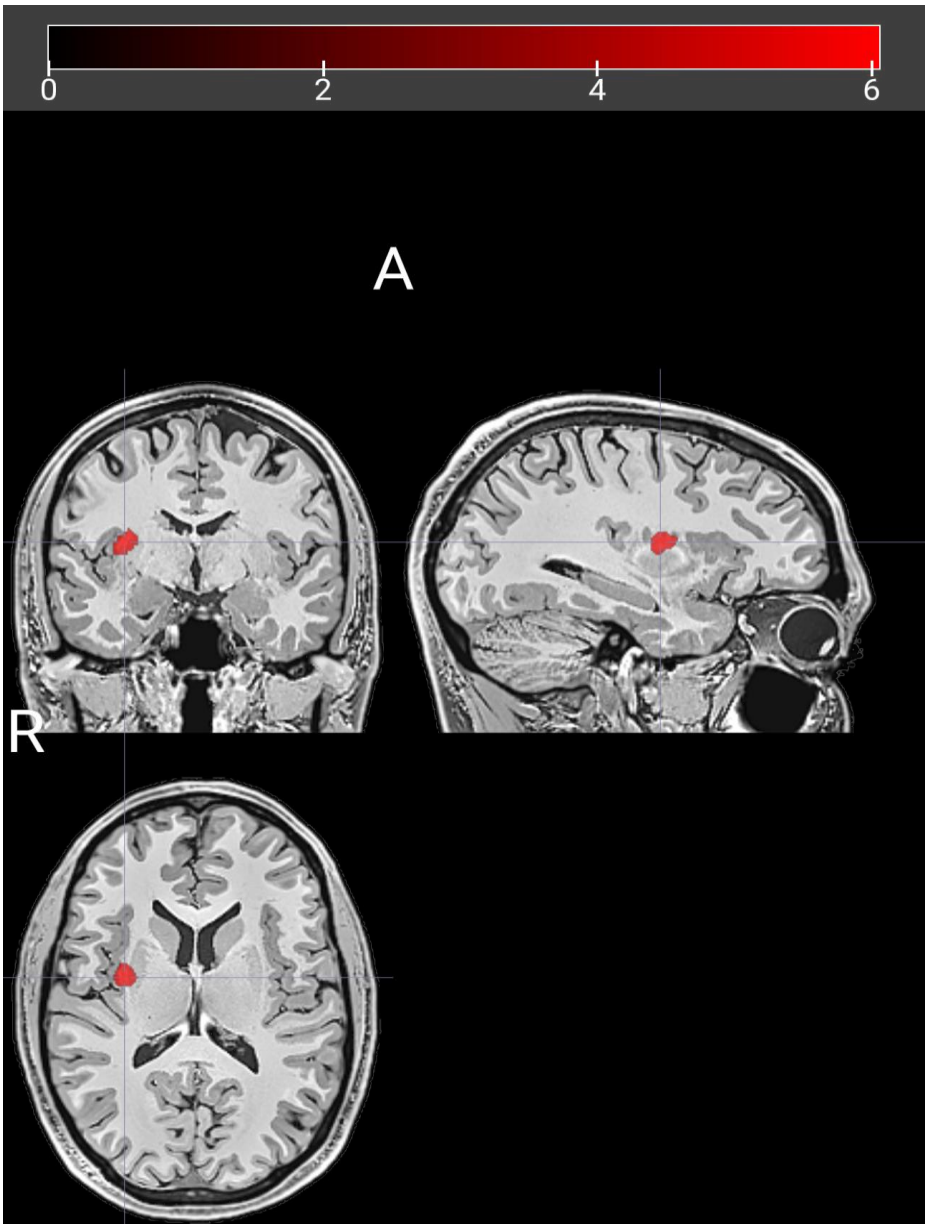


Figure 6 Image- Voxel Based Morphometry comparing Controls with MCI

Unilateral reduction of the right putamen volume was the only difference between the controls and the MCI group following correction for multiple comparisons. This would not be considered a typical finding in prodromal Alzheimer's.

To refine the analysis, I proceeded to run a repeat voxel-based morphometric analysis on the ADMCI cohort (as discussed in the methodology section). The same thresholds for significance and cluster size were applied.

Table 6 presents the results of voxel-based morphometry (VBM) comparing healthy controls (HC) with participants with Alzheimer's disease-related mild cognitive impairment (ADMCI). Regions with significant grey matter volume differences are listed, including MNI coordinates, T-values, Z-scores, cluster sizes, and corrected p-values. Only peak voxels for each cluster are reported here.

Table 6 Voxel-Based Morphometry Comparing Controls with ADMCI

Region	*MNI Coordinates (x, y, z)*	*T-Value*	*Z-Score*	*Cluster Size (voxels)*	*Cluster pFWE*	*Peak pFWE*
Right Putamen	30, -6, 8	7.50	6.22	1325	< .001	< .001
Left Parahippocampal Gyrus	-21, -34, 0	7.31	6.10	730	< .001	< .001
Left Lentiform Nucleus	-27, -14, -8	7.18	6.02	615	< .001	< .001
Right Thalamus	18, -20, 14	6.99	5.90	1007	< .001	< .001
Left Caudate	-16, -4, 22	5.75	5.08	67	.001	.012
Left Occipital Lobe	-14, -70, 5	5.87	5.16	56	.002	.008
Right Posterior Cingulate	8, -55, 24	6.08	5.30	43	.003	.004
Right Anterior Cingulate	10, 42, 8	5.80	5.11	47	.003	.010
Left Parietal Lobe	-24, -64, 46	6.29	5.45	37	.004	.002
Right Occipital Lobe	26, -66, -10	6.03	5.27	42	.004	.005
Left Temporal Lobe	-30, -40, -14	5.85	5.15	27	.007	.009

Note. Coordinates are given in MNI space. Cluster p-values reflect family-wise error correction at the cluster level. Peak p-values refer to correction at the voxel level.

Again, the largest cluster was observed in the area of the right putamen. The left parahippocampal gyrus was significantly smaller in the ADMCI group compared to healthy controls (peak voxel MNI coordinate x = -21, y = -34, z = 0; cluster size 730, pFWE < 0.001). Significant differences were also observed in the right posterior cingulate gyrus and to a slightly lesser extent the right anterior cingulate gyrus.

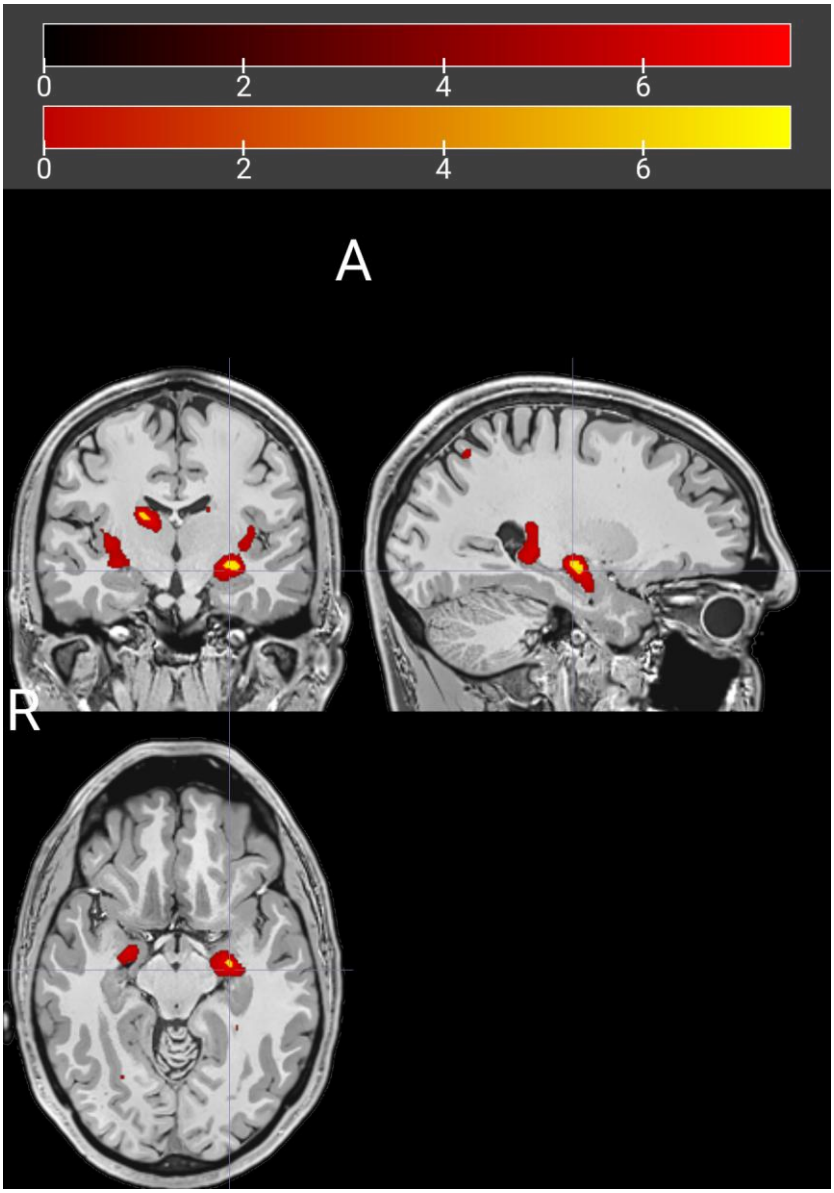


Figure 7 Statistical Parametric Map (VBM) Controls compared to ADMCI.

Figure 7 shows a statistical parametric map highlighting the region of significantly reduced grey matter volume in the ADMCI group when compared with the control group. The areas in red reflect differences in volume meeting statistical significance value of < 0.05. The overlaid areas in yellow demonstrate which clusters survive the more stringent significance values of < 0.001.

The relative atrophy of the left parahippocampal gyrus, right posterior cingulate, left temporal lobe and left parietal lobe would all be consistent with expectations in prodromal Alzheimer's. However, the changes observed in the occipital lobes and deep grey matter structures (most consistently the right putamen) would not be considered typical for Alzheimer's disease.

The fact that areas of difference known to be associated with Alzheimer's disease emerged as statistically significant when the ADMCI group was selected for analysis

suggests that (excluding the unexpected findings, the potential reasons for which will be discussed in a later section) these patients have some typical features for Alzheimer's disease and that comparing structural features (such as hippocampal volume) has some validity.

The left temporal lobe is strongly associated with semantic function, with the parahippocampal gyrus and parietal lobes also contributing towards semantic control and support mechanisms (Binder & Desai, 2011).

Given concerns regarding the reliability of the voxel-based imaging approach, efforts I decided to rely only on the hippocampal volumes (obtained through automated hippocampal segmentation, as described in methodology section (Apostolova et al., 2006)(Jorge Cardoso et al., 2013)

4.3.2.2 Comparisons between semantic Features and Structural Imaging Features

To explore the relationship between semantic fluency features derived from the category fluency test and structural imagine features (hippocampal volume), partial correlation analyses were conducted (these partial correlations controlled for age and sex, selected on the basis that these values differed significantly between groups). As a significant number of the variables violated the conditions of normality, a non-parametric approach was chosen (Spearman Rho). False Discovery Rate (FDR) correction was applied to p-values to control for multiple comparisons.A subset of features showed significant associations with left, right, and combined hippocampal volumes, as reported in Table 7.

Table 7 Significant Partial Correlations Between Category Fluency Features and Hippocampal Volumes

Category Fluency Feature	*Structural Feature*	*Correlation (r)*	*p-value*
Category Fluency Serial Recall Typicality	Combined Hippocampal Volume: TIV	0.25*	1.54e-02
Category Fluency Serial Recall Typicality	Left Hippocampus: TIV	0.25*	1.69e-02
Category Fluency Serial Recall Typicality	Right Hippocampus: TIV	0.24*	1.93e-02
Category Fluency Correct Score	Right Hippocampus: TIV	0.23*	2.85e-02
Category Fluency Correct Score	Combined Hippocampal Volume: TIV	0.22*	3.71e-02
Category Fluency Repetitions	Left Hippocampus: TIV	-0.2*	4.76e-02
* p < .05, ** p < .01, *** p < .001			

Among the semantic features derived from the category fluency task, serial recall typicality was the only variable to show a significant correlation with hippocampal volumes. The correlation was stronger with the left hippocampus than with the right.

The number of repetitions was negatively correlated with left hippocampal volume, suggesting that individuals with more left hippocampal atrophy tended to produce more repeated responses. In contrast, the number of correct category fluency responses was positively correlated with right hippocampal volume, indicating better performance was associated with greater right hippocampal integrity.

With the exception of serial recall typicality, no other semantic features showed statistically significant correlations with hippocampal volume. However, several features exhibited non-significant negative correlations, including overall valence, typicality, frequency, prevalence, and the number of repetitions. The strongest of these was between category fluency repetitions and left hippocampal volume.

Positive, non-significant correlations were observed between the cluster related features, age of acquisition, arousal, dominance, recognition time and concreteness.

The serial recall order coefficients for age of acquisition and dominance were negatively correlated with hippocampal volumes whilst the serial recall order coefficients for body-interaction, recognition time, arousal and valence were negative. These correlations did not meet statistical significance. Figure 8 shows all of the correlations in a heatmap.

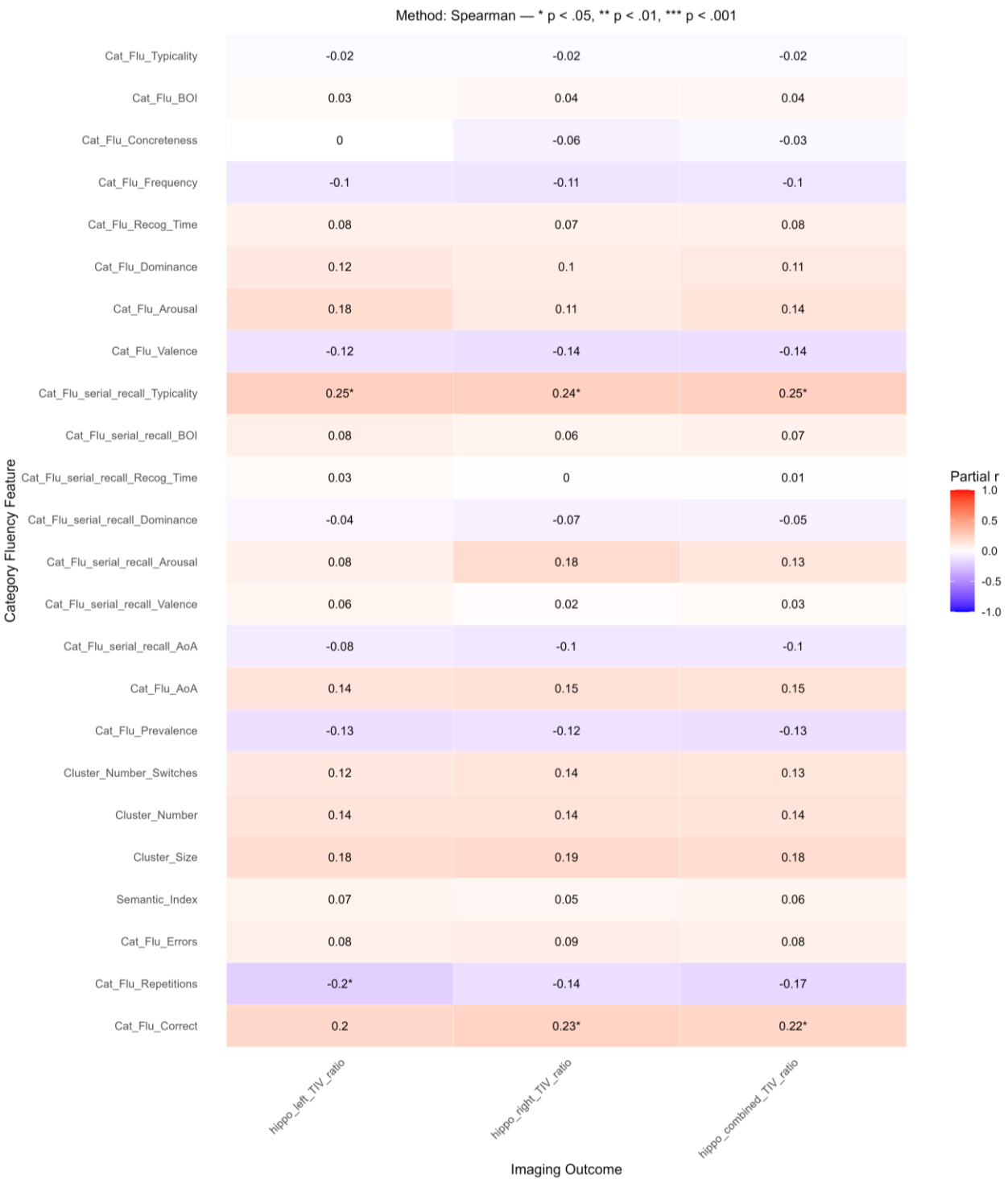


Figure 8 Heatmap of Partial Correlations between Category Fluency Features and Hippocampal Volumes

4.3.2 Comparisons between various Verbal Fluency Derived Semantic Features

To explore the relationships between the various semantic fluency features derived from the category fluency test, correlation analyses were conducted (controlling for age and sex). As a significant number of the variables violated the conditions of normality, a non-parametric approach was chosen (Spearman Rho). False Discovery Rate (FDR) correction was applied to p-values to control for multiple comparisons.

The results are summarised in Figure 9 .The colour gradient reflects the strength and direction of the correlation (blue = negative, red = positive, white = neutral). Correlation values are shown in each cell, with an asterisk (*) indicating statistical significance ($p < .05$).

The heatmap demonstrates a generally high level of correlation between the semantic features. The strongest negative correlation was between age of acquisition and frequency (Spearman's $p = -0.88$, $p = <0.001$). Age of acquisition also featured in the strongest positive correlation with recognition time (Spearman's $p = 0.73$, $p = <0.001$). This makes intuitive sense, with increasing age of acquisition and recognition time one would expect to see more complex verbiage, which is encountered less frequently in day-to-day parlance.

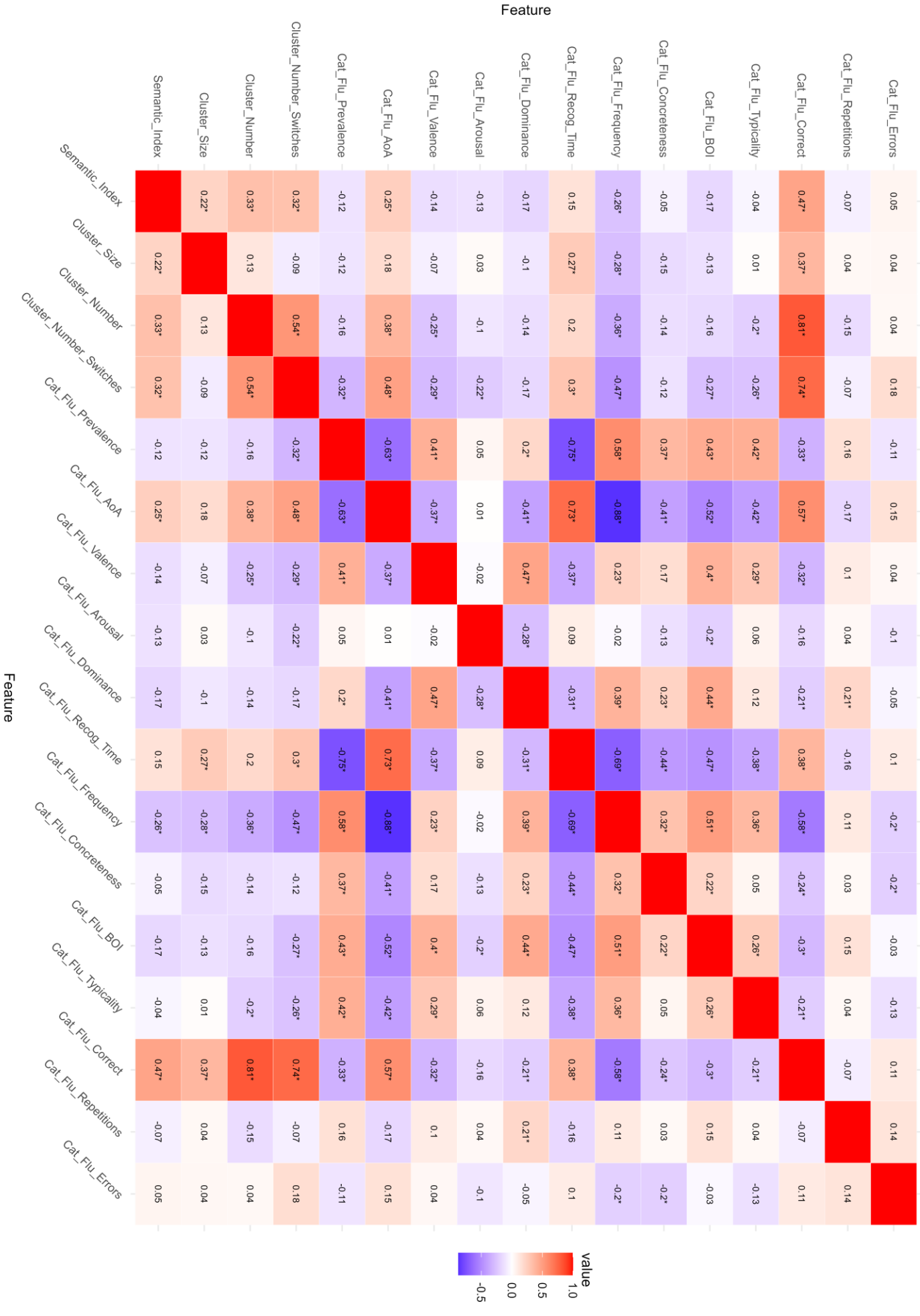


Figure 9 Heatmap Demonstrating Correlations Between Semantic Features (VFT)

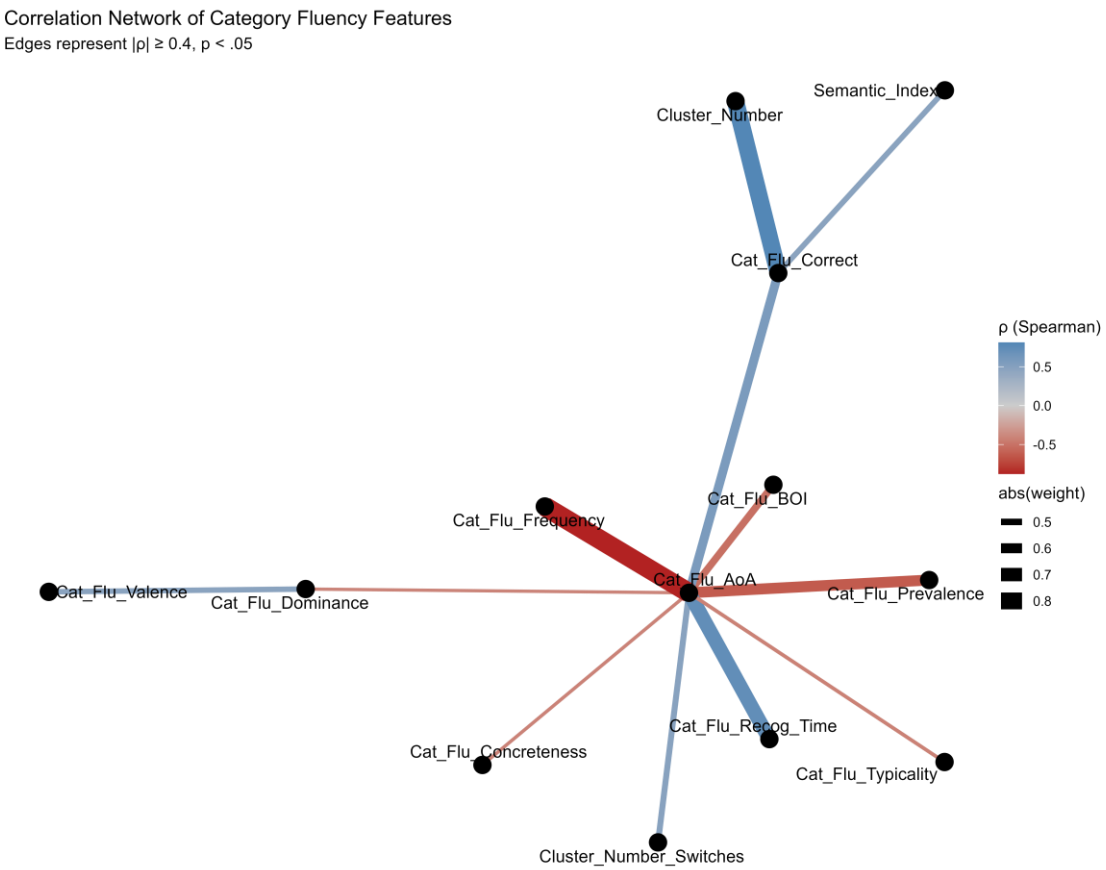


Figure 10 Correlation network of category fluency-derived semantic features.

To visualize the structure of interrelationships among semantic fluency features, a correlation network graph was generated (Figure 10). This graph revealed several tightly interconnected clusters, particularly among lexical-semantic variables such as age of acquisition, frequency, and recognition time. These variables formed a coherent subnetwork. Emotional-semantic features, such as valence and dominance, were moderately connected to lexical features, though more strongly correlated to one-another. In contrast, measures such as cluster size and number of switches appeared more peripheral. These results underscore the multidimensional nature of semantic fluency, with discernable clusters of features representing lexical, emotional, and structural systems.

4.3.2.4 Comparisons between Semantic Features Derived from the Category Fluency Task and from the Whole Interview

To explore the relationships between the semantic fluency features derived from the category fluency test and those derived from the participants responses to the rest of interview questions, correlation analyses were conducted (controlling for age and sex). As a significant number of the variables violated the conditions of normality, a non-parametric approach was chosen (Spearman Rho). False Discovery Rate (FDR) correction was applied to p-values to control for multiple comparisons.

Table 8 Comparison of Semantic Features from Category Fluency and Whole Interview Tasks

Semantic Feature	*Cat Fluency Mean ± SD*	*Whole Interview Mean ± SD*	*Spearman's ρ*	*p-value*	*Sig.*
Age of Acquisition	5.34 ± 0.82	5.84 ± 0.39	0.31	0.0018	*
Valence	5.96 ± 0.28	5.78 ± 0.17	0.08	0.4600	
Arousal	4.26 ± 0.34	3.98 ± 0.15	-0.08	0.4300	
Dominance	5.08 ± 0.26	5.60 ± 0.14	-0.12	0.2300	
Concreteness	4.87 ± 0.13	3.05 ± 0.18	-0.05	0.6300	
Body-Object Interaction	4.61 ± 0.46	3.11 ± 0.30	0.21	0.0370	*
Recognition Time	879.86 ± 41.19	886.61 ± 32.27	0.19	0.0640	
Frequency	4.03 ± 0.27	5.11 ± 0.16	0.07	0.5200	

Note. * Indicates p < .05 (two-tailed Spearman correlation).

As demonstrated in Table 8, statistically significant correlations were observed

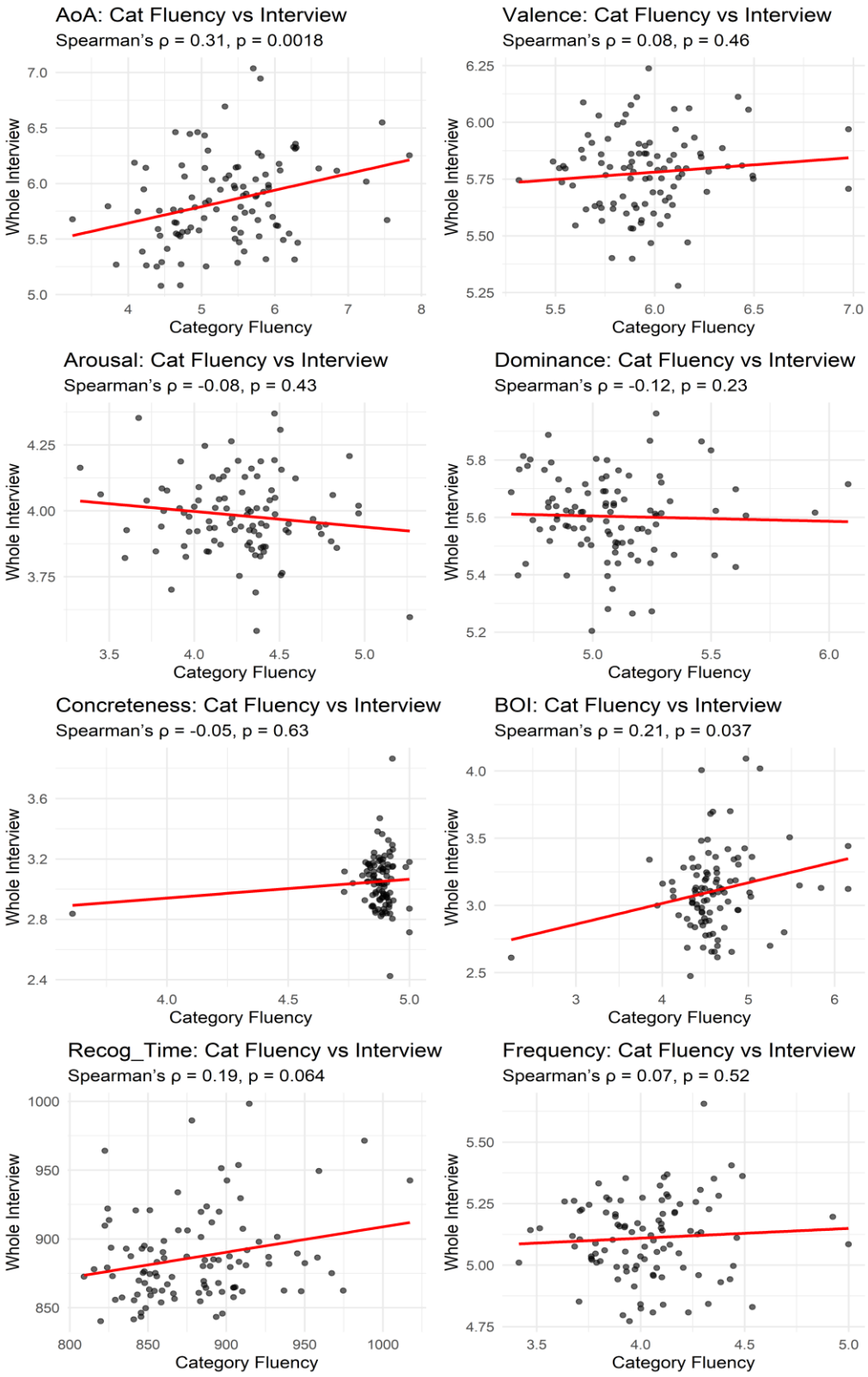


Figure 11 Scatterplots Comparing Matched Semantic Features Between Category Fluency and Whole Interview Tasks

between the age of acquisition derived from the category fluency test and the whole interview (Spearman's p = 0.31, p = 0.002), and the body-object interaction (Spearman's p = 0.21, p = 0.037). The rest of the features did not correlate significantly. The scatter plots seen in Figure 11 illustrate the distribution of values across the features examined.

These results suggest that values for age of acquisition and body-object interaction extracted from free speech may demonstrate similar performance in later tasks as values derived from the category fluency test. Many patients find the timed nature of the verbal fluency tasks to be unpleasant, reporting anxiety over performance. If meaningful values could be derived from natural, unstructured speech this could be attractive.

4.3.3 Comparison of Features between groups

To explore the potential diagnostic utility of the various semantic and imaging features I evaluated how they compared between groups. The following section begins to address the questions three, four and five. The majority of the group comparisons were conducted using linear models with diagnosis as the predictor and age and sex as covariates. Separate models were fitted for each feature. For features violating model assumptions of normality, robust linear models were used. FDR correction was applied across each feature group.

4.3.3.1 Group comparison of Hippocampal Volumes

To investigate the structural differences between the groups further, hippocampal volumes were quantified (according to previously described methods) and compared across groups using the Mann-Whitney U Test (as the assumption for parametric testing were violated).

Table 9 provides the average hippocampal volumes for the control and MCI groups. The average left hippocampal volumes for the control group is 2.63 cm^3 whilst for the MCI group it is significantly smaller at 2.29. Similar patterns were observed for the right hippocampal volumes and total intracranial volumes.

Interestingly the effect size was more pronounced for the difference in right hippocampal (r = 0.429) than the left (r = 0.371). This is somewhat contrary to what might be expected in MCI due to Alzheimer's, where typically left hippocampal atrophy exceeds right early in the disease process (Ferreira et al., 2011).

Table 9 Hippocampal Volumes in Control and MCI groups

	Healthy Controls (Median)	*MCI (Median)*	*U Statistic*	*p-value*	*Effect Size (r)*
Left Hippocampal Volume (cm^3)	2.56	2.29	1787.5	0.0	0.371
Right Hippocampal Volume (cm^3)	2.63	2.29	1872.0	0.0	0.429
Total Intracranial Volume (cm^3)	1634.74	1494.89	1896.0	0.0	0.445

Note. Medians are presented for each group. Effect sizes (r) interpreted as: small ≥ 0.1, medium ≥ 0.3, large ≥ 0.5.

Table 10 demonstrates that this pattern persisted when the ADMCI sub-cohort was compared with the matched control group. The effect sizes were somewhat smaller and the p values somewhat higher. This is more likely to reflect the smaller cohort than a meaningful difference between the cohorts; it may suggest that a significant proportion of the MCI cohort who did not have a confirmed diagnosis of Alzheimer's disease (based on biomarkers or progression) nonetheless have the disease. Neither of the groups demonstrate the left side preponderance more commonly reported.

Table 10 Hippocampal Volumes in Control and ADMCI Groups

	Healthy Controls (Median)	*ADMCI (Median)*	*U Statistic*	*p-value*	*Effect Size (r)*
Left Hippocampal Volume (cm^3)	2.62	2.36	656.5	0.013	0.315
Right Hippocampal Volume (cm^3)	2.73	2.36	678.0	0.006	0.353
Total Intracranial Volume (cm^3)	1714.01	1477.17	785.0	0.0	0.544

Note. Medians are presented for each group. Effect sizes (r) interpreted as: small ≥ 0.1, medium ≥ 0.3, large ≥ 0.5.

4.3.3.2 Group comparison of Cognitive Assessment and Basic Verbal Fluency Scores

To provide a context for the semantic features, a linear model was applied to compare the differences in MMSE and verbal fluency scores between the control and MCI group, controlling for age and sex.

Table 11 Comparison of established features between MCI and control groups and ADMCI and control groups

Feature	*MCI Estimate*	*MCI Std. Error*	*MCI FDR-adj p*	*ADMCI Estimate*	*ADMCI Std. Error*	*ADMCI FDR-adj p*
MMSE	-1.37	0.17	0.000	-1.05	0.17	0.000
Category Fluency Correct	-0.86	0.20	0.000	-0.65	0.27	0.057
Category Fluency Repetitions	0.20	0.22	0.699	0.19	0.22	0.611
Category Fluency Errors	-0.11	0.22	0.807	-0.38	0.28	0.364
Phonemic Fluency Correct	-0.96	0.19	0.000	-0.93	0.26	0.003
Phonemic Fluency Repetitions	-0.16	0.22	0.760	0.06	0.21	0.899
Phonemic Fluency Errors	0.04	0.21	0.933	0.03	0.28	0.927
Semantic Index	0.02	0.22	0.933	0.21	0.30	0.642

As Table 11 demonstrates, significant differences were observed in MMSE scores, with both MCI (Estimate = -1.37, p<.001, FDR-adjusted) and ADMCI groups (Estimate = -1.05, p<.001) scoring lower than healthy controls. Category fluency total correct responses were also significantly reduced in the MCI group (p<.001) and

showed a trend toward significance in the ADMCI group (p=.057). Phonemic fluency correct scores were significantly lower in both MCI (p<.001) and ADMCI (p=.003) groups compared to controls. However, error and repetition counts for both fluency tasks, as well as semantic index scores, did not differ significantly across groups after FDR correction. These findings suggest that core fluency measures (MMSE, total correct responses on verbal fluency tests) are sensitive to early cognitive decline, while error-based metrics offer limited discriminative utility.

Figure 12 illustrates how the features differ between the controls and the MCI and ADMCI groups. Similar patterns are observed in both the MCI groups. The MMSE is feature that demonstrates the largest estimate size in both groups, though in the ADMCI group the number of correct responses on the phonemic fluency test is quite close. This is unexpected, as patients with prodromal Alzheimer's disease typically perform worse on the category fluency test that the phonemic fluency test (Henry & Crawford, 2004). This results in the so called "Semantic advantage" observed in controls (and in this study labelled the semantic index). Given that phonemic fluency was more impaired than might be expected in the MCI groups, it is not surprising that the semantic index does not differ significantly between the controls and MCI groups.

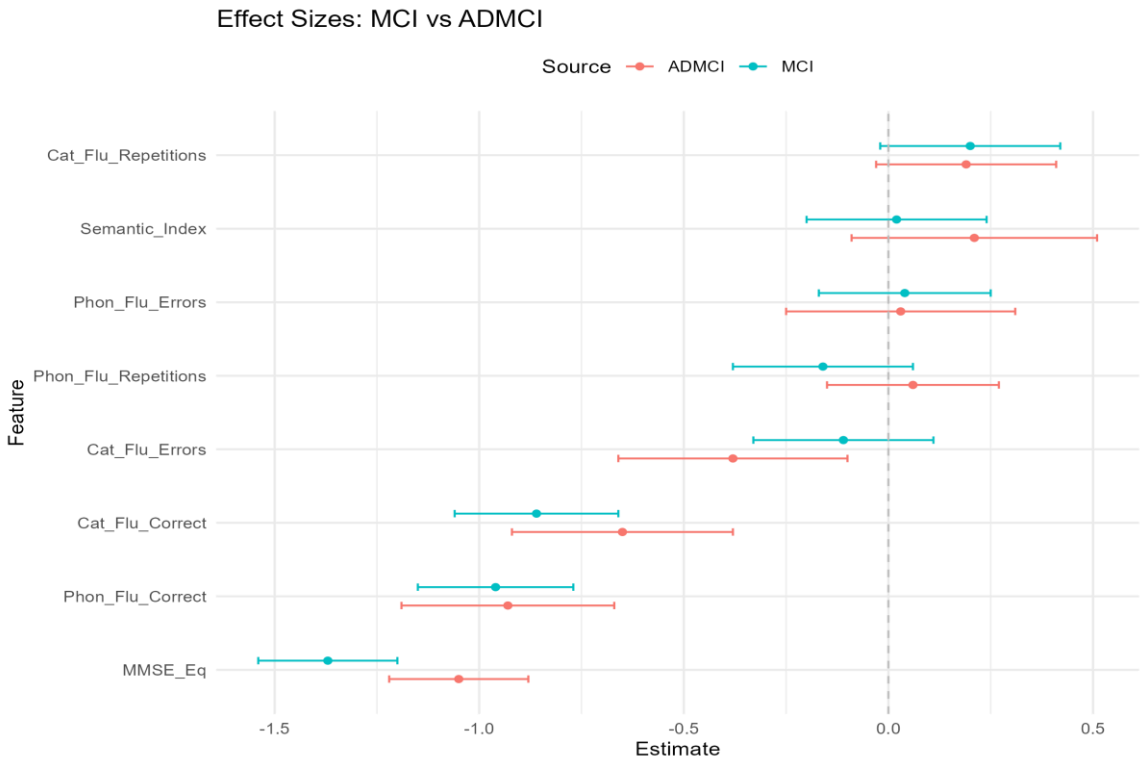


Figure 12 Overlay Forest Plot: Established Features in Controls vs MCI/ADMCI

4.3.3.3 Group comparison of verbal fluency derived semantic features.
Linear models were built to compare the various semantic features between groups. The features were all derived from the responses to the category fluency task. As with previous models, age and sex were included as covariates and FDR correction was applied to correct for multiple comparisons.

Table 12 demonstrates the mean values for the various features across the different groups. Related features have been grouped together, categories Lexical/ Semantic- Age of Acquisition, Recognition Time, Frequency, Prevalence and Typicality; Affective/ Sensorimotor- Valence, Arousal, Dominance, Concreteness and Body-Object Interaction; Structural- Cluster Size, Number of Clusters, Number of Switches.

Table 12 Descriptive comparison of VFT semantic features between groups

Feature	*Healthy Controls*	*MCI*	*ADMCI*
Age of Acquisition	5.58 ± 0.70	5.10 ± 0.87	5.11 ± 0.87
Recognition Time	890.47 ± 40.90	869.26 ± 39.06	870.14 ± 38.48
Frequency	3.95 ± 0.18	4.11 ± 0.31	4.11 ± 0.31
Prevalence	2.29 ± 0.17	2.35 ± 0.13	2.36 ± 0.11
Typicality	4.39 ± 0.60	4.82 ± 0.80	4.84 ± 0.69
Valence	5.92 ± 0.20	6.00 ± 0.33	6.01 ± 0.30
Arousal	4.23 ± 0.29	4.30 ± 0.37	4.33 ± 0.40
Dominance	5.06 ± 0.22	5.09 ± 0.29	5.08 ± 0.30
Concreteness	4.85 ± 0.18	4.89 ± 0.05	4.89 ± 0.04
Body-Object Interaction	4.50 ± 0.42	4.72 ± 0.47	4.71 ± 0.40
Cluster Size	3.02 ± 0.67	2.76 ± 0.89	2.82 ± 1.00
Number of Clusters	5.08 ± 1.88	3.72 ± 2.01	4.03 ± 2.29
Number of Switches	8.54 ± 2.87	6.30 ± 3.76	6.52 ± 3.95

Values represent Mean ± SD.

The MCI and ADMCI groups, on average, employ words with higher age of acquisition, recognition time, valence, arousal, dominance, concreteness and body-object interaction values than the healthy controls. It is important to note that in some cases the difference is minimal.

Fewer clusters were produced by the MCI and ADMCI groups compared to the healthy controls. These clusters were, on average, smaller. There was a concomitant reduction in switches observed also.

Frequency, prevalence and typicality scores were higher across the MCI and ADMCI groups.

For the most part the MCI and ADMCI groups behave very similarly. The most striking differences are observed in the cluster-related categories, with ADMCI patients producing smaller clusters and more clusters and switches.

Linear Model: Semantic and Imaging Features compared between control and MCI groups.

Figure 13 presents a forest plot displaying the estimated group differences across semantic and imaging features between healthy control and MCI groups. Each horizontal line represents the regression estimate and 95% confidence interval for a feature from the linear model, adjusted for age and sex. Positive estimates indicate higher values in the MCI group, while negative estimates reflect lower values in MCI relative to controls. Features with confidence intervals that do not cross zero are statistically significant after adjusting for multiple comparisons using the false discovery rate (FDR) method.

Table 13 shows the full results of the linear model. This demonstrates a significant difference between groups in the ratio between the right hippocampus and the total intracranial volume. The estimate is noted to be small. This suggests a subtle but significant difference in the hippocampal values between the control and MCI groups.

Of the lexical/semantic features, neither typicality nor prevalence were significantly different between the groups. Age of acquisition and recognition time demonstrated negative estimates and were significantly different between groups.

Of the affective/ sensorimotor features, only body-object interaction demonstrated a significant between group difference demonstrating a positive estimate.

The differences in clustering features were all significant, with the MCI group producing fewer clusters and switches alongside smaller clusters.

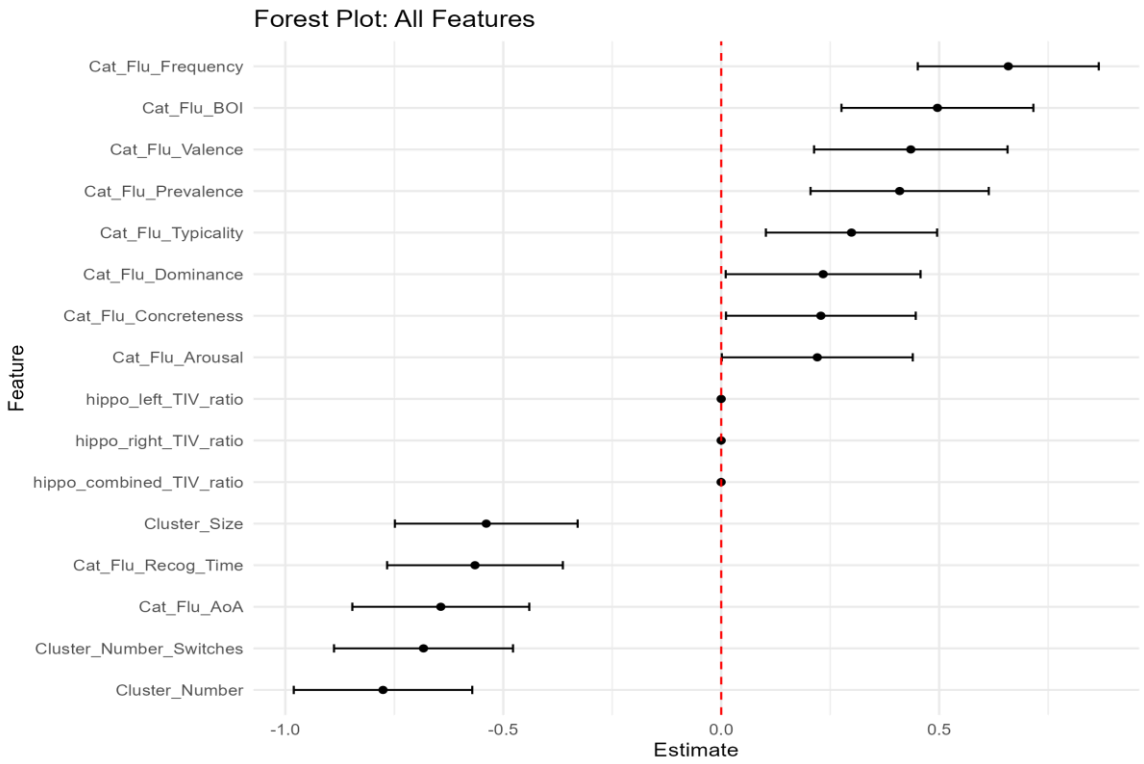


Figure 13 Forest Plot: Linear Model comparing Imaging and Semantic Features between Healthy Controls and MCI

Table 13 Full results of Linear Model comparing Imaging and Semantic Features Between healthy controls and MCI

Feature	*Estimate*	*Std. Error*	*p-value*	*FDR-adjusted p*
Left Hippocampus: TIV	-0.00	0.00	0.065	0.086
Right Hippocampus: TIV	-0.00	0.00	0.008	0.021
Combined Hippocampus: TIV	-0.00	0.00	0.023	0.045
Age of Acquisition	-0.64	0.20	0.002	0.008
Recognition Time	-0.56	0.20	0.006	0.020
Frequency	0.66	0.21	0.002	0.008
Prevalence	0.41	0.20	0.048	0.077
Typicality	0.30	0.20	0.131	0.162
Valence	0.43	0.22	0.053	0.077
Arousal	0.22	0.22	0.317	0.317
Dominance	0.23	0.22	0.298	0.317
Concreteness	0.23	0.22	0.296	0.317
Body-Object Interaction	0.50	0.22	0.027	0.047
Cluster Size	-0.54	0.21	0.012	0.027
Cluster Number	-0.78	0.20	0.000	0.004
Cluster Number Switches	-0.68	0.21	0.001	0.008

Linear Model: Semantic and Imaging Features compared between control and MCI groups.

A linear regression model was run to compare the ADMCI group to healthy controls, using semantic and hippocampal features while controlling for age and sex. The results are shown in full in Figure 14. Although none of the features reached statistical significance after FDR correction, the direction of effects was consistent with findings from the HC vs MCI comparison (demonstrated by comparing Figure 13 and Figure 14). In particular, features such as age of acquisition, recognition time, and hippocampal volume measures continued to demonstrate associations in the expected direction, indicating increased semantic impairment and structural atrophy in the ADMCI group. The absence of statistically significant effects likely reflects reduced statistical power due to the smaller ADMCI sample size.

This was recognized to be a consistent pattern. Similar trends were demonstrated across both groups, with the features in the ADMCI cohort routinely failing to meet statistical significance. For the purposes of brevity, I will continue to provide results for the MCI group only unless significant differences were observed.

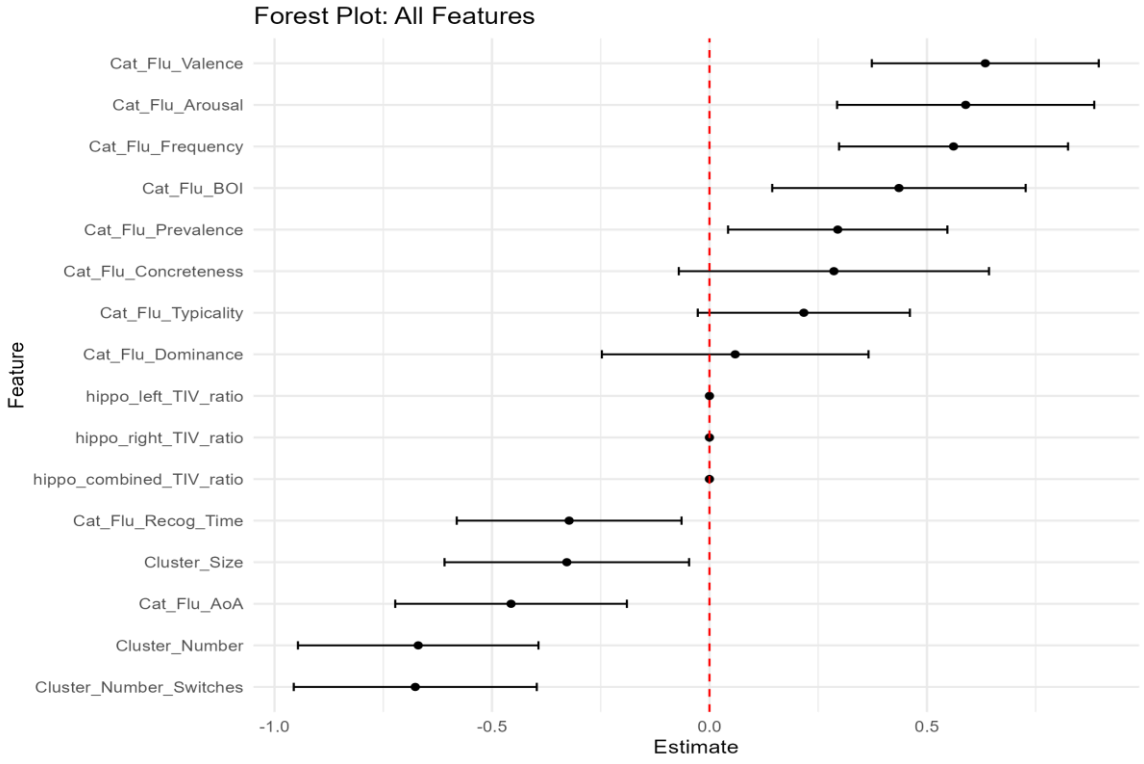


Figure 14 Forest Plot: Linear Model comparing Imaging and Semantic Features between Healthy Controls and ADMCI

Table 14 Full results of Linear Model comparing Imaging and Semantic Features Between healthy controls and ADMCI

Feature	*Estimate*	*Std. Error*	*p-value*	*FDR-adjusted p*
Left Hippocampus: TIV	-0.00	0.00	0.815	0.848
Right Hippocampus: TIV	-0.00	0.00	0.323	0.470
Combined Hippocampus: TIV	-0.00	0.00	0.542	0.619
Age of Acquisition	-0.46	0.27	0.092	0.245
Recognition Time	-0.32	0.26	0.217	0.397
Frequency	0.56	0.26	0.037	0.149
Prevalence	0.29	0.25	0.247	0.397
Typicality	0.22	0.24	0.377	0.503
Valence	0.63	0.26	0.018	0.099
Arousal	0.59	0.30	0.051	0.164
Dominance	0.06	0.31	0.848	0.848
Concreteness	0.29	0.36	0.426	0.524
Body-Object Interaction	0.44	0.29	0.140	0.321
Cluster Size	-0.33	0.28	0.248	0.397
Cluster Number	-0.67	0.28	0.019	0.099
Cluster Number Switches	-0.68	0.28	0.019	0.099

Linear Model Comparing Semantic Serial Recall Order Features between Healthy Controls and MCI

Linear models were constructed for each of the serial recall order coefficients. Each model examined the effect of group, while controlling for age and sex. The outcome variables reflected the linear trend (slope) of each semantic attribute—such as age of acquisition or valence—over the course of the fluency task, effectively capturing how word characteristics evolved as the task progressed.

Table 15 Full results of Linear Model comparing Semantic Serial Recall Features Between healthy controls and MCI.

Feature	*Estimate*	*Std. Error*	*p-value*	*FDR-adjusted p*
Serial Recall Order Age of Acquisition	0.51	0.21	0.018	0.129
Serial Recall Order Recognition Time	0.08	0.23	0.708	0.892
Serial Recall Order Typicality	-0.02	0.23	0.940	0.940
Serial Recall Order Valence	0.26	0.22	0.244	0.569
Serial Recall Order Arousal	-0.16	0.23	0.477	0.835
Serial Recall Order Dominance	0.07	0.23	0.765	0.892
Serial Recall Order Body-Object Interaction	-0.33	0.23	0.151	0.528

Table 15 demonstrates the results of the linear model. Among the serial recall features evaluated, age of acquisition showed the strongest trend toward group differentiation. The model yielded an uncorrected *p*-value of 0.018 and a positive estimate of 0.51, indicating that participants in the MCI group produced increasingly *later-acquired* words over the course of the fluency task compared to healthy controls. However, this effect did not survive correction for multiple comparisons (*FDR-adjusted p* = 0.129), suggesting the trend should be interpreted with caution.

Other features, including recognition time, typicality, valence, and dominance, did not show significant group differences. A small negative trend was observed for body-

object interaction (estimate = −0.33) but this also did not reach statistical significance.

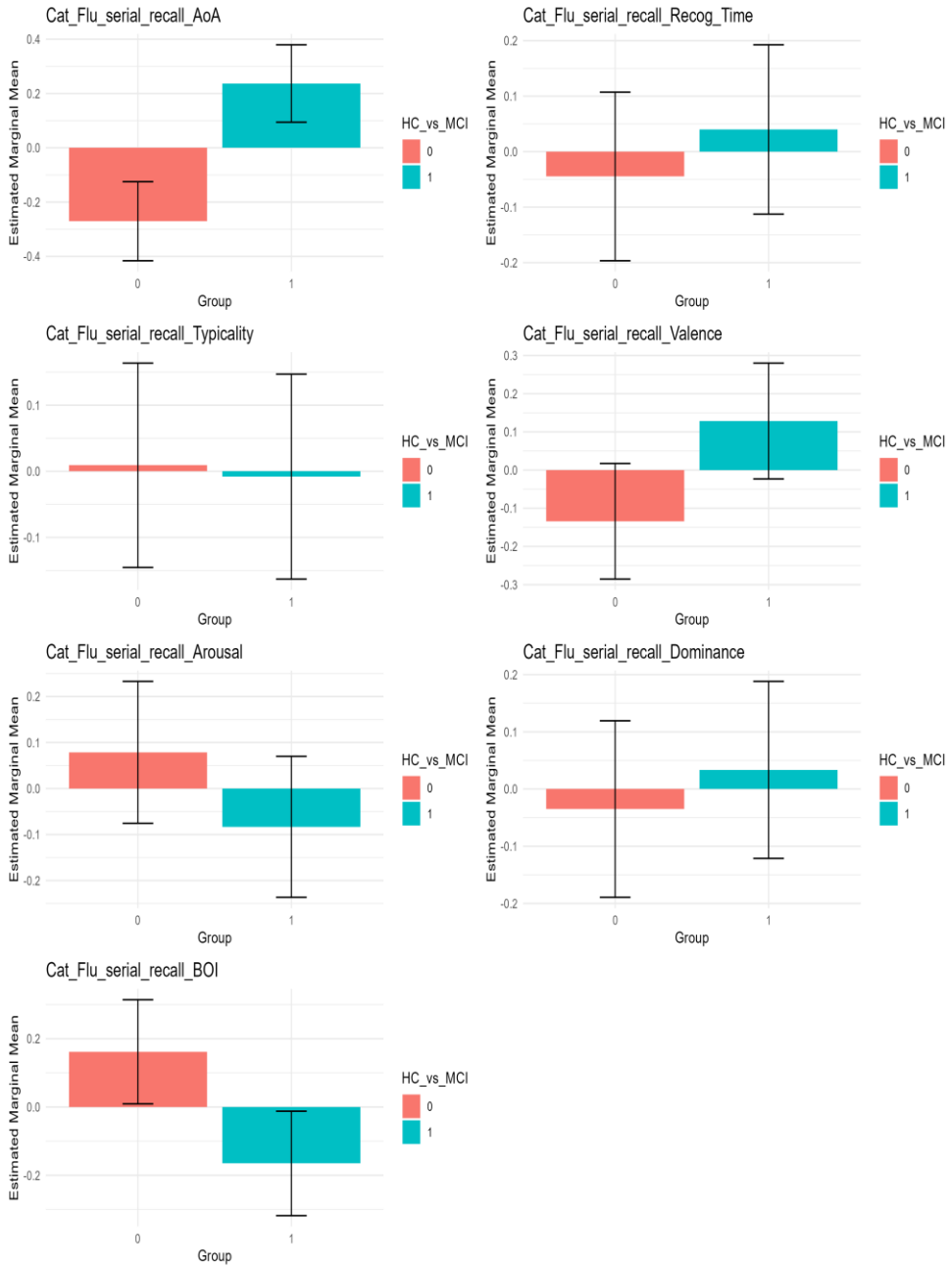


Figure 15 Marginal Means Plot: Comparing Serial Recall Order between control and MCI groups.

The widely overlapping error bars in typicality, recognition time, and arousal suggest against a significant difference between groups. Although not significant, the means plots for age of acquisition, valence and body-object interaction might suggest that these features could display group level differences with larger groups of larger sample size.

4.3.3.4 Group comparison of whole interview derived semantic features.

To explore the differences between whole interview derived semantic features between groups (and therefore inform whether these features might present a viable alternative to those derived from the verbal fluency tasks). I ran a linear regression model, conforming to the design described previously.

This did not detect any differences between the groups which survived FDR correction.

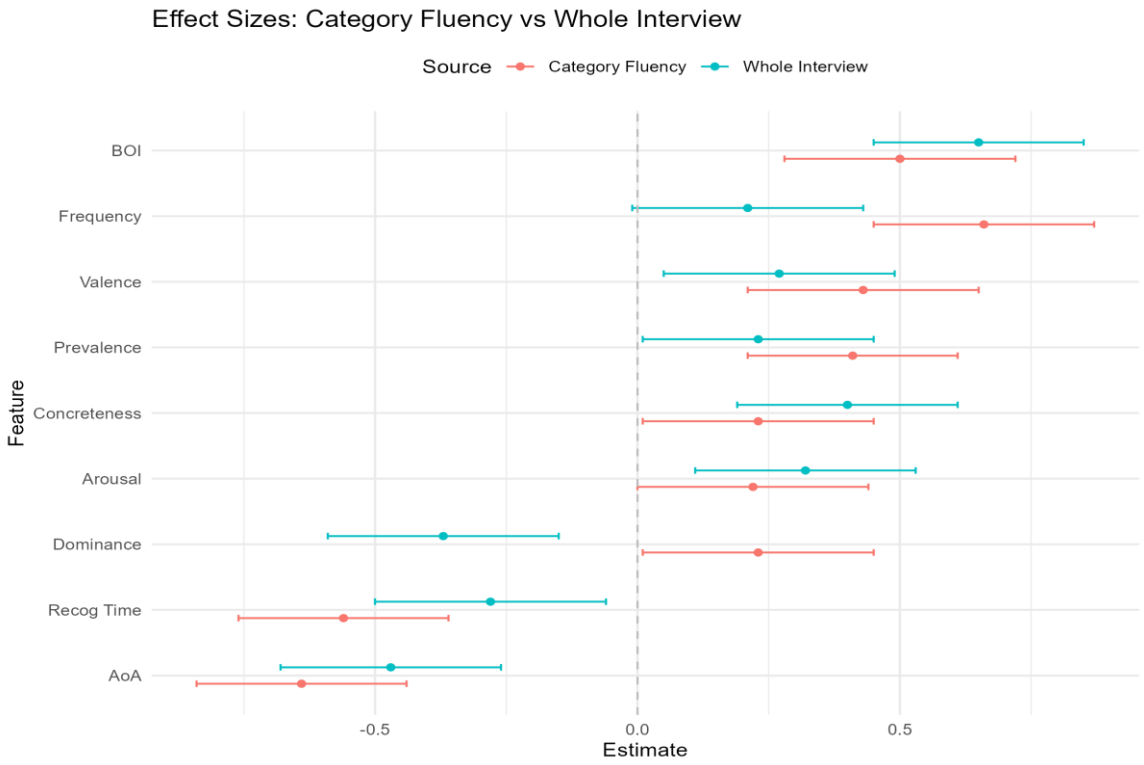


Figure 16 Overlay Forest Plot: Comparing Differences in Category Fluency vs Whole Interview between Control and MCI groups.

To visually compare the strength and direction of associations across tasks, I constructed overlaid forest plots (see Figure 16) combining the results from linear models of semantic features from the verbal fluency tasks and the whole interview.

Aside from dominance, all the features showed consistent trends across both tasks. The magnitude of effects tended to be more pronounced in features derived from the verbal fluency task. None of the features derived from the whole interview were significantly different between groups.

4.3.3.5 Group comparison of word count and filler word proportion

It has been frequently observed that patients with Alzheimer's disease acquire an "empty" quality of speech. This is distinct from the agrammatism encountered in non-fluent primary progressive aphasia or the frequent word retrieval issues reported in logopoenic variant primary progressive aphasia. We know that semantic deficits account for some of this characteristic. I wondered if overall verbosity (number of words produced throughout the whole interview) and number of filler words (representing lexical retrieval (Fagundo et al., 2008) would vary significantly between groups.

Linear models were constructed to compare the word count and proportion of filler words over the course of the whole interview (excluding the verbal fluency tasks). The design of these models is identical to those above.

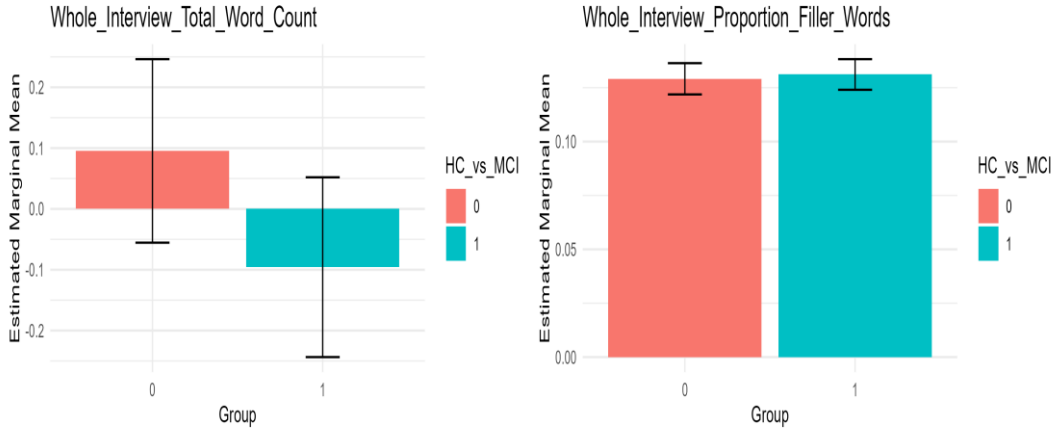


Figure 17 Marginal Means Plot: Comparing Word Count and Filler word proportion between control and MCI groups.

As Figure 17 shows, there was no significant difference in word count and filler word proportion between the two groups, suggesting against this features having particular discriminant value.

4.3.4 Summary of between feature and between group comparisons

With regards to how the findings relate to the questions underpinning this section:

1. Do the linguistic features' relationship to one another and selected regions of interest on structural imaging support their use as measures of semantic memory and processes?

Hippocampal volumes correlated with correct category fluency responses and serial recall order typicality. Whilst the rest of the correlations between semantic features and hippocampal volume did not reach significance, the trends reflected expectations based on established literature.

There was strong internal correlation between the semantic features. This largely reflected the different categories of semantic test (e.g. lexical, affective).

2. Does the analysis of semantic function applied to participants' responses to standardized interview questions (administered using the same automated system) resemble the analysis of the verbal fluency tasks?

Values for age of acquisition and body-object interaction derived from the whole interview and the verbal fluency task correlated significantly. Whilst the strength of the correlation between the rest of the features did not meet statistical significance, the trends appeared to be identical, suggesting that with larger samples, more significant findings may be observed.

3. Does the semantic content of a participant's speech, as collected using automated verbal fluency tests, enable accurate classification between Healthy Controls and patients with MCI?

MMSE showed robust and consistent group differences**,** with both MCI and ADMCI groups scoring significantly lower than healthy controls, highlighting its continued clinical relevance for detecting early cognitive decline.

A coherent trend was observed for features demonstrating significant difference between controls and MCI and controls and ADMCI. The features explored in the ADMCI group generally did not meet the threshold for statistical significance. It is posited that this may represent an effect of reduced sample size and that similarity in their behaviour across both groups supports the hypothesis that these features meaningfully differ between both the ADMCI and MCI groups and healthy controls.

Category and phonemic fluency total correct scores were significantly reduced in the MCI group, and phonemic fluency remained significantly impaired in the ADMCI group.

Category fluency errors, repetitions, and the semantic index did not differ significantly between groups, suggesting these conventional error-based metrics may lack sensitivity for early-stage disease differentiation.

Semantic features derived from verbal fluency, particularly age of acquisition, typicality, frequency, and cluster structure, revealed subtle but consistent differences between MCI and control participants. These trends were also observed in the ADMCI comparison, though often with reduced statistical significance, likely reflecting the smaller sample size.

Verbosity features, such as total word count and proportion of filler words, did not significantly differ between groups.

Overall, significant differences in several of the semantic features were observed between the groups. This does not confirm predictive value but highlights these features as potentially promising.

4.3.5 Exploring the predictive value of individual features.

Work in the previous sections has identified group-level differences in cognitive test scores, imaging and a variety of semantic features using linear models. This approach primarily assess whether features differ significantly between diagnostic categories. However, significant group differences do not necessarily imply that a feature has strong discriminative power when applied to individual-level classification.

The following section primarily aims to address the following questions:

3 Does the semantic content of a participant's speech, as collected using automated verbal fluency tests, enable accurate classification between Healthy Controls and patients with MCI?

4 Is the discriminant value of these semantic features comparable to established structural imaging features?

To do so, I conducted a series of binary logistic regression models diagnosis defined as the dependent variable (coded 0 = HC, 1 = MCI). Each model tested a single predictor of interest, entered independently to avoid multicollinearity, with age and sex included as covariates to account for demographic variability. As with the previous models, continuous predictors were z-scored prior to analysis to facilitate comparability of effect sizes across features. Model performance was evaluated using odds ratios (ORs) with 95% confidence intervals, and area under the receiver operating characteristic curve (AUC) to estimate classification accuracy. P-values were adjusted using the false discovery rate (FDR) correction to account for multiple comparisons across the set of predictors. Reporting these results by feature group further supports interpretation and aligns with the structure of preceding analyses.

4.3.5.1 Assessing the predictive value of cognitive assessment and basic verbal fluency features in predicting diagnosis.

As demonstrated in Table 16, MMSE was the most powerful discriminator (OR = 0.011, AUC = 0.954, p < .001) between diagnostic categories. The number of correct words produced on phonemic and category fluency and Category Fluency Correct were also significant predictors. Features such as fluency repetitions, errors, and the semantic index did not significantly contribute to group classification after FDR correction.

Table 16 Logistic regression demonstrating predictive performance of cognitive assessment and basic verbal fluency scores.

Feature	*OR*	*CI_low*	*CI_high*	*AUC*	*p.value*	*p.adj*
MMSE	0.011	0.002	0.069	0.954	< 0.001	
Category Fluency Correct	0.358	0.207	0.620	0.820	< 0.001	
Category Fluency Repetitions	1.279	0.787	2.080	0.740	0.320	0.891
Category Fluency Errors	0.898	0.591	1.365	0.735	0.615	0.891
Phonemic Fluency Correct	0.297	0.164	0.538	0.839	< 0.001	
Phonemic Fluency Repetitions	0.845	0.550	1.299	0.733	0.444	0.891
Phonemic Fluency Errors	1.053	0.666	1.665	0.735	0.824	0.891
Semantic Index	1.032	0.658	1.619	0.731	0.891	0.891

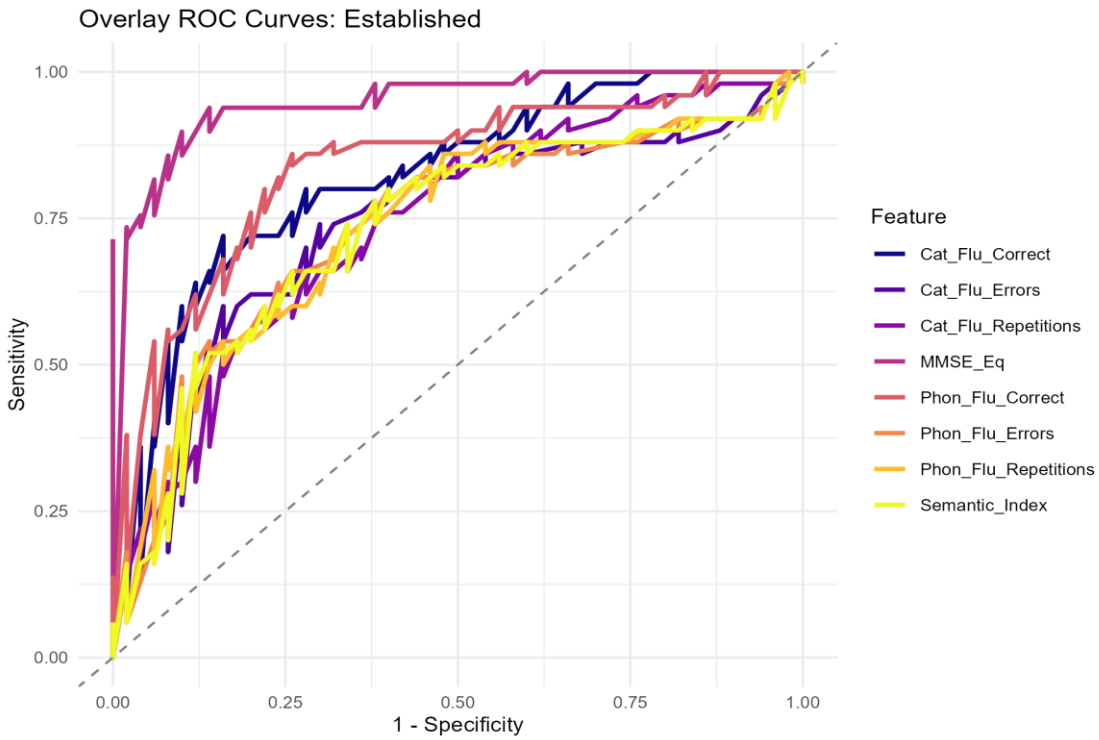


Figure 18 Overlay Receiver Operating Characteristic (ROC) Curves for Logistic Models Based on Established Cognitive assessment and basic fluency Features

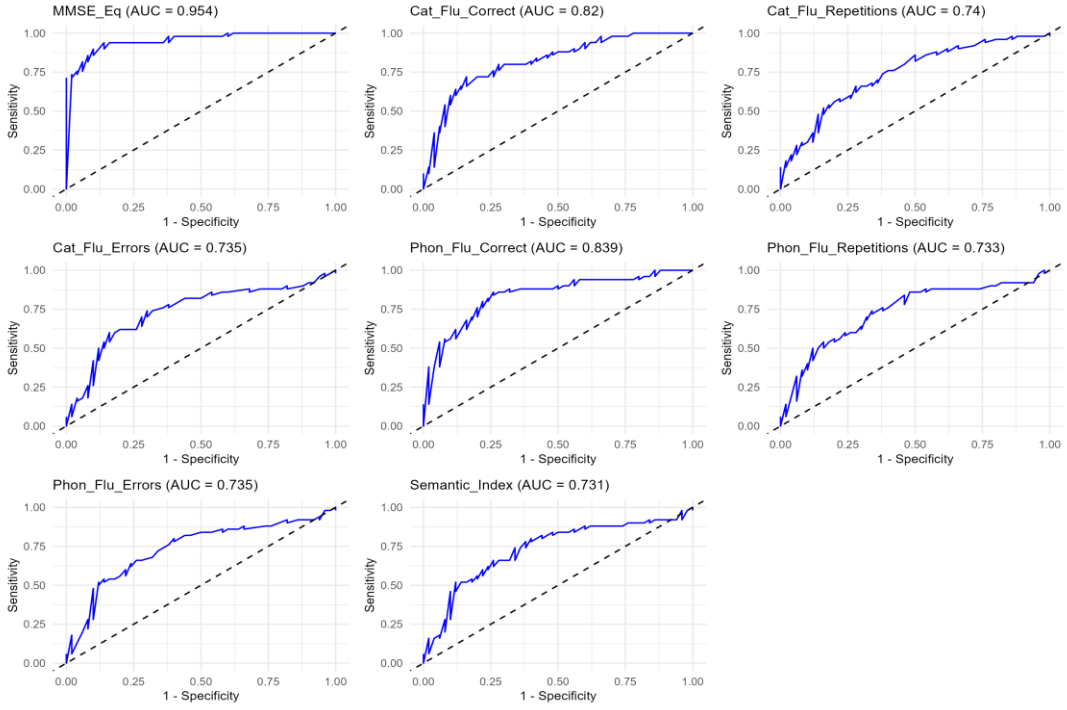


Figure 19 Individual Receiver Operating Characteristic (ROC) Curves for Logistic Models Based on Established Cognitive assessment and basic fluency Features.

Figure 18 and Figure 19 are ROC curves demonstrating the classification performance of overlayed and individual logistic regression models using conventional features: MMSE, Category Fluency (Correct, Repetitions, Errors), Phonemic Fluency (Correct, Repetitions, Errors), and the Semantic Index. Although only MMSE, phonemic and category fluency scores met the threshold for significant predictive performance, the ROC curves for the other features remain convex, albeit to a letter extent than the aforementioned significant features.

4.3.5.2 Assessing the predictive value of structural imaging and semantic features in predicting diagnosis.

Several imaging and semantic fluency features demonstrated meaningful discriminatory power in differentiating individuals with mild cognitive impairment (MCI) from healthy controls, as demonstrated in Table 17.

The right hippocampal volume ratio emerged as a significant predictor (p = 0.011, FDR-adjusted p = 0.033), achieving an AUC of 0.779, while the combined hippocampal volume ratio also reached significance (p = 0.026, FDR-adjusted p = 0.039, AUC = 0.772). The left hippocampal ratio showed a comparable AUC (0.760), but its association did not reach statistical significance after correction (p = 0.068). ROC curves are shown in Figure 20**Error! Reference source not found.**

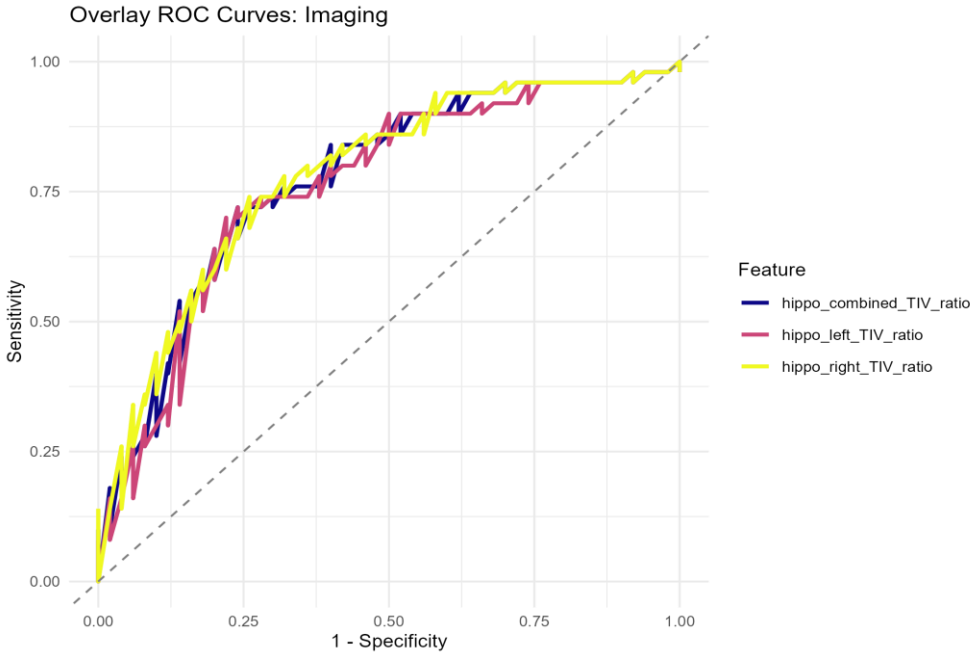


Figure 20 ROC Curves demonstrating predictive power of imaging features.

Table 17 Logistic Regression demonstrating the predictive power of imaging and semantic features (derived from verbal fluency task)

Feature	*OR*	*CI_low*	*CI_high*	*AUC*	*p.value*	*p.adj*
Left Hippocampus: TIV	< 0.001	< 0.001	> 10000	0.760	0.068	0.068
Right Hippocampus: TIV	< 0.001	< 0.001	< 0.001	0.779	0.011	0.033
Combined Hippocampal: TIV	< 0.001	< 0.001	< 0.001	0.772	0.026	0.039
Age of Acquisition	0.473	0.285	0.783	0.788	0.004	0.020
Recognition Time	0.519	0.319	0.845	0.777	0.008	0.024
Frequency	2.068	1.247	3.427	0.777	0.005	0.020
Prevalence	1.591	0.979	2.587	0.743	0.061	0.091
Typicality	1.480	0.893	2.454	0.761	0.128	0.160
Valence	1.584	1.001	2.507	0.777	0.050	0.086
Arousal	1.272	0.801	2.020	0.751	0.307	0.307
Dominance	1.278	0.814	2.005	0.757	0.286	0.307
Concreteness	2.847	0.726	11.161	0.745	0.133	0.160
Body-Object Interaction	1.839	1.077	3.138	0.772	0.026	0.052
Cluster Size	0.533	0.323	0.881	0.792	0.014	0.034
Cluster Number	0.424	0.259	0.695	0.806	< 0.001	
Cluster Number of Switches	0.455	0.274	0.756	0.787	0.002	0.020

Among semantic features derived from the category fluency task, the lexical features Age of Acquisition, Recognition Time, and Frequency emerged as significant predictors after FDR correction (adjusted p values < 0.05), with AUCs ranging from 0.777 to 0.788. Figure 21 displays the individual ROC curves for semantic features.

Cluster Size and Cluster Number also demonstrated significant predictive value (p = 0.014 and < 0.001, respectively), with Cluster Number achieving the highest AUC of 0.806 among semantic features. These clustering metrics may reflect underlying semantic network organization, which becomes disrupted in early cognitive decline. Although Prevalence, Typicality, and Valence trended toward significance, only Valence approached the corrected threshold (p = 0.050, adjusted p = 0.086), suggesting some emotional-linguistic properties may contribute diagnostically, albeit less robustly.

The ROC curves in Figure **21** demonstrate the diagnostic performance of individual logistic regression models using semantic features. All of the features demonstrate somewhat convex curves suggesting reasonable discriminant value (though bearing in mind that only age of acquisition, recognition time, frequency and the clustering variables retained significance when corrected for multiple comparisons).

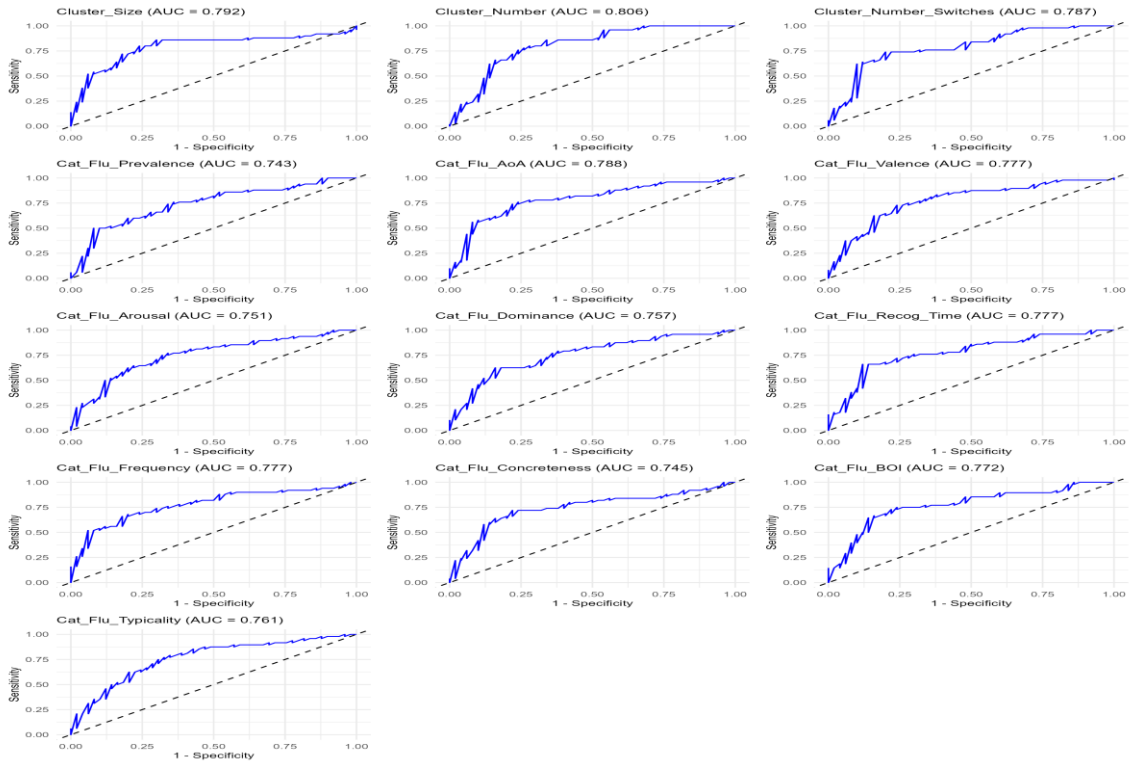


Figure 21 Receiver Operating Curves demonstrating predictive power of individual semantic features.

4.3.5.3 Summary of individual feature predictive value

These results highlight the diagnostic value of several individual features across cognitive, linguistic, and neuroimaging domains. The top performing features are summarized in Table 18.

MMSE demonstrated the strongest discriminative performance (AUC = 0.954), followed by phonemic and category fluency correct scores. Among the semantic fluency features, clustering metrics such as Cluster Number (AUC = 0.806) and Cluster Size (AUC = 0.792) were particularly informative. Lexical-semantic properties including Age of Acquisition and Recognition Time also showed significant predictive value, while affective dimensions such as Valence approached significance (with reasonable AUC = 0.77) but did not survive correction.

Table 18 Features with highest value in predicting diagnosis.

Feature	*Domain*	*Odds Ratio (OR)*	*AUC*	*FDR-adjusted p*
MMSE	Cognitive	0.011	0.954	< 0.001
Phonemic Fluency Correct	Fluency	0.297	0.839	< 0.001
Category Fluency Correct	Fluency	0.358	0.820	< 0.001
Cluster Number	Semantic Structure	0.424	0.806	< 0.001
Cluster Size	Semantic Structure	0.533	0.792	0.034
Cluster Number of Switches	Semantic Structure	0.455	0.787	0.020
Age of Acquisition	Lexical	0.473	0.788	0.020
Recognition Time	Lexical	0.519	0.777	0.024
Frequency	Lexical	2.068	0.777	0.020
Right Hippocampus: TIV	Imaging	< 0.001	0.779	0.033
Combined Hippocampus: TIV	Imaging	< 0.001	0.772	0.039

Note: This table is illustrative. Formal comparison of model strength has not been performed at this time.

In the imaging domain, right and combined hippocampal volume ratios reached statistical significance, with AUCs above 0.77.

As with previous analyses, the results for the Healthy Controls vs ADMCI group demonstrated similar trends to those observed in the HC vs. MCI comparison, without reaching the threshold for statistical significance. They are therefore not reported in detail here. Likewise, the semantic features derived from the whole interview yielded comparable AUC's but did not reach statistical significance.

In response to the questions underpinning this section:

3 Does the semantic content of a participant's speech, as collected using automated verbal fluency tests, enable accurate classification between Healthy Controls and patients with MCI?

When tested individually several of the semantic features demonstrated acceptable to good classification accuracy.

4 Is the discriminant value of these semantic features comparable to established structural imaging features?

Semantic features such as category and phonemic fluency, cluster size, number of switches and age of acquisition exceeded the AUC of the best performing structural imaging features (right hippocampal volume: total intracranial volume ratio).

While individual features provide valuable insights into early cognitive decline, clinical (and indeed all complex) decision-making usually benefits from considering multiple sources of information in synchrony. The following section explores whether the predictive power of these individual features can be enhanced through the application of multivariate logistic regression models, assessing how feature sets perform collectively in distinguishing MCI from healthy controls.

4.3.6 Exploring the predictive value of combined feature sets.

This section aims to primarily address the final and most salient question:

5 Do the language features, cognitive test scores and imaging features combined contribute to greater diagnostic accuracy than the information typically available after an initial appointment?

This is the question that most directly addresses the hypothesis that Semantic features extracted from speech will meaningfully contribute towards the identification of prodromal Alzheimer's disease. Throughout this section, I compare two predictive models: a model based on features typically available to healthcare professionals (clinician-informed model), and a data-driven model comprising the best-performing lexical and structural language features identified in earlier analyses (feature-optimized model; potentially amenable to automatic sampling).

To systemically compare the predictive performance of the competing logistic regression models, DeLong's test for two correlated ROC curves was use. This method provides a non-parametric comparison of the Area Under the Curve (AUC) values for two models fitted to the same dataset. Comparisons were conducted between:

4.3.6.1 Model Definitions

Clinician-Informed Model

This model is based on what is available to most specialist memory clinics and some general practice surgeries. Though the ubiquity of the MMSE is declining, in this instance it has been considered to be a representative multidomain cognitive assessment (though ironically one that does not employ a cognitive fluency test as a component). The MMSE was chosen because some MCI patients had not undergone the more detailed Addenbrooke's Cognitive Assessment (ACE) but had recently completed an MMSE.

Category and phonemic fluency have been included as both are commonly administered as part of a standard cognitive evaluation. Both feature in the Addenbrook's Cognitive Examination whilst the phonemic fluency test is administered in isolation as part of the Montreal Cognitive Assessment.

Whilst it is unlikely that normalised values for hippocampal volume would be available to either specialist or generalist, I have included this as a substitute for an expert report of a high-quality structural MRI. Though this is far from universal, many expert memory services in the United Kingdom would consider this a routine facet of the assessment.

Feature-Optimised Model

This model has been populated by the semantic features which demonstrated greatest discriminant value between the control and MCI groups. All of the semantic features could potentially be collected via a remotely administered category fluency test. To test if these can contribute towards the diagnosis, features that would be available to a clinician have also been included.

Table 19 Predictive Model Features

Clinician-Informed Model	*Feature-Optimized Model*
MMSE	MMSE
Category Fluency Correct	Category Fluency Correct
Phonemic Fluency Correct	Phonemic Fluency Correct
Right Hippocampus: TIV	Category Fluency Age of Acquisition
	Category Fluency Recognition Time
	Category Fluency Frequency Valence
	Cluster Number
	Cluster Number Switches
	Cluster Size
	Right Hippocampus: TIV

Statistical Analysis

To evaluate the diagnostic utility of selected feature sets, binary logistic regression models were employed with group membership as the dependent variable. All models included age and sex as covariates to control for demographic variability. Performance metrics of note include area under the ROC, sensitivity, specificity and overall accuracy of the model.

In previous analyses, the right hippocampal volume (normalized for TIV) was retained in its original scale to align with conventional anatomical reporting. However, during model development, this variable consistently produced unstable coefficients, and extreme odds ratio estimates due to its large scale relative to other predictors. To address this, a z-score normalization was applied to the hippocampal volume variable at the model-fitting stage. This standardization does not alter the underlying associations but brings the feature into a comparable range with other

predictors, thereby improving numerical stability and interpretability of the regression outputs.

4.3.6.2 Predictive strength of clinician Informed features

The clinician-informed model demonstrated good predictive power in distinguishing between the control and MCI groups with an AUC of 0.963. As shown in Figure 22, the overall accuracy of the model was 89% with a sensitivity of 90% and a specificity of 88%.

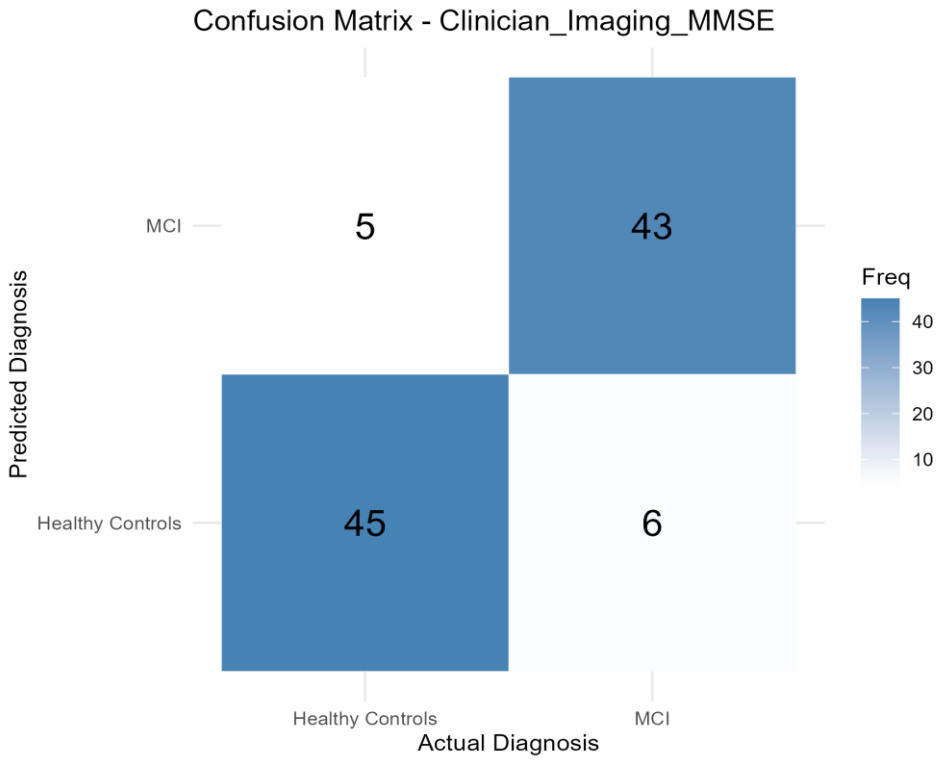


Figure 22 Confusion Matrix for Clinician-Informed Model

Table 20 Summary of Logistic Regression Model Using Clinician-Informed Features

Feature	*Estimate*	*Std. Error*	*Odds Ratio*	*CI Lower*	*CI Upper*	*AUC*
(Intercept)	8.54	3.62	5,130.00	4.28	>1e6	0.963
Age	-0.12	0.05	0.89	0.81	0.98	0.963
Sex	1.32	0.78	3.73	0.81	17.2	0.963
MMSE	-5.00	1.13	0.01	0.00	0.06	0.963
Category Fluency Correct	-0.83	0.45	0.44	0.18	1.06	0.963
Phonemic Fluency Correct	0.54	0.51	1.71	0.63	4.62	0.963
Normalized Hippocampus: TIV	0.06	0.48	1.06	0.41	2.7	0.963

As Table 20 demonstrates, the MMSE was the strongest predictor in the model, with a large negative coefficient ($\beta = -5.00$, $p < .001$) and an odds ratio (OR) of 0.01, indicating that for each unit increase in MMSE score, the odds of being classified as MCI were dramatically reduced.

Although only included as a covariate, age also contributed significantly ($\beta = -0.12$, OR = 0.89), suggesting that increasing age was associated with higher likelihood of MCI classification.

Category fluency score had a negative association ($\beta = -0.83$, OR = 0.44), implying that better performance was associated with reduced odds of MCI. Somewhat surprisingly, phonemic fluency score had a positive coefficient ($\beta = 0.54$, OR = 1.71) suggesting an opposite effect. However, the wide confidence interval (CI: 0.63–4.62) prompts caution regarding interpretation.

The z-scored right hippocampal volume ($\beta = 0.06$, OR = 1.06) did not emerge as a significant predictor in this model (CI: 0.41–2.70), suggesting its contribution may be minimal when MMSE and fluency measures are included. Overall, the model highlights the dominant role of MMSE, while also suggesting potential contributions from verbal fluency measures.

4.3.6.3 Predictive strength of feature-optimized model

The feature-optimized demonstrated excellent performance in distinguishing between groups, with an AUC of 0.972. As shown in Figure 23 the overall accuracy was 93%, with a sensitivity of 92% and a specificity of 94%. Only 3 patients with MCI were misallocated into the control group, compared to 6 with the clinician informed model.

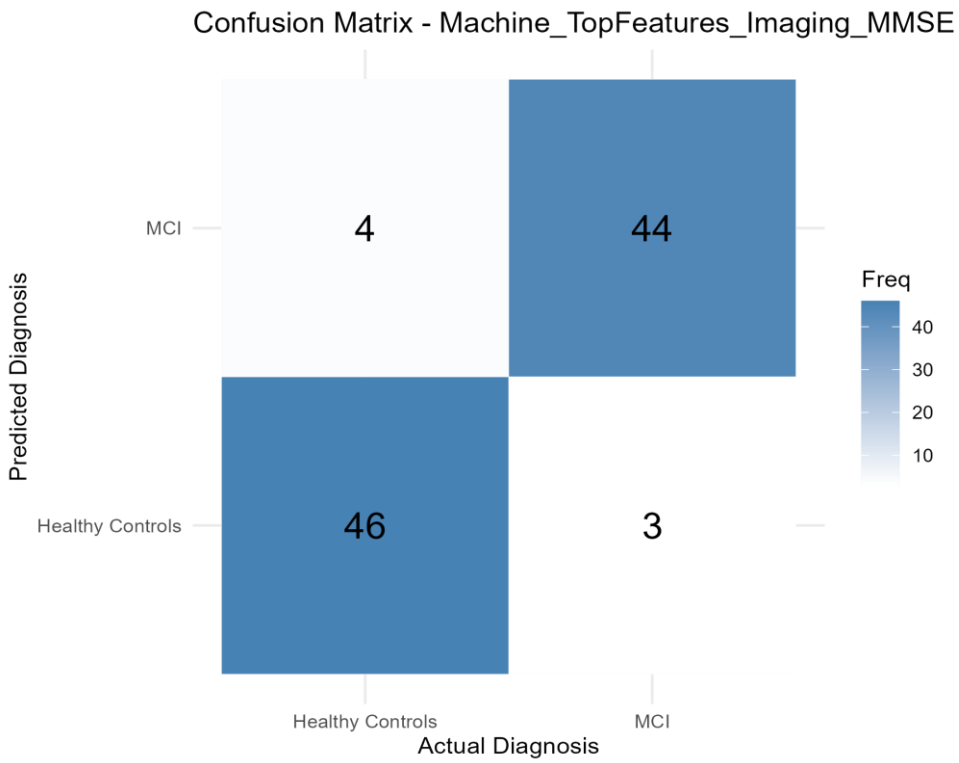


Figure 23 Confusion matrix for features-optimized model.

Table 21 Summary of Logistic Regression Model using Feature-Optimized Model

Feature	*Estimate*	*Std. Error*	*Odds Ratio*	*CI Lower*	*CI Upper*	*AUC*
(Intercept)	16.30	6.74	>1e6	21.80	>1e6	0.972
Age	-0.23	0.09	0.8	0.67	0.96	0.972
Sex	1.02	0.99	2.78	0.40	19.5	0.972
MMSE	-5.61	1.40	0	0.00	0.06	0.972
Category Fluency Correct	-1.86	2.41	0.16	0.00	17.5	0.972
Phonemic Fluency Correct	0.28	0.61	1.32	0.40	4.36	0.972
Category Age of Acquisition	1.02	1.44	2.76	0.16	46.3	0.972
Category Fluency Recognition Time	0.33	0.96	1.39	0.21	9.08	0.972
Category Fluency Frequency	0.46	1.40	1.58	0.10	24.8	0.972
Category Fluency Valence	0.28	0.57	1.33	0.43	4.09	0.972
Cluster Number	1.20	1.64	3.33	0.13	83.7	0.972
Cluster Number of Switches	-0.39	1.14	0.68	0.07	6.39	0.972
Cluster Size	-0.46	1.06	0.63	0.08	5	0.972
Normalized Right Hippocampus: TIV	-0.31	0.62	0.74	0.22	2.48	0.972

The performance of the individual features within this model is summarized in Table 21. As expected, the MMSE score emerged as the most powerful individual predictor ($\beta = -5.61$, OR ≈ 0.00), reflecting a strong inverse association with MCI diagnosis; higher MMSE scores were strongly associated with lower odds of being classified as MCI.

Other variables, including fluency and semantic features, did not reach statistical significance but showed varying degrees of association. For example, Category Fluency Correct was associated with reduced odds of MCI (OR = 0.16), but the large standard error and wide confidence interval (0.00–17.5) indicate low precision in this estimate. Similarly, age of acquisition (OR = 2.76), frequency (OR = 1.58), valence (OR = 1.33) and normalized right hippocampal volume (β = -0.31, OR = 0.74) did not significantly predict MCI status (CI: 0.22–2.48), all showed positive but non-significant associations with MCI status, suggesting these features contribute only modestly when MMSE is already included in the model.

Overall, the results again emphasize the dominance of MMSE in distinguishing diagnostic groups, with other features offering potential but largely non-significant incremental value in the presence of MMSE.

4.3.6.4 Summary of the competing performance of the clinician informed and feature-optimized models.

To objectively evaluate whether the machine-based model outperformed the clinician-informed model in distinguishing between healthy controls (HC) and individuals with mild cognitive impairment (MCI), a DeLong's test was conducted to compare their respective areas under the curve (AUC). The clinician-informed model, which included MMSE, fluency, and imaging features, yielded an AUC of 0.963. The machine-based model, which included MMSE along with a set of data-driven, top-performing semantic and imaging features, achieved a higher AUC of 0.972. This difference was not found to be statistically significant, (t (188.19) = –0.437, p = 0.663) indicating that whilst the features-model appeared to improve accuracy, this boost did not meet the threshold for statistical significance.

I would regard the trend toward improved classification performance with the feature-optimized model to be in line with the findings from earlier sections. To address

5. Do the language features, cognitive test scores and imaging features combined contribute to greater diagnostic accuracy than the information typically available after an initial appointment?

Accepting that the results do not meet the threshold for significance, I would suggest that the trend suggests potential benefits from the feature-optimized model and invites further, more highly powered study.

4.3.7 Exploring the potential benefits of machine learning models.

Whilst the choice of features for the feature-optimized model above was informed by feature performance in earlier stages, it is important to acknowledge the myriad of complex, potentially nonlinear interactions between linguistic, cognitive, and neuroanatomical features. Though I spent time experimenting with various models, the features-optimized model performed best and was composed in a rational and evidence informed fashion. Exhaustive re-configurations of the feature-drive model may yield a more satisfying confusion matrix but risks degrading the generalizability of the model through over-fitting.

Of note, I did experiment with a refined feature-optimized model after testing for collinearity amongst the features by generating variance-inflation metrics and a heatmap of correlations. This demonstrated a high level of collinearity between age of acquisition and frequency (corresponding to an intuitive understanding of these features). However, running a refined/ tuned variation of the model did not yield increased classification accuracy.

This section aims to expand the answer to:

5. Do the language features, cognitive test scores and imaging features combined contribute to greater diagnostic accuracy than the information typically available after an initial appointment?

Unlike traditional regression models, machine learning approaches can flexibly combine diverse features and identify subtle patterns that may not be apparent through linear associations. To answer this question accurately, alternative methods of classification (using the same features) must be explored.

A random forest model was chosen on the basis that they are robust against non-linear data, collinearity, over-fitting and handle complex interactions well. They have the added benefit over other machine learning approaches in that they are amenable to interpretation (producing a meaningful output of feature importance alongside classification)(Dimitriadis et al., 2018; Sapir-Pichhadze & Kaplan, 2020).

The random forest used the top performing features included in the aforementioned logistic regression model and was trained and evaluated using cross-validation and performance metrics such as accuracy, sensitivity, specificity, and area under the ROC curve (AUC).

4.3.7.1 Random Forest Classification Using High-Performing Features Model Description

The model was trained using 10-fold cross-validation to ensure generalizability, and performance was evaluated in terms of area under the receiver operating characteristic curve (AUC), classification accuracy, sensitivity (true positive rate for MCI), and specificity (true negative rate for HC).

It is particularly robust to overfitting, capable of capturing nonlinear relationships, and allows estimation of feature importance based on how much each variable contributes to improving classification purity (Mean Decrease in Gini Impurity).

The model achieved an AUC of 0.89, indicating a good ability to discriminate between groups. The overall classification accuracy was 87%, which represents a substantial improvement over chance performance. The model demonstrated a sensitivity of 88% and a specificity of 86%. The classification accuracy is demonstrated in Figure 24.

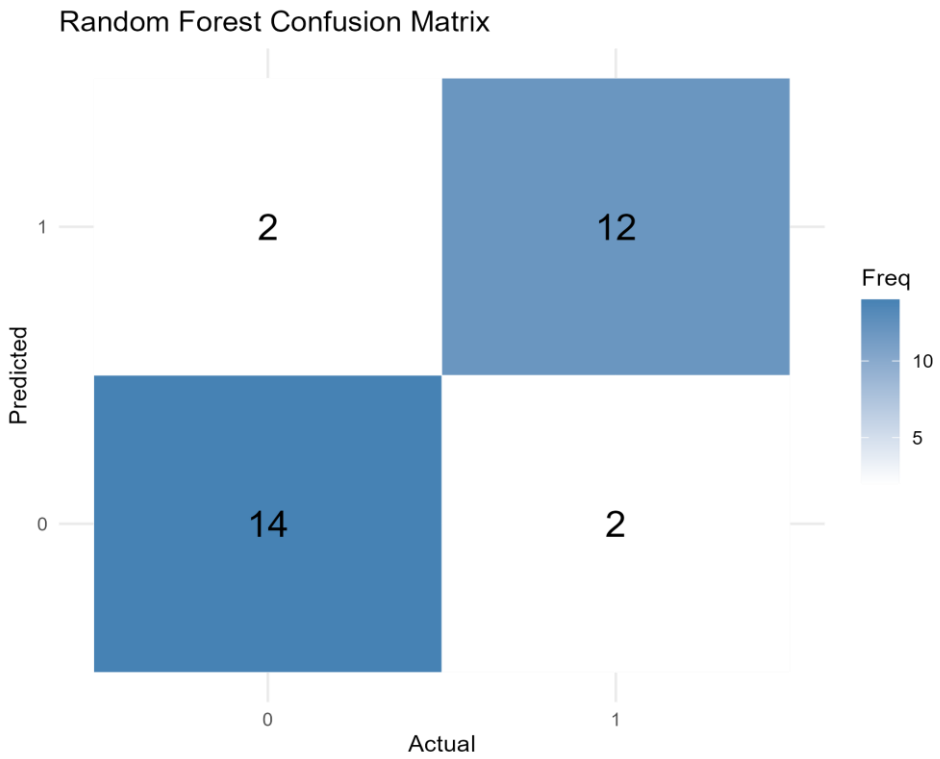


Figure 24 Confusion Matrix demonstrating Random Forest Classification Performance

Table 22 Random Forest Variable Importance

Feature	MCI	Controls	Mean Decrease Gini
MMSE	25.8243694	24.2537264	11.575316
Frequency	0.7825945	5.2538669	2.908779
Normalized Right Hippo:TIV	1.1281134	0.2007284	2.752121
Cluster Number of Switches	8.9212366	1.0065421	2.631046
Phonemic Fluency	0.9079976	0.7483841	2.457955
Age of Acquisition	1.1801312	2.8355825	2.282820
Valence	0.2920271	2.1800669	2.174520
Category Fluency	3.6689992	1.8017177	1.933331
Recognition	-1.6852800	1.7222689	1.775427
Cluster Size	-1.2793809	-1.1066739	1.256050
Cluster Number	-0.2077367	3.7413349	1.252756

Note: Higher figure indicate higher importance in the classification. For group variables, this indicates how important a feature is in accurate classification of that group.

Predictably MMSE emerged as the most important variable in classification. The right hippocampus: TIV ration contributed significantly to the overall accuracy of the model. Frequency was the most important semantic variable followed by the number of cluster switches. All of the variables contributed towards accurate classification of MCI except for recognition time, cluster size and cluster number. Inclusion of cluster size was also deleterious the accuracy of control prediction.

The performance demonstrates the potential utility of incorporating machine-derived lexical and structural features into diagnostic tools. Future work may explore further optimization of the feature set and integration with clinical workflows to enhance diagnostic precision.

4.3.7.1 Comparing the performance of the logistic regression and random forest models

To evaluate whether a machine learning approach could outperform traditional statistical methods in classifying between healthy controls (HC) and individuals with mild cognitive impairment (MCI), I directly compared the performance of the Random

Forest model to that of a logistic regression model using the same set of high-performing features.

Table 23 Comparing performance between logistic regression and random forest model.

	Logistic Regression	***Random Forest***
AUC	0.96	0.89
Accuracy	0.93	0.87
Sensitivity (MCI)	0.90	0.88
Specificity (HC)	0.88	0.86

The logistic regression model demonstrated superior overall classification performance, achieving an accuracy of 93% and an AUC of 0.96, indicating excellent discriminatory ability. In contrast, the Random Forest model yielded a slightly more modest performance, with an accuracy of 87% and AUC of 0.89.

Sensitivity and specificity were slightly better for the logistic regression model. Of note, I later decided to run the model again to perform a DeLong test. On this iteration of the random forest model the AUC was 0.95 and there were no significant differences between the performance of the two.

This demonstrates that the two models are comparable in terms of accuracy, sensitivity and specificity. The logistic regression appears to predict diagnosis with slightly greater accuracy, though this was not statistically significant.

4.3.7.2 Exploratory random forest model comparison

To further explore the potential benefits of data-driven approach I implemented a series of Random Forest (RF) classification models. Feature sets were derived from previously defined superordinate groups, including clinician-accessible scores (e.g., MMSE, verbal fluency), affective semantic features (e.g., valence, arousal), structural semantic metrics (e.g., clustering metrics), imaging measures (e.g., hippocampal volume), serial recall order metrics, and whole-interview features. Random Forest models were chosen for their ability to model complex nonlinear relationships, resistance to overfitting, and interpretability via variable importance. The design, cross validation and model performance metrics remain unchanged from the previous model.

Comparative ROC curves (see Figure 25) were plotted to permit visual comparison of the various models.

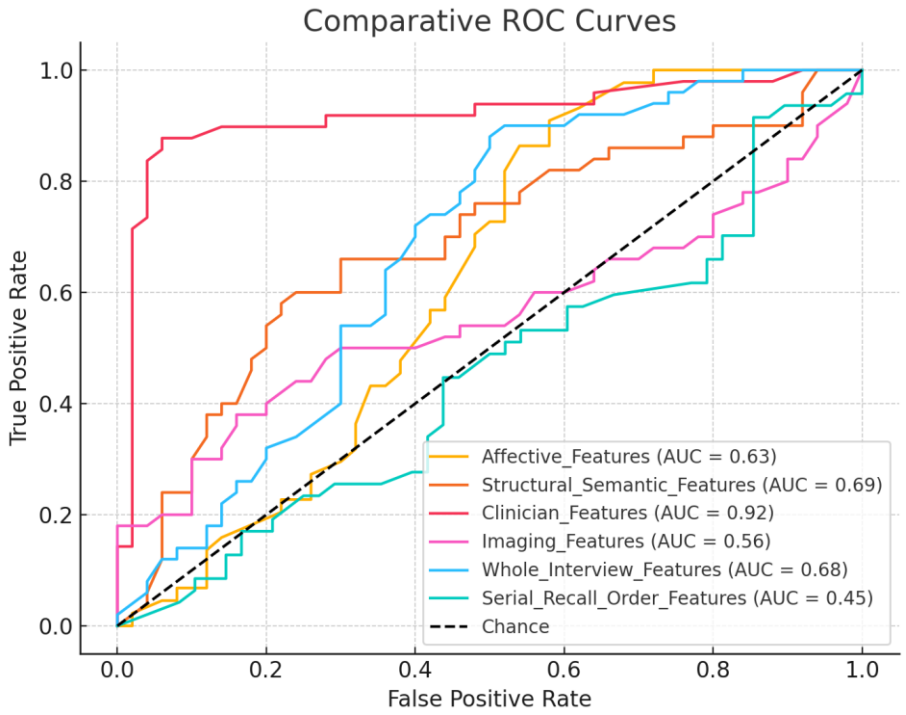


Figure 25 ROC Comparative Random Forest Models

Of the various feature-set models, the clinician-informed features demonstrated superior predictive power. Again, this is likely to reflect the overwhelming effect of the MMSE (this was the only group in which it was included).The performance of the rest of the models, exploring the potential utility of the whole interview features, affective features and imaging features (in isolation), cluster-based features (in isolation) did not match the accuracy of the feature-optimized version. The serial recall feature were observed to perform worse than might be expected by chance.

4.3.7.3 Machine learning models summary

I tested the predictive performance of a random forest machine learning classification model compared to the logistic regression model previously used. This did not improve classification accuracy when provided with the feature-optimized variables. Likewise, testing a range of other variables sets did not identify a group with classification accuracy comparable to the feature-optimized model.

5. Do the language features, cognitive test scores and imaging features combined contribute to greater diagnostic accuracy than the information typically available after an initial appointment?

The feature-optimized model consisting of a number of sematic features appears to contribute towards diagnostic accuracy. Applying this feature set with a machine learning classification system did not impact positively on its performance.

4.3.8. Results Summary

Throughout this section I have explored the characteristics and diagnostic value of a selection of imaging and semantic features (the latter obtained through participant interaction with a virtual clinician).

The analyses have been directed towards exploring the hypothesis:

- Semantic features extracted from speech will meaningfully contribute towards the identification of prodromal Alzheimer's disease.

To summarise how the results have informed the validity of this hypothesis I will return to the questions originally posed at the beginning of the section.

1. Do the linguistic features' relationship to one another and selected regions of interest on structural imaging support their use as measures of semantic memory and processes?

Voxel-Based Morphometry based analysis demonstrated significant between group differences. However, differences that might be expected in prodromal Alzheimer's were apparent only when the confirmed AD MCI group was assessed. Persistent, significant differences in the volume of the basal ganglia raise questions as to the reliability of the data.

Hippocampal volumes correlated with correct category fluency responses and serial recall order typicality. Whilst the rest of the correlations between semantic features and hippocampal volume did not reach significance, the trends reflected expectations based on established literature. The answer to this question is **possibly**. These trends invite further exploration.

There was strong internal correlation between the semantic features. This largely reflected the different categories of semantic test (e.g. lexical, affective).

2. Does the analysis of semantic function applied to participants' responses to standardized interview questions (administered using the same automated system) resemble the analysis of the verbal fluency tasks?

Values for age of acquisition and body-object interaction derived from the whole interview and the verbal fluency task correlated significantly. Whilst the strength of the correlation between the rest of the features did not meet statistical significance, the trends appeared to be identical, suggesting that with larger samples, more significant findings may be observed. When the whole interview features were included in a random forest classification model their predictive performance was poor (AUC 0.68). The answer to this question is **possibly**. The trend for these features appeared to closely resemble those for the category fluency test. Given the

potential strengths of being able to extract semantic features such as these from free speech further investigation is warranted.

In line with Hypothesis 1, semantic features derived from verbal fluency tasks showed significant differences between HC and MCI groups. Several semantic variables—including age of acquisition, valence, word frequency, and clustering metrics—were significantly associated with diagnosis and displayed predictive value, both independently and when used in combination with traditional cognitive tests.

3 Does the semantic content of a participant's speech, as collected using automated verbal fluency tests, enable accurate classification between Healthy Controls and patients with MCI?

When tested individually several of the semantic features demonstrated acceptable to good classification accuracy, these included the category fluency score, number of clusters, cluster size, number of cluster switches, age of acquisition recognition time and frequency (with valence approaching, but not meeting, significance). This suggests that **yes**, semantic features extracted from a participant's speech (specifically the responses to the category fluency test) enable accurate classification between control and MCI groups.

4 Is the discriminant value of these semantic features comparable to established structural imaging features?

Semantic features such as category and phonemic fluency, cluster size, number of switches and age of acquisition exceeded the AUC of the best performing structural imaging features (right hippocampal volume: total intracranial volume ratio). The right hippocampal volume to total intracranial volume ratio performed best (contrary to expectations) and was included in the multivariate logistic regression models and random forest classifiers. It contributed significantly to the prediction accuracy of the random forest model but was less impactful in the logistic regression. On balance, **yes,** the semantic discriminant value of the semantic features and structural imaging features was comparable.

5. Do the language features, cognitive test scores and imaging features combined contribute to greater diagnostic accuracy than the information typically available after an initial appointment?

The diagnostic performance of two competing models was compared. One was populated with features that would be available to a clinician in memory clinic, and one was populated with additional features (chosen based on their performance on individual logistic regression models). This latter, feature-optimised, model appeared to outperform the clinician-informed model, with an overall accuracy of 93%

compared to 88%. Accepting the failure to meet the threshold for significance, I would suggest that the trend is promising and invites further investigation.

A machine learning classifier was trailed to explore whether further improvement in accuracy could be achieved. A random forest model using the same feature-optimized feature set achieved comparable accuracy to the logistic regression approach. It is important to note that comparing the performance of the clinician informed model and feature-optimized model did not demonstrate a statistically significant difference. However, based on the performance of individual features on and the positive trend observed, I feel that the answer to this is justifiably **yes**.

4.4 Advanced Semantic Feature Analysis for MCI Detection: Discussion

4.4.1 Introduction to Discussion

This section discusses the findings of the advanced semantic feature analysis work in relation to the hypothesis:

- Semantic features extracted from speech will meaningfully contribute towards the identification of prodromal Alzheimer's disease.

This discussion is organized to complement the structure of the results section and aims to address key trends, methodological considerations, theoretical implications, and the relevance of each finding to the broader literature. I will first discuss the imaging and language features in more detail. In doing so, this chapter aims to interpret how the results inform our understanding of the role of language derived features in diagnostic classification.

4.4.2 Methodological Considerations surrounding the composition of the MCI group.

As discussed in the methodology section, of the 50 patients with MCI, only 31 were ultimately confirmed to have Alzheimer's disease. All of the patients were categorized as amnestic MCI at the time of recruitment to the study. Of the 50 participants, 12 underwent CSF biomarker testing. This confirmed Alzheimer's disease in every case. 19 further patients were confirmed to progress in a manner consistent with Alzheimer's disease at follow up, confirming the diagnosis. Table 24 summarizes the diagnoses of all the MCI participants. This is accurate at the time of writing.

Table 24 Diagnosis at time of writing

Diagnosis	
Alzheimer's Disease	31
Mild Cognitive Impairment	10
Lewy Body Disease	3
Frontotemporal Dementia	2
Corticobasal Degeneration	1
Progressive Supranuclear Palsy	2
Parkinson's Disease Dementia	1

Cohort heterogeneity of this nature introduces issues with the interpretation of the results. All of the analyses discussed in the previous section were applied to both the overall MCI group and the confirmed ADMCI cohort. The patterns of behavior in the features were observed to be identical, with the only distinction being that none of the features, aside from MMSE were found to be statistically different between the ADMCI and control groups. Given the preservation of the same patterns, but loss of significance, I attribute this more to loss of sample size than reflecting a change in effect between the groups.

The fact that the same patterns were observed despite the addition of the 19 other patients suggests that either the effect of the variables is very powerful (this is not particularly supported by the data, none of the variables approached the predictive power of the MMSE for example) or that of the 10 patients with amnestic MCI a significant proportion have Alzheimer's disease (but have neither undergone biomarker testing nor have progressed sufficiently over the interval since study enrolment to have met the diagnostic criteria).

The NIA-AA research framework proposes that the diagnosis of Alzheimer’s disease is made on the basis of biological criteria (ATN) rather than syndromic features (Jack et al., 2017). There has been an increasing move towards the inclusion of imaging, CSF and serum biomarkers in clinical trials to, particularly therapeutic trails where, in addition to improving confidence in the homogeneity of the study cohorts, they have the added benefit of demonstrating target engagement (Cummings, 2019).

The participants in this study were recruited from the specialist memory clinic and engaged with the virtual clinician remotely. All of their investigations aside from this, including the structural MRI, formed part of their routine clinical assessment. I consider this to be an attractive aspect of the study design, in that it is non-invasive and reflects real world practice.

We have historically been slightly reticent about pursuing a biological diagnosis. Locally, the only option available in clinical practice is CSF testing for Aβ42/40 and phosphorylated tau at present. The logistical challenges of arranging a lumbar puncture on patients with cognitive complaints, in the absence of any meaningful treatment have tended to limit enthusiasm. Recent years have seen the emergence of effective blood-based biomarkers (Mantellatto Grigoli et al., 2024) with phosphorylated tau-based markers such p-Tau181 and p-Tau217 particularly demonstrating strong correlations between CSF and imaging based biomarkers and exhibiting superior accuracy in distinguishing prodromal Alzheimer’s disease from MCI of other aetiology and control groups (Dasari et al., 2025; Lai et al., 2024). These serum based approaches promise to be accurate, cost effective and accessible (Chohan et al., 2022b; Mattke et al., 2020).

Future iterations of this study will ensure that a biological diagnosis of prodromal Alzheimer’s disease features in the inclusion criteria. The most practical means of

achieving this will be the addition of a blood-based biomarker such as p-Tau217 to the standard clinical procedure.

4.4.3 Voxel Based Morphometry

The Voxel Based Morphometry study between the control and MCI groups did not demonstrate areas of significant difference that would be considered typical on reviewing the extant literature. They are reported as they reflect the output of a robust analysis process. Given the variance from findings reported elsewhere I reflect below on potential causes for this.

VBM studies have consistently demonstrated significant differences between control and Alzheimer's disease groups (even those early in the disease course. Atrophy is observed in the hippocampus, amygdala, insula, lateral temporal and cingulate cortex, precuneus and caudate nuclear. Anterior cingulate atrophy may be particularly observed early in the disease course (Frisoni et al., 2002). Progressive changes in patterns of atrophy correspond to the Braak stages of neurofibrillary tangle deposition with the very earliest changes being observed in the medial temporal lobe structures (Matsuda, 2013). Smaller hippocampal volume, particularly within subfield CA1 predict conversion from MCI to AD (Apostolova et al., 2006). Healthy controls typically have right-larger-than-left asymmetry, which may lessen in the early stages of the disease course (reflecting right sided atrophy) (Barnes et al., 2005). The left hippocampus demonstrates greater involvement in semantic processing, whilst the right is associated with visual and spatial memory retrieval (Dalton et al., 2016).

The most marked area of atrophy demonstrated by the VBM study was in the right putamen. On close inspection of the imaging, it appears to overlap with the right insula cortex.

Some studies have reported decreased putamen volumes in Alzheimer's disease, with some asymmetry observed. (De Jong et al., 2008). Right insular atrophy has been observed as an early feature of Alzheimer's disease (Fathy et al., 2020). However, these changes were observed alongside a pattern of atrophy more typical for Alzheimer's disease, with hippocampal atrophy being reported as a consistent early feature.

On comparing the ADMCI and control groups, areas of decreased volume more in line with expectations emerged. Atrophy in areas including the left parahippocampal gyrus, right posterior cingulate, left temporal lobe, left parietal lobe and the left caudate have all been reported in Alzheimer's disease (Frisoni et al., 2002). The right putamen atrophy remained the largest cluster.

To determine if the putamen finding was artefactual or due to methodological issues the raw, segmented and smoothed images were re-reviewed. Registration appeared

to be generally acceptable. There were no gross lateralizing abnormalities that might skew the study. On reviewing the raw images, the tissue contrasts appeared markedly different between some of the scans. This may represent a source of artificial group difference, raising spurious areas of apparent atrophy whilst obscuring subtle but real changes that reflect early disease. Statistical harmonization techniques such as ComBat can be employed to reduce the impact of inter-scanner variability (Radua et al., 2020), this has been reported to ameliorate this effect in multisite studies (Bonthrone et al., 2025).

Although I considered whether the putamen atrophy may reflect effect of the patients with Lewy Body Disease, Corticobasal degeneration syndrome and Parkinson's disease dementia (Cui et al., 2020) (Messina et al., 2011). However, the persistence of the observation after removing these cases from the analysis would suggest strongly against that.

Cerebrovascular risk factors such as smoking, uncontrolled hypertension and hypercholesterolemia are associated with damage to the deep grey matter structures as well as increased risk of Alzheimer's disease (Knopman & Roberts, 2010). However, if vascular disease were responsible for the changes I would anticipate bilateral involvement.

Collecting information on vascular risk factors and including this as a covariate for the T-test may enable more accurate interpretation. I will also consider the use of a statistical harmonization technique such as ComBat to mitigate against the impact of varying scan properties.

I will also consider supplementing this analysis with a surface based morphometric technique such as FreeSurfer. Though the performance of FreeSurfer and VBM is reported to be comparable, the methods applied are sufficiently distinct that cross-validation of findings would aid interpretation and increase confidence (Goto et al., 2022). This has not been performed at this time due to the significant time expense associated with the FreeSurfer preprocessing protocols.

Conversely, the comparative advantage of using FreeSurfer is the ability to completely automate the (admittedly slow) imaging processing pipelines and segment and extract regions of interest on an individual level. This would enable imaging features of highly discriminant value (such as those listed above) to be extracted and included in a classification model.

4.4.4 Hippocampal Volumes

The hippocampal volumes were significantly smaller between the control and MCI groups. The right appeared to be disproportionately affected compared to the left. This is considered consistent with the available literature (Barnes et al., 2005).

Serial recall order typicality was found to correlate significantly, albeit weakly, both hippocampal: total intracranial volume ratios, with the left being marginally stronger than the right. Serial recall order has been reported to be positively associated with the left entorhinal cortex.

Despite the lack of significant correlations, the patterns observed between the features and hippocampal volumes were largely as one might expect. Positive correlations were observed between hippocampal volume and cluster related features, age of acquisition and recognition time.

The trends were consistent with expectations and support the putative connection between these imaging and semantic features, albeit tentatively and pending confirmation with further evaluation.

Semi and fully automated techniques for the reporting of structural imaging in Alzheimer's disease have been developed, with varying results reported. Some models have achieved levels of accuracy comparable to or exceeding human reporters (Klöppel et al., 2008) whilst others have reported relatively poor performance (Akazawa et al., 2019). One study by Tanabe et al. assessed the impact on classification accuracy of providing a volumetric report to neuroradiologists to aid in reporting. The group reported no significant improvement in performance, with a post-hoc analysis suggesting that the report may have biased readers towards a diagnosis of neurodegeneration in cognitively healthy adults (Tanabe et al., 2024).

The average left and right hippocampal volumes in our control group were 2.56 cm^3 and 2.63cm^3 respectively. Estimates on the normal range vary, with hippocampal volumes exceeding 3cm^3 being considered definitely normal and volumes less than 2cm^3 being considered definitely abnormal. Boccardi et al. reported mean hippocampal volumes of 2.44 cm^3 in controls; 1.99 cm^3 in patients with amnestic MCI and 1.80cm^3 in patients with Alzheimer's disease (mean age of all groups 76) (Boccardi et al., 2015; Csernansky et al., 2005). It is plausible that the same figure can be interpreted quite differently depending on the accuracy of the reporters estimates for normality.

The individual logistic classification model found that the right hippocampal value performed fairly in predicting MCI diagnosis, with an AUC of 0.76. On this basis it was included in the feature-optimized analysis. To ensure that semi-automatic or automatic diagnostic systems are trusted, the reporting must be transparent and interpretable. In future work exploring the integration of the semantic and imaging features into an automated classification system, this will need to be carefully considered. At a minimum I would suggest providing descriptive data with robust

estimates on the parameters of normality as well as clear and meaningful information on how the imaging features contributed towards the classification.

4.4.5 Language Based Features

The language features included in this study were selected on the basis that they were:-

- Related to various facets of semantic function.
- Available from the response to a category fluency task
- Amenable to automatic/ semi-automatic processing
- Interpretable, with a tangible quality that can assist in clinical decision-making relationship.

As discussed in the introduction section, semantic impairment is an early, consistent and measurable feature of Alzheimer's disease (Snowdon & Nun Study, 2003).

The category fluency task is an attractive medium to collect language-based data for a number of reasons:-

- Robust- Consistent and significant deficits in task performance are identified early in the disease course when compared to controls (Auriacombe et al., 2006; Cerhan et al., 2002; McDonnell et al., 2020),
- Scalable- This pattern is preserved across languages (Monsch et al., 1997b)
- Cheap- There is no copyright applied to the category fluency test.
- Quick- In its simplest incarnation it is administered over 60 seconds.
- Amenable to remote and automated administration (O'Malley et al., 2020)

There has been increasing interest in the automated analysis of the verbal fluency tasks (both category and phonemic). In part this is likely to reflect the broad appeal of the aforementioned features. However, automatic analysis of this task also holds the promise of mitigating against some of the issues inherent in the interpretation of the task in its traditional form. Though it is often interpreted as an index of semantic integrity, overall performance is likely to be related to a number of supporting processes, including encoding, retrieval, psychomotor speed and semantic control (rather than semantic knowledge)(De Marco et al., 2023). Without elaborating on the analysis, it is difficult to be certain what is being measured.

The work of Gonzalez-Recober et al and Cho et al has confidently established the feasibility of applying completely automated approaches to the extraction of a number of additional features, including age of acquisition, ambiguity, concreteness, frequency, word length and semantic distance from the preceding word) from the verbal fluency tasks (Cho et al., 2021; Gonzalez-Recober et al., 2023). Perez et al have assessed the performance of lexical/ semantic features and acoustic features in the accurate identification of patients with MCI compared to controls. They found that classification accuracy was superior when based on word properties such as

frequency (AUC = 0.72) compared to temporal features (AUC = 0.63). They also demonstrated neuroanatomical correlates to these features, with atrophy in the left temporal regions (Pérez et al., 2024)

Table 25 Expected and observed behavior of language-based features in MCI.

Feature	*Category*	*Expected behaviour in Prodromal Alzheimer's Disease*	*Observed behaviour in MCI cohort*
Category Fluency Score	Semantic (Cerhan et al., 2002)	Reduced	Reduced*
Category Fluency Repetitions	Semantic	Increased	Increased
Category Fluency Errors	Semantic (Pekkala et al., 2008)	Increased	Reduced
Age of acquisition	Semantic (Forbes-McKay et al., 2005)	Reduced	Reduced*
Typicality	Semantic	Increased	Increased
Frequency	Semantic	Increased	Increased*
Prevalence	Semantic	Increased	Increased
Recognition Time	Semantic	Reduced	Reduced
Cluster Size	Semantic/ Executive	Reduced	Reduced*
Cluster Number	Semantic/ Executive	Reduced	Reduced*
Number of switches	Semantic/ Executive (Troyer, Moscovitch, Winocur, Leach, et al., 1998b)	Reduced	Reduced*
Dominance	Affective	Increased	Increased
Valence	Affective	Increased	Increased
Arousal	Affective (Paek, 2021b)	Increased	Increased
Body-Object Interaction	Semantic/ Sensorimotor (Siakaluk et al., 2008)	Increased	Increased*
Concreteness	Conceptual (Giffard et al., 2015)	Increased	Increased

*Note: * Indicates that the difference between groups was statistically significant.*

The language-based features demonstrated a high level of inter-feature correlation, with strongest correlations observed between features of the same category.

Whilst not all of the between groups met the threshold for statistical significance, all of the features (excepting category fluency errors) behaved as might be anticipated based on the extant literature.

The predictive value of the language-based features, particularly age of acquisition and those related to clustering, was reflected in their robust AUC values when examined on an individual level with logistic regression. In both the feature-optimized logistic classifier and the random forest model the cluster related metrics and age of acquisition contributed significantly towards the classification accuracy.

The features did not appear to have the same predictive value when derived from the whole interview, though continued to demonstrate similar trends. This, therefore, cannot lend support for the application of the features (extracted from free speech) in diagnostic classification at this time.

The strength of these features in accurately predicting diagnosis, combined with their ubiquity in medical practice, ease of administration and clear interpretability (with a robust supporting literature and clear neuroanatomical correlates) make the semantic and semantic/executive features particularly attractive for incorporation into a broader automated cognitive assessment system.

4.4.6 Machine Learning Classifiers

I acknowledge a lack of expertise in machine learning techniques. The machine learning aspect of my experiment and results section was permitted by the intuitive packages available in r and generative AI assistance in debugging code errors. Machine learning is increasingly ubiquitous in both the clinical and research domains. There is a pervading sense that “machine learning” in a general sense can be applied unselectively to any problem to improve outcomes (or prospects of publication). This introduction to the subject has been a revelation, to both the capabilities and limitations associated with this emergent set of technologies.

Random Forest (RF) is a machine learning method increasingly popular in medical diagnostics due to its ability to handle high-dimensional data (many features relative to sample size), model complex interactions, and provide robust predictive performance with minimal tuning. Introduced by Breiman in 2001 (Breiman, 2001) the RF algorithm constructs a large number of decision trees during training, each based on a bootstrapped sample of the data (meaning that a large number of sample groups are constructed by randomly assembling a group and randomly replacing a member of that group to create a new sample group in an iterative fashion) and a random subset of predictors. Predictions are made by aggregating the outputs

across all trees (essentially a majority vote determines the classification, as opposed to linear regression models which are defined by the mean), thereby reducing the risk of overfitting (the creation of a model that conforms too closely to the idiosyncrasies of a particular sample, lacking generalizability) (Biau & Scornet, 2016).

Random forests are popular in medical diagnostics as they combine the robustness to non-linear relationships, collinearity and multidimensionality of other models (such as the single vector machine) with a higher degree of interpretability. A frequently voiced concern regarding machine learning classifiers is “black box” nature of the process. It can be difficult to trust a decision without any insight as to how it was alighted upon. Random forest models provide built-in estimates of feature importance—such as Mean Decrease in Gini Impurity or Mean Decrease in Accuracy—which offer interpretable insights into which variables most influence classification performance (Lipton, 2018).

Numerous studies have successfully applied RF classifiers to support diagnostic tasks, including early detection of Alzheimer's disease (Kolte et al., 2023). Dating back over a decade, random forest models have been demonstrating strong performance in distinguishing mild cognitive impairment (MCI) from healthy ageing by integrating neuropsychological, neuroimaging, and language-derived features (Gray et al., 2013).

In my experiment the random forest model performed with comparable precision to the traditional logistic regression approach. It’s interpretable output and relative robustness to high-dimensional data sets makes it a potentially attractive means of analyzing an increasing number of variables over the course of future work.

4.4.7 Considerations for further work

Many features were observed to repeatedly demonstrate trends and estimates that might suggest discriminant value, without reaching statistical significance. In part, this is likely attributable to the heterogeneity of the MCI group, the relatively low numbers and the high data dimensionality. However, by applying a data-drive approach to feature selection, with choices informed by previous studies and a theoretical groundwork, a feature set has been assembled that suggests better discriminant value compared to tools readily available to the typical clinician. To explore this further I intend to build a larger cohort of patients with biologically proven prodromal Alzheimer’s disease. Consolidating further features with recognized discriminant value (such as performance on picture description tasks (Hernández-Domínguez et al., 2018) and acoustic features (Ding et al., 2024) and applying exploratory machine learning techniques may contribute towards the development of an automatic system capable of accurate, language based diagnostic classification.

5. Conclusion

5.1 Summary of Key Findings

The aim of this thesis has been to explore whether semantic and linguistic features, derived from an automated and remotely administered assessment tool, could contribute meaningfully to the identification of prodromal AD.

In doing so, I have addressed the following hypotheses:

1. Unsupervised and remote interaction with a virtual clinical agent will permit the administration of verbal fluency tasks and provide data suitable for further analysis,

The CognoSpeak system appears capable of remotely administering and recording verbal fluency tasks as part of a more comprehensive assessment. Although high word error rate precluded the use of an entirely automated approach (using ASR based outputs), there is a clear precedent for how this can be addressed.

The CognoSpeak system demonstrates classification accuracy comparable to other automated systems. However, interpretability of the result remains limited and may negatively impact on the user experience.

The assessment of language, and in particular, lexical-semantic features is feasible. Integration of these features, with their intuitive nature and grounding in neuropathological and psychological therapy may improve classification accuracy as well as contribute towards a more meaningful clinical report.

2. Semantic features extracted from speech will contribute towards the identification of prodromal Alzheimer's disease.

The trends observed in the semantic features were consistent with previously reported findings and some demonstrate discriminant value between AD and control groups. The addition of semantic features to information that might be available to a clinician in a typical appointment improved classification accuracy and resulted in three fewer MCI patients being misclassified.

5.2 Interpretation of Key Findings

The findings are in line with the extant literature, and consistent with our understanding of the pathophysiology and distribution of AD. They support the clinical utility of semantic features.

The automated system demonstrated the capability to collect all the required information in an unsupervised and remote manner. With minor adjustments a completely automated approach is achievable.

Integration of the semantic features into the CognoSpeak tool will feasibly improve classification accuracy for prodromal AD. Using the data informed “Features-Optimized” model, which included semantic feature such as age of acquisition, cluster number and number of switches (alongside imaging metrics that were extracted automatically and an MMSE) an overall accuracy of 93% was achieved. Given that the CognoSpeak system has classification accuracy rates comparable to the MMSE, we might infer that a completely automated system with these features embedded (plus or minus a focused MRI protocol and automated reporting) could achieve similar predictive success.

This could be enormously impactful considering the challenges that memory clinics are likely to face in the years ahead. With more accurate biomarkers and more effective treatments becoming available it is likely that memory clinics will see a significant increase in the number of referrals (a trend that has been observed for years prior to these developments (O’Malley et al., 2020). It may be neither feasible, nor appropriate to review everyone in clinic and undertake a biological classification of their syndrome. An automated system such as described would have value in stratifying risk, informing the need for referral and permitting ongoing monitoring in the community.

5.3 Future Work

To build on this work I would like to apply these techniques on a larger, refined cohort of biomarker proven patients with prodromal AD. I expect that the trends observed in the data so far will meet significance and demonstrate even stronger discriminant value. This will permit a more confident interpretation of the results and potential promise for clinical application.

I plan to work with the broader CognoSpeak towards the integration of the semantic features into an experimental version of the CognoSpeak system and refine the ASR to achieve greater accuracy in the verbal fluency tasks. This will further inform the viability of a fully automated approach.

Whilst achieving any level of expertise in machine learning is a remote prospect, I think there is value in understanding the limitations of certain systems and the need for interpretable outputs. I will work with the team towards developing and refining such an output.

I intend to investigate automated imaging pipelines further and explore a stratified approach to the assessment, determining when a focused T1 MRI protocol capturing specific areas of interest (and analyzed automatically) may be indicated on the basis of the automated assessment.

5.4 Final Reflections

It is an exciting time to be working in the field of dementia. New, powerful diagnostic tools are at our disposal, and there is a tangible prospect of effective disease-modifying therapies in the near future. These developments have the potential to fundamentally reshape the way memory clinics operate and how patients live with Alzheimer's disease. Such changes will undoubtedly come with attendant challenges, and it will require innovation and collaboration to rise to them.

As we depart the era of evidence-based medicine and embrace the epoch of data driven personalized care it is important that we retain an understanding of the tools we employ.

Appendix A

Faculty of Medicine Dentistry & Health

Professor Chris Newman

Interim Pro-Vice-Chancellor

Medical School Office
Beech Hill Road
Sheffield
Telephone: +44-(0) 114-2713430
Fax: +44 (0) *114 2713158*
E-mail: a.venneri@sheffield.ac.uk

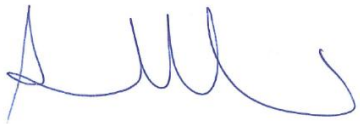
29th May 2020

Study title:	**Optimising detection of cognitive decline and associated symptoms**
REC reference:	**19/WS/0177**
IRAS project ID:	**244064**

I am writing with reference to the above study which received ethical and HRA approval on the 21st November 2019. This study has been significantly impacted by the current COVID-19 approval, and we have been unable to proceed with recruitment. After careful consideration we are now requesting an amendment to the original plan to be able to proceed with at least one segment of the project in which we ask the patient permission to donate their clinical MRI scan and clinical neuropsychological assessment to be used for research while the halting of non-COVID-19 related research is still in place. This would allow us to complete one segment of the project which does not require any active new data collection. We would like to recruit patients by sending them the Patient Information Sheet and Consent form by post and provide them with a return addressed and stamped envelope to post their signed consent back to us. These forms have been modified to be suitable for this purpose and they are attached to this request. We have also prepared an accompanying letter which will be posted to patients together with these forms.

Thank you for considering this request and we look forward to hearing from you in due course.

Yours sincerely,



Annalena Venneri

Professor of Clinical Neuropsychology

Honorary Consultant Clinical Neuropsychologist

Date ../../….

Invitation Letter

Dear Mr./Mrs.

We would like to invite you to participate in a research study entitled '**Optimizing detection of cognitive decline and associated symptoms**'

You have been contacted because you have attended the memory services at Sheffield Teaching Hospital and you have had an MRI scan and neuropsychological assessment as part of your clinical care. I am enclosing an information sheet and consent form that explain the purpose of the study and what it involves for you if you decided to take part.

We emphasized that participation in this study is entirely voluntary and there are no consequences for your treatment or follow-up.

If you need more information or any clarification about this study you can contact Dr Ronan O'Marley (contact details XXXX) who will be happy to answer any queries you may have

Yours sincerely,

Prof. Annalena Venneri.

CONSENT FORM 1

Title of project: Optimizing detection of cognitive decline and associated symptoms

Names of Researchers: *Prof A. Venneri, Dr D. Blackburn, Dr M. De Marco, Dr P. Sarrigiannis, Dr R. O'Malley, Dr R. Manca, Ms L. Wright, Mr J. M. Valera Bermejo*

Please initial box

1. I confirm that I have read and understood the information sheet (PIS_1 Version 3, 29 May 2020) for the above study and have had the opportunity to ask questions. ☐

2. I understand that my participation is voluntary and that I am free to withdraw at any time, without giving a reason. ☐

3. I agree to take part in the above study. ☐

4. I agree that the researchers can request my clinical brain MRI exams and use them for neurological research aimed at optimizing detection of cognitive decline. ☐

5. I agree that they can share my anonymized scans with other researchers for the purpose of this or other scientific research in optimizing methods to improve clinical diagnosis ☐

6. I agree that the researchers can access my clinical neuropsychological assessment and use it for neurological research aimed at optimizing detection of cognitive decline. ☐

7. I agree for all my data to be stored in an anonymized database to be used in future neurological research aimed at optimizing detection of cognitive decline. ☐

_____	_____	_____
Name of Participant	Date	Signature
_____	_____	_____
Name of Person taking consent	Date	Signature
_____	_____	_____
Name and Designation of Researcher	Date	Signature

Patient Information Sheet 1 **Invitation to participate in a research study on**

"Optimizing detection of cognitive decline and associated symptoms"

Invitation

We would like to invite you to take part in the study described below. Before you decide we would like you to understand why the research is being done and what it would involve for you. One of our team will go through the information with you and answer any questions you may have. Please ask if there is anything that is not clear.

Purpose of study

The objective of this research is to improve our ability to detect abnormal worsening of mental abilities such as attention, memory and language. Moderate decline of these abilities is normal as we age and it does not necessarily mean that something is wrong. However, it is important for those who work in a memory clinic to be able to recognize when this type of decline is cause for concern, in order to plan the most appropriate treatment as early as possible. In this study we would like to understand if a certain combination of examinations will improve the detection of abnormal decline of mental abilities.

Procedure

At this stage you have already had a brain MRI and you have already completed (or are about to complete) a cognitive assessment. We would like to ask you to:

- Enable us to have access to your brain MRI and to your cognitive assessment to be used in this research;

Access to your brain MRI and to your cognitive assessment

If you agree that we can use your brain MRI as part of this study, we will be able to repurpose clinical exams for medical research. This might involve collaborations with other researchers in other research teams we are collaborating with to optimize methods of analysis.

Do I have to take part?

No. It is up to you to decide to join the study. If you agree to take part, you will be asked to sign a Consent Form. Your personal results will be completely anonymous, will be assigned a study code and will be kept strictly confidential. The results will be stored in a restricted access database for approximately 10 years after the end of the study. You have, of course, the right to refuse to participate, and you may withdraw from the research at any time.

If you choose to withdraw from the study at any point, any data collected from you up until that time will continue to be used.

What will happen to me if I take part?

After having read this information sheet if you are happy to donate your existing MRI and test results for research you will have to indicate so by signing a dating the consent form and return this to us in the envelope provided. This marks the end of your participation.

What are the possible risks of taking part in this research?

There are no direct risks involved in this research.

Taking part in the research

By carrying out this research, we hope to gain a better understanding of the way memory other abilities decline with brain disease, and so improve early detection of abnormal changes in the ageing population.

How will we use information about you?

We will need to use information from your medical records for this research project.

This information will include your initials, hospital number, and name. People will use this information to do the research or to check your records to make sure that the research is being done properly. People who do not need to know who you are will not be able to see your name or contact details. Your data will have a code number instead. We will keep all information about you safe and secure.

Once we have finished the study, we will keep some of the data so we can check the results. We will write our reports in a way that no-one can work out that you took part in the study.

Original summary sheets with your scores without any identifiable information but only your study number will be stored in a safe and secured locked archive under the custody of the Principal Investigator for at least 10 years so that we can comply with the right of scientific journals to inspect the original data our published studies are based on.

What are your choices about how your information is used?

- You can stop being part of the study at any time, without giving a reason, but we will keep information about you that we already have.
- We need to manage your records in specific ways for the research to be reliable. This means that we won't be able to let you see or change the data we hold about you.
- If you agree to take part in this study, you will have the option to take part in future research using your data saved from this study.

Where can you find out more about how your information is used?

You can also ask any member of the research team and you can contact them using the phone or email (contact details are at the end of this sheet).

You can also contact Peter Wilson, the Data Protection Officer and Information Governance Manager of Sheffield Teaching Hospitals NHS Trust (Tel: 0114 2265151 or Email: sth.infogov@nhs.net)

You can find out more about how your information is used at the following websites:
www.hra.nhs.uk/information-about-patients/

https://www.hra.nhs.uk/planning-and-improving-research/policies-standards-legislation/data- protection-and-information-governance/gdpr-guidance/templates/template-wording-for- generic-information-document/

If you would like this information to be printed please ask a member of the research team (contact details below).

What if something goes wrong?

In the event that something does go wrong and you are harmed during the research and this is due to someone's negligence then you may have grounds for a legal action for compensation against Sheffield Teaching Hospitals NHS Foundation Trust but you may have to pay your legal costs.

If you want independent advice or have a complaint, you may contact the Patient Services Team (PST) at 01142712400 or by email at sth.pals@nhs.net

Contact Details

For further information about our research, or if you have a concern about any aspect of this study, please feel free to contact Professor Venneri of the research team. Details are given below.

Prof. Annalena Venneri E-mail: a.venneri@sheffield.ac.uk Telephone: 0114 2713430

Appendix B

Table: VPH Dare Healthy Controls and Study Controls

VPH Dare (1) vs Control (2)	***ID***	***MMSE***	***Age***	***Educ***	***Sex***
1	Pros-001	29	56	18	1
1	Pros-002	27	59	9	1
1	Pros-003	30	55	18	1
1	Pros-004	29	67	13	1
1	Pros-005	29	61	20	1
1	Pros-006	30	69	17	1
1	Pros-007	27	66	17	2
1	Pros-008	28	68	8	2
1	Pros-009	28	59	16	1
1	Pros-011	27	65	8	2
1	Pros-012	27	79	10	2
1	Pros-013	29	73	16	2
1	Pros-014	29	64	13	2
1	Pros-015	30	72	12	2
1	Pros-016	30	57	16	2
1	Pros-019	27	72	9	1
1	Pros-033	29	69	17	1
1	Pros-036	30	69	13	1
1	Pros-073	30	67	10	2
1	SH_DARE_G2_001	29	60	16	2
1	SH_DARE_G2_002	27	71	10	1
1	SH_DARE_G2_003	26	68	11	1
1	SH_DARE_G2_004	29	64	19	2
1	SH_DARE_G2_005	29	60	17	2
1	SH_DARE_G2_006	29	57	15	2
1	SH_DARE_G2_008	29	71	18	1
1	SH_DARE_G2_009	28	72	18	2
1	SH_DARE_G2_010	25	80	12	2

1	SH_DARE_G2_011	25	82	17	1
1	SH_DARE_G2_012	28	85	12	2
1	SH_DARE_G2_013	30	67	15	2
1	SH_DARE_G2_014	28	85	14	1
1	SH_DARE_G2_015	28	75	15	2
1	SH_DARE_G2_016	30	76	13	2
1	SH_DARE_G2_017	28	77	15	1
1	SH_DARE_G2_018	28	76	14	2
1	SH_DARE_G2_019	30	79	12	2
1	SH_DARE_G2_020	30	58	10	2
1	SH_DARE_G2_021	30	53	14	2
1	SH_DARE_G2_022	29	52	17	2
1	SH_DARE_G2_024	28	70	13	1
1	SH_DARE_G2_025	29	51	16	1
1	SH_DARE_G2_026	28	80	21	2
1	SH_DARE_G2_027	27	52	16	1
1	SH_DARE_G2_028	28	52	14	1
1	SH_DARE_G2_029	27	53	13	2
1	SH_DARE_G2_031	29	65	11	1
1	SH_DARE_G2_032	28	49	15	2
1	SH_DARE_G2_034	29	66	12	2
1	SH_DARE_G2_035	28	81	16	2
1	SH_DARE_G2_037	27	76	17	1
1	SH_DARE_G2_038	26	71	12	2
1	SH_DARE_G2_040	30	43	22	1
1	SH_DARE_G2_042	24	100	14	2
1	SH_DARE_G2_043	25	63	16	2
1	SH_DARE_G2_044	24	56	11	1
1	SH_DARE_G2_045	25	48	20	1
1	VL_DARE_001_3T	29	81	8	1
1	VL_DARE_003_3T	28	70	13	1
1	VL_DARE_004_3T	29	72	17	2
1	VL_DARE_005_3T	26	66	8	2
1	VL_DARE_006_3T	30	74	17	1
1	VL_DARE_007_3T	29	69	11	2
1	VL_DARE_010_3T	28	68	10	1
1	VL_DARE_012_3T	27	79	15	2
1	VL_DARE_015_3T	29	73	10	2
1	VL_DARE_016_3T	30	69	10	2
1	VL_DARE_018_3T	29	77	10	2

1	VL_DARE_020_3T	30	66	11	2
1	VL_DARE_021_3T	30	67	14	2
1	VL_DARE_024_3T	30	74	12	2
1	VL_DARE_025_3T	29	77	17	1
1	VL_DARE_026_3T	28	81	12	2
1	VL_DARE_028_3T	29	77	8	1
1	VL_DARE_029_3T	28	78	10	2
1	VL_DARE_031_3T	29	72	9	2
1	VL_DARE_032_3T	27	74	5	2
1	VL_DARE_036_3T	27	88	13	2
1	VL_DARE_041_3T	30	71	8	2
2	HCR2ODCAS001	29.13	75	21	2
2	HCR2ODCAS002	29.75	77	21	1
2	HCR2ODCAS003	29.75	67	15	2
2	HCR2ODCAS004	29.96	62	16	1
2	HCR2ODCAS006	28.53	86	16	2
2	HCR2ODCAS007	30	69	18	2
2	HCR2ODCAS008	27.94	74	14	2
2	HCR2ODCAS009	29.75	69	22	2
2	HCR2ODCAS010	29.75	82	15	1
2	HCR2ODCAS011	27.75	72	15	2
2	HCR2ODCAS012	29.13	73	18	1
2	HCR2ODCAS013	29.96	68	15	2
2	HCR2ODCAS014	27.94	77	15	1
2	HCR2ODCAS015	28.53	91	14	2
2	HCR2ODCAS016	29.75	72	18	2
2	HCR2ODCAS017	29.54	68	18	2
2	HCR2ODCAS018	29.96	71	16	2
2	HCR2ODCAS019	29.33	80	15	1
2	HCR2ODCAS020	26.77	76	16	2
2	HCR2ODCAS021	29.33	76	15	2
2	HCR2ODCAS022	28.73	80	18	1
2	HCR2ODCAS023	28.33	88	16	1
2	HCR2ODCAS024	28.14	82	18	2
2	HCR2ODCAS025	29.54	77	18	2
2	HCR2ODCAS026	29.75	64	18	2
2	HCR2ODCAS027	29.75	71	18	1
2	HCR2ODCAS028	30	79	19	1
2	HCR2ODCAS029	27.75	66	16	1
2	HCR2ODCAS030	30	76	21	2

2	HCR2ODCAS031	29.75	62	17	1
2	HCR2ODCAS032	28.93	70	17	2
2	HCR2ODCAS033	30	68	21	2
2	HCR2ODCAS034	29.96	74	15	2
2	HCR2ODCAS035	29.96	72	22	2
2	HCR2ODCAS036	29.75	58	18	2
2	HCR2ODCAS037	30	74	20	2
2	HCR2ODCAS038	29.33	72	15	2
2	HCR2ODCAS039	27.94	74	15	1
2	HCR2ODCAS040	30	75	21	2
2	HCR2ODCAS041	30	55	19	1
2	HCR2ODCAS042	29.33	71	17	2
2	HCR2ODCAS043	29.13	59	19	2
2	HCR2ODCAS044	29.54	62	21	2
2	HCR2ODCAS045	29.54	63	23	1
2	HCR2ODCAS046	29.96	78	20	2
2	HCR2ODCAS047	29.33	86	18	1
2	HCR2ODCAS048	28.73	69	21	1
2	HCR2ODCAS049	28.73	69	18	2
2	HCR2ODCAS050	30	65	21	2
2	HCR2ODCAS051	29.75	63	18	2
2	HCR2ODCAS052	29.13	66	15	2

Table Matched VPH Dare Participants

ID	*Diagnosis*	*MMSE*	*Age*	*Educ*	*Sex*
Pros-001	0	29	56	18	1
Pros-003	0	30	55	18	1
Pros-004	0	29	67	13	1
Pros-005	0	29	61	20	1
Pros-006	0	30	69	17	1
Pros-007	0	27	66	17	2
Pros-009	0	28	59	16	1
Pros-012	0	27	79	10	2
Pros-013	0	29	73	16	2
Pros-014	0	29	64	13	2
Pros-015	0	30	72	12	2

Pros-016	0	30	57	16	2
Pros-033	0	29	69	17	1
Pros-036	0	30	69	13	1
SH_DARE_G2_001	0	29	60	16	2
SH_DARE_G2_004	0	29	64	19	2
SH_DARE_G2_005	0	29	60	17	2
SH_DARE_G2_006	0	29	57	15	2
SH_DARE_G2_008	0	29	71	18	1
SH_DARE_G2_009	0	28	72	18	2
SH_DARE_G2_013	0	30	67	15	2
SH_DARE_G2_014	0	28	85	14	1
SH_DARE_G2_015	0	28	75	15	2
SH_DARE_G2_016	0	30	76	13	2
SH_DARE_G2_017	0	28	77	15	1
SH_DARE_G2_018	0	28	76	14	2
SH_DARE_G2_019	0	30	79	12	2
SH_DARE_G2_021	0	30	53	14	2
SH_DARE_G2_022	0	29	52	17	2
SH_DARE_G2_024	0	28	70	13	1
SH_DARE_G2_025	0	29	51	16	1
SH_DARE_G2_026	0	28	80	21	2
SH_DARE_G2_034	0	29	66	12	2
SH_DARE_G2_035	0	28	81	16	2
SH_DARE_G2_037	0	27	76	17	1
SH_DARE_G2_040	0	30	43	22	1
VL_DARE_003_3T	0	28	70	13	1
VL_DARE_004_3T	0	29	72	17	2
VL_DARE_006_3T	0	30	74	17	1
VL_DARE_007_3T	0	29	69	11	2
VL_DARE_012_3T	0	27	79	15	2
VL_DARE_015_3T	0	29	73	10	2
VL_DARE_018_3T	0	29	77	10	2
VL_DARE_020_3T	0	30	66	11	2
VL_DARE_021_3T	0	30	67	14	2
VL_DARE_024_3T	0	30	74	12	2
VL_DARE_025_3T	0	29	77	17	1
VL_DARE_026_3T	0	28	81	12	2
VL_DARE_029_3T	0	28	78	10	2

VL_DARE_036_3T	0	27	88	13	2

Appendix C

Healthy Controls Matched with ADMCI nested cohort

Code	*ID*	*Diagnosis*	*Age*	*Educ*	*Sex*
1	Pros-001	0	56	18	1
2	Pros-003	0	55	18	1
3	Pros-004	0	67	13	1
4	Pros-005	0	61	20	1
5	Pros-006	0	69	17	1
6	Pros-007	0	66	17	2
7	Pros-009	0	59	16	1
9	Pros-013	0	73	16	2
12	Pros-016	0	57	16	2
13	Pros-033	0	69	17	1
14	Pros-036	0	69	13	1
15	SH_DARE_G2_001	0	60	16	2
16	SH_DARE_G2_004	0	64	19	2
17	SH_DARE_G2_005	0	60	17	2
18	SH_DARE_G2_006	0	57	15	2
19	SH_DARE_G2_008	0	71	18	1
20	SH_DARE_G2_009	0	72	18	2
21	SH_DARE_G2_013	0	67	15	2
23	SH_DARE_G2_015	0	75	15	2
25	SH_DARE_G2_017	0	77	15	1
28	SH_DARE_G2_021	0	53	14	2
29	SH_DARE_G2_022	0	52	17	2
31	SH_DARE_G2_025	0	51	16	1
32	SH_DARE_G2_026	0	80	21	2
34	SH_DARE_G2_035	0	81	16	2
35	SH_DARE_G2_037	0	76	17	1
38	VL_DARE_004_3T	0	72	17	2
39	VL_DARE_006_3T	0	74	17	1
41	VL_DARE_012_3T	0	79	15	2

45	VL_DARE_021_3T	0	67	14	2
47	VL_DARE_025_3T	0	77	17	1
51	R2ODCAS3	1	58	16	1
52	R2ODCAS6	1	58	16	1
53	R2ODCAS9	1	63	15	2
54	R2ODCAS109	1	63	22	1
55	R2ODCAS12	1	52	16	1
56	R2ODCAS13	1	78	16	2
57	R2ODCAS41	1	69	22	1
58	R2ODCAS42	1	72	16	1
59	R2ODCAS63	1	66	21	2
60	R2ODCAS64	1	64	15	1
61	R2ODCAS65	1	88	18	2
62	R2ODCAS67	1	62	15	2
63	R2ODCAS69	1	51	17	1
64	R2ODCAS70	1	78	18	2
65	R2ODCAS71	1	65	18	2
66	R2ODCAS72	1	51	21	2
67	R2ODCAS74	1	68	21	1
68	R2ODCAS81	1	84	24	1
69	R2ODCAS83	1	56	16	1
70	R2ODCAS86	1	78	20	2
71	R2ODCAS88	1	61	16	2
72	R2ODCAS89	1	61	16	1
73	R2ODCAS91	1	67	18	2
74	R2ODCAS93	1	75	15	2
75	R2ODCAS94	1	37	16	2
76	R2ODCAS96	1	54	16	2
77	R2ODCAS99	1	68	16	2
78	R2ODCAS100	1	63	18	1
79	R2ODCAS101	1	82	18	1
80	R2ODCAS102	1	77	16	1
81	R2ODCAS108	1	65	15	2

Bibliography

Aerts, L., Heffernan, M., Kochan, N. A., Crawford, J. D., Draper, B., Trollor, J. N., Sachdev, P. S., & Brodaty, H. (2017). Effects of MCI subtype and reversion on progression to dementia in a community sample. *Neurology*, *88*(23), 2225–2232. https://doi.org/10.1212/wnl.0000000000004015

Ahmed, S., Haigh, A.-M. F., de Jager, C. A., & Garrard, P. (2013). Connected speech as a marker of disease progression in autopsy-proven Alzheimer's disease. *Brain*, *136*(12), 3727–3737. https://doi.org/10.1093/brain/awt269

Aisen, P. S., Cummings, J., Jack, C. R., Morris, J. C., Sperling, R., Frölich, L., Jones, R. W., Dowsett, S. A., Matthews, B. R., Raskin, J., Scheltens, P., & Dubois, B. (2017). On the path to 2025: Understanding the Alzheimer's disease continuum. *Alzheimers Res Ther*, *9*(1), 60. https://doi.org/10.1186/s13195-017-0283-5

Akazawa, K., Sakamoto, R., Nakajima, S., Wu, D., Li, Y., Oishi, K., Faria, A. V., Yamada, K., Togashi, K., Lyketsos, C. G., Miller, M. I., & Mori, S. (2019). Automated Generation of Radiologic Descriptions on Brain Volume Changes From T1-Weighted MR Images: Initial Assessment of Feasibility. *Frontiers in Neurology*, *10*, 7. https://doi.org/10.3389/fneur.2019.00007

Alacam, Ö., Schüz, S., Wegrzyn, M., Kißler, J., & Zarrieß, S. (2022). Exploring Semantic Spaces for Detecting Clustering and Switching in Verbal

Fluency. In N. Calzolari, C.-R. Huang, H. Kim, J. Pustejovsky, L. Wanner, K.-S. Choi, P.-M. Ryu, H.-H. Chen, L. Donatelli, H. Ji, S. Kurohashi, P. Paggio, N. Xue, S. Kim, Y. Hahm, Z. He, T. K. Lee, E. Santus, F. Bond, & S.-H. Na (Eds.), *Proceedings of the 29th International Conference on Computational Linguistics* (pp. 178–191). International Committee on Computational Linguistics. https://aclanthology.org/2022.coling-1.16/

Apostolova, L. G., Dutton, R. A., Dinov, I. D., Hayashi, K. M., Toga, A. W., Cummings, J. L., & Thompson, P. M. (2006). Conversion of Mild Cognitive Impairment to Alzheimer Disease Predicted by Hippocampal Atrophy Maps. *Archives of Neurology*, *63*(5), 693. https://doi.org/10.1001/archneur.63.5.693

Arango-Lasprilla, J. C., Cuetos, F., Valencia, C., Uribe, C., & Lopera, F. (2007). Cognitive changes in the preclinical phase of familial Alzheimer's disease. *Journal of Clinical and Experimental Neuropsychology*, *29*(8), 892–900. https://doi.org/10.1080/13803390601174151

Auriacombe, S., Lechevallier, N., Amieva, H., Harston, S., Raoux, N., & Dartigues, J.-F. (2006). A Longitudinal Study of Quantitative and Qualitative Features of Category Verbal Fluency in Incident Alzheimer's Disease Subjects: Results from the PAQUID Study. *Dementia and Geriatric Cognitive Disorders*, *21*(4), 260–266. https://doi.org/10.1159/000091407

Baldo, J. V., Schwartz, S., Wilkins, D., & Dronkers, N. F. (2006). Role of frontal versus temporal cortex in verbal fluency as revealed by voxel-based lesion symptom mapping. *Journal of the International Neuropsychological Society: JINS*, *12*(6), 896–900. https://doi.org/10.1017/S1355617706061078

Barnes, J., Scahill, R. I., Schott, J. M., Frost, C., Rossor, M. N., & Fox, N. C. (2005). Does Alzheimer's Disease Affect Hippocampal Asymmetry? Evidence from a Cross-Sectional and Longitudinal Volumetric MRI Study. *Dementia and Geriatric Cognitive Disorders*, *19*(5–6), 338–344. https://doi.org/10.1159/000084560

Bateman, R. J., Xiong, C., Benzinger, T. L., Fagan, A. M., Goate, A., Fox, N. C., Marcus, D. S., Cairns, N. J., Xie, X., Blazey, T. M., Holtzman, D. M., Santacruz, A., Buckles, V., Oliver, A., Moulder, K., Aisen, P. S., Ghetti, B., Klunk, W. E., McDade, E., … Morris, J. C. (2012). Clinical and biomarker changes in dominantly inherited Alzheimer's disease. *N Engl J Med*, *367*(9), 795–804. https://doi.org/10.1056/NEJMoa1202753

Biau, G., & Scornet, E. (2016). A random forest guided tour. *TEST*, *25*(2), 197–227. https://doi.org/10.1007/s11749-016-0481-7

Binder, J. R., & Desai, R. H. (2011). The neurobiology of semantic memory. *Trends in Cognitive Sciences*, *15*(11), 527–536. https://doi.org/10.1016/j.tics.2011.10.001

Biundo, R., Gardini, S., Caffarra, P., Concari, L., Martorana, D., Neri, T. M., Shanks, M. F., & Venneri, A. (2011). Influence of APOE status on lexical-semantic skills in mild cognitive impairment. *Journal of the International Neuropsychological Society: JINS*, *17*(3), 423–430. https://doi.org/10.1017/S135561771100021X

Blackburn, D. J., Wakefield, S., Shanks, M. F., Harkness, K., Reuber, M., & Venneri, A. (2014). Memory difficulties are not always a sign of incipient dementia: A review of the possible causes of loss of memory efficiency. *Br Med Bull*. https://doi.org/10.1093/bmb/ldu029

Blennow, K., & Zetterberg, H. (2018). Biomarkers for Alzheimer's disease: Current status and prospects for the future. *J Intern Med*, *284*(6), 643–663. https://doi.org/10.1111/joim.12816

Boccardi, M., Bocchetta, M., Ganzola, R., Robitaille, N., Redolfi, A., Duchesne, S., Jack, C. R., Frisoni, G. B., & EADC-ADNI Working Group on The Harmonized Protocol for Manual Hippocampal Segmentation and for the Alzheimer's Disease Neuroimaging Initiative. (2015). Operationalizing protocol differences for EADC-ADNI manual hippocampal segmentation. *Alzheimer's & Dementia: The Journal of the Alzheimer's Association*, *11*(2), 184–194. https://doi.org/10.1016/j.jalz.2013.03.001

Bonthrone, A. F., Blesa Cábez, M., Edwards, A. D., Hajnal, J. V., Counsell, S. J., & Boardman, J. P. (2025). Harmonizing multisite neonatal diffusion-weighted brain MRI data for developmental neuroscience. *Developmental*

Cognitive Neuroscience, *71*, 101488. https://doi.org/10.1016/j.dcn.2024.101488

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. https://doi.org/10.1007/s00401-006-0127-z

Breiman, L. (2001). Random Forests. *Machine Learning*, *45*(1), 5–32. https://doi.org/10.1023/A:1010933404324

Brysbaert, M., & Biemiller, A. (2017). Test-based age-of-acquisition norms for 44 thousand English word meanings. *Behavior Research Methods*, *49*(4), 1520–1523. https://doi.org/10.3758/s13428-016-0811-4

Brysbaert, M., & Ellis, A. W. (2016). Aphasia and age of acquisition: Are early-learned words more resilient? *Aphasiology*, *30*(11), 1240–1263. https://doi.org/10.1080/02687038.2015.1106439

Brysbaert, M., Mandera, P., McCormick, S. F., & Keuleers, E. (2019). Word prevalence norms for 62,000 English lemmas. *Behavior Research Methods*, *51*(2), 467–479. https://doi.org/10.3758/s13428-018-1077-9

Bushnell, J., Svaldi, D., Ayers, M. R., Gao, S., Unverzagt, F., Gaizo, J. D., Wadley, V. G., Kennedy, R., Goñi, J., & Clark, D. G. (2022). A comparison of techniques for deriving clustering and switching scores from verbal fluency word lists. *Frontiers in Psychology*, *13*. https://doi.org/10.3389/fpsyg.2022.743557

Campbell, N. L., Unverzagt, F., LaMantia, M. A., Khan, B. A., & Boustani, M. A. (2013). Risk factors for the progression of mild cognitive impairment to dementia. *Clinics in Geriatric Medicine*, *29*(4), 873–893. https://doi.org/10.1016/j.cger.2013.07.009

Cerhan, J. H., Ivnik, R. J., Smith, G. E., Tangalos, E. C., Petersen, R. C., & Boeve, B. F. (2002). Diagnostic Utility of Letter Fluency, Category Fluency, and Fluency Difference Scores in Alzheimer's Disease. *The Clinical Neuropsychologist*, *16*(1), 35–42. https://doi.org/10.1076/clin.16.1.35.8326

Chiotis, K., Saint-Aubert, L., Rodriguez-Vieitez, E., Leuzy, A., Almkvist, O., Savitcheva, I., Jonasson, M., Lubberink, M., Wall, A., Antoni, G., & Nordberg, A. (2018). Longitudinal changes of tau PET imaging in relation to hypometabolism in prodromal and Alzheimer's disease dementia. *Mol Psychiatry*, *23*(7), 1666–1673. https://doi.org/10.1038/mp.2017.108

Cho, S., Nevler, N., Parjane, N., Cieri, C., Liberman, M., Grossman, M., & Cousins, K. A. Q. (2021). Automated Analysis of Digitized Letter Fluency Data. *Frontiers in Psychology*, *12*. https://doi.org/10.3389/fpsyg.2021.654214

Chohan, P., Dashwood, M., Theodoulou, G., Reed, H., & Kuruvilla, T. (2022a). Blood-based biomarkers for Alzheimer's disease. *Progress in Neurology and Psychiatry*, *26*(4), 10–14. https://doi.org/10.1002/pnp.764

Chohan, P., Dashwood, M., Theodoulou, G., Reed, H., & Kuruvilla, T. (2022b). Blood-based biomarkers for Alzheimer's disease. *Progress in Neurology and Psychiatry*, *26*(4), 10–14. https://doi.org/10.1002/pnp.764

Cipriani, G., Dolciotti, C., Picchi, L., & Bonuccelli, U. (2011). Alzheimer and his disease: A brief history. *Neurological Sciences: Official Journal of the Italian Neurological Society and of the Italian Society of Clinical Neurophysiology*, *32*(2), 275–279. https://doi.org/10.1007/s10072-010-0454-7

Clark, C. M., Pontecorvo, M. J., Beach, T. G., Bedell, B. J., Coleman, R. E., Doraiswamy, P. M., Fleisher, A. S., Reiman, E. M., Sabbagh, M. N., Sadowsky, C. H., Schneider, J. A., Arora, A., Carpenter, A. P., Flitter, M. L., Joshi, A. D., Krautkramer, M. J., Lu, M., Mintun, M. A., & Skovronsky, D. M. (2012). Cerebral PET with florbetapir compared with neuropathology at autopsy for detection of neuritic amyloid-beta plaques: A prospective cohort study. *Lancet Neurol*, *11*(8), 669–678. https://doi.org/10.1016/s1474-4422(12)70142-4

Coon, D. W., & Edgerly, E. S. (1999). The Personal and Social Consequences of Alzheimer Disease. *Genetic Testing*, *3*(1), 29–36. https://doi.org/10.1089/gte.1999.3.29

Csernansky, J. G., Wang, L., Swank, J., Miller, J. P., Gado, M., McKeel, D., Miller, M. I., & Morris, J. C. (2005). Preclinical detection of Alzheimer's disease: Hippocampal shape and volume predict dementia onset in the

elderly. *NeuroImage*, *25*(3), 783–792. https://doi.org/10.1016/j.neuroimage.2004.12.036

Cuetos, F., Herrera, E., & Ellis, A. W. (2010). Impaired word recognition in Alzheimer's disease: The role of age of acquisition. *Neuropsychologia*, *48*(11), 3329–3334. https://doi.org/10.1016/j.neuropsychologia.2010.07.017

Cui, X., Li, L., Yu, L., Xing, H., Chang, H., Zhao, L., Qian, J., Song, Q., Zhou, S., & Dong, C. (2020). Gray Matter Atrophy in Parkinson's Disease and the Parkinsonian Variant of Multiple System Atrophy: A Combined ROI- and Voxel-Based Morphometric Study. *Clinics (Sao Paulo, Brazil)*, *75*, e1505. https://doi.org/10.6061/clinics/2020/e1505

Cummings, J. (2019). *The Role of Biomarkers in Alzheimer's Disease Drug Development*. *1118*, 29–61. https://doi.org/10.1007/978-3-030-05542-4_2

Cummings, J., Jones, R., Wilkinson, D., Lopez, O., Gauthier, S., Waldemar, G., Zhang, R., Xu, Y., Sun, Y., Richardson, S., & Mackell, J. (2010). Effect of donepezil on cognition in severe Alzheimer's disease: A pooled data analysis. *Journal of Alzheimer's Disease: JAD*, *21*(3), 843–851. https://doi.org/10.3233/JAD-2010-100078

Cummings, J. L., Tong, G., & Ballard, C. (2019). Treatment Combinations for Alzheimer's Disease: Current and Future Pharmacotherapy Options. *Journal of Alzheimer's Disease: JAD*, *67*(3), 779–794. https://doi.org/10.3233/JAD-180766

Dalton, M. A., Hornberger, M., & Piguet, O. (2016). Material specific lateralization of medial temporal lobe function: An fMRI investigation. *Human Brain Mapping*, *37*(3), 933–941. https://doi.org/10.1002/hbm.23077

Damasio, A. R. (1989). Time-locked multiregional retroactivation: A systems-level proposal for the neural substrates of recall and recognition. *Cognition*, *33*(1–2), 25–62. https://doi.org/10.1016/0010-0277(89)90005-x

Dasari, M., Kurian, J. A., Gundraju, S., Raparthi, A., & Medapati, R. V. (2025). Blood-Based β-Amyloid and Phosphorylated Tau (p-Tau) Biomarkers in Alzheimer's Disease: A Systematic Review of Their Diagnostic Potential. *Cureus*. https://doi.org/10.7759/cureus.79881

Davies, R. R., Graham, K. S., Xuereb, J. H., Williams, G. B., & Hodges, J. R. (2004). The human perirhinal cortex and semantic memory. *European Journal of Neuroscience*, *20*(9), 2441–2446. https://doi.org/10.1111/j.1460-9568.2004.03710.x

De Jong, L. W., Van Der Hiele, K., Veer, I. M., Houwing, J. J., Westendorp, R. G. J., Bollen, E. L. E. M., De Bruin, P. W., Middelkoop, H. A. M., Van Buchem, M. A., & Van Der Grond, J. (2008). Strongly reduced volumes of putamen and thalamus in Alzheimer's disease: An MRI study. *Brain*, *131*(12), 3277–3285. https://doi.org/10.1093/brain/awn278

de Leon, M. J., Convit, A., Wolf, O. T., Tarshish, C. Y., DeSanti, S., Rusinek, H., Tsui, W., Kandil, E., Scherer, A. J., Roche, A., Imossi, A., Thorn, E., Bobinski, M., Caraos, C., Lesbre, P., Schlyer, D., Poirier, J., Reisberg, B.,

& Fowler, J. (2001). Prediction of cognitive decline in normal elderly subjects with 2-[(18)F]fluoro-2-deoxy-D-glucose/poitron-emission tomography (FDG/PET). *Proc Natl Acad Sci U S A*, *98*(19), 10966–10971. https://doi.org/10.1073/pnas.191044198

De Marco, M., Blackburn, D. J., & Venneri, A. (2021). Serial Recall Order and Semantic Features of Category Fluency Words to Study Semantic Memory in Normal Ageing. *Frontiers in Aging Neuroscience*, *13*, 678588. https://doi.org/10.3389/fnagi.2021.678588

De Marco, M., & Venneri, A. (2021). Serial Recall Order of Category Fluency Words: Exploring Its Neural Underpinnings. *Frontiers in Psychology*, *12*, 777838. https://doi.org/10.3389/fpsyg.2021.777838

De Marco, M., Vonk, J. M. J., & Quaranta, D. (2023). Editorial: The mechanistic and clinical principles of item-level scoring methods applied to the category fluency test and other tests of semantic memory. *Frontiers in Psychology*, *14*. https://doi.org/10.3389/fpsyg.2023.1152574

DeCarli, C., Frisoni, G. B., Clark, C. M., Harvey, D., Grundman, M., Petersen, R. C., Thal, L. J., Jin, S., Jack, C. R., & Scheltens, P. (2007). Qualitative estimates of medial temporal atrophy as a predictor of progression from mild cognitive impairment to dementia. *Arch Neurol*, *64*(1), 108–115. https://doi.org/10.1001/archneur.64.1.108

Diamond, M. I. (2024). Travels with tau prions. *Cytoskeleton*, *81*(1), 83–88. https://doi.org/10.1002/cm.21806

DiBello, J. R., Lu, Y., Swartz, J., Bortnichak, E. A., Liaw, K.-L., Zhong, W., & Liu, X. (2023). Patterns of use of symptomatic treatments for Alzheimer's disease dementia (AD). *BMC Neurology*, *23*(1), 400. https://doi.org/10.1186/s12883-023-03447-5

Didic, M., Barbeau, E. J., Felician, O., Tramoni, E., Guedj, E., Poncet, M., & Ceccaldi, M. (2011). Which memory system is impaired first in Alzheimer's disease? *J Alzheimers Dis*, *27*(1), 11–22. https://doi.org/10.3233/jad-2011-110557

Dimitriadis, S. I., Liparas, D., & Initiative, for the A. D. N. (2018). How random is the random forest? Random forest algorithm on the service of structural imaging biomarkers for Alzheimer's disease: from Alzheimer's disease neuroimaging initiative (ADNI) database. *Neural Regeneration Research*, *13*(6), 962. https://doi.org/10.4103/1673-5374.233433

Ding, H., Lister, A., Karjadi, C., Au, R., Lin, H., Bischoff, B., & Hwang, P. H. (2024). Detection of Mild Cognitive Impairment From Non-Semantic, Acoustic Voice Features: The Framingham Heart Study. *JMIR Aging*, *7*, e55126. https://doi.org/10.2196/55126

Doran, S. J., & Sawyer, R. P. (2024). Risk factors in developing amyloid related imaging abnormalities (ARIA) and clinical implications. *Frontiers in Neuroscience*, *18*, 1326784. https://doi.org/10.3389/fnins.2024.1326784

Dubois, B., Feldman, H. H., Jacova, C., Hampel, H., Molinuevo, J. L., Blennow, K., DeKosky, S. T., Gauthier, S., Selkoe, D., Bateman, R., Cappa, S.,

Crutch, S., Engelborghs, S., Frisoni, G. B., Fox, N. C., Galasko, D., Habert, M. O., Jicha, G. A., Nordberg, A., … Cummings, J. L. (2014). Advancing research diagnostic criteria for Alzheimer's disease: The IWG-2 criteria. *Lancet Neurol*, *13*(6), 614–629. https://doi.org/10.1016/s1474-4422(14)70090-0

Dubois, B., Villain, N., Schneider, L. S., et al. (2024). Alzheimer disease is a clinical-biological construct: An International Working Group (IWG) recommendation. JAMA Neurology.

Duff, M. C., Covington, N. V., Hilverman, C., & Cohen, N. J. (2020). Semantic Memory and the Hippocampus: Revisiting, Reaffirming, and Extending the Reach of Their Critical Relationship. *Frontiers in Human Neuroscience*, *13*. https://doi.org/10.3389/fnhum.2019.00471

Elias-Sonnenschein, L. S., Viechtbauer, W., Ramakers, I. H., Verhey, F. R., & Visser, P. J. (2011). Predictive value of APOE-epsilon4 allele for progression from MCI to AD-type dementia: A meta-analysis. *J Neurol Neurosurg Psychiatry*, *82*(10), 1149–1156. https://doi.org/10.1136/jnnp.2010.231555

Endel Tulving. (2002). Episodic Memory: From Mind to Brain. *Annual Review of Psychology*, *53*(1), 1–25. https://doi.org/10.1146/annurev.psych.53.100901.135114

Erichsen, P. A. (2024). Amyloid-related imaging abnormalities in antiamyloid therapy for Alzheimer's disease: A narrative review. *Adverse Drug*

Reaction Bulletin, *348*(1), 1351–1354. https://doi.org/10.1097/FAD.0000000000000077

Esmaeili, S., Obeidat, A. Z., & Zabeti, A. (2023). Molecular biomarkers and cognitive impairment in multiple sclerosis: A review. *Biomarkers in Neuropsychiatry*, *9*, 100077. https://doi.org/10.1016/j.bionps.2023.100077

Fagan, A. M., Mintun, M. A., Shah, A. R., Aldea, P., Roe, C. M., Mach, R. H., Marcus, D., Morris, J. C., & Holtzman, D. M. (2009). Cerebrospinal fluid tau and ptau(181) increase with cortical amyloid deposition in cognitively normal individuals: Implications for future clinical trials of Alzheimer's disease. *EMBO Mol Med*, *1*(8–9), 371–380. https://doi.org/10.1002/emmm.200900048

Fagundo, A. B., López, S., Romero, M., Guarch, J., Marcos, T., & Salamero, M. (2008). Clustering and switching in semantic fluency: Predictors of the development of Alzheimer's disease. *International Journal of Geriatric Psychiatry*, *23*(10), 1007–1013. https://doi.org/10.1002/gps.2025

Fan, Z., Liu, X., Liu, J., Chen, C., & Zhou, M. (2023). Neurofilament Light Chain as a Potential Biomarker in Plasma for Alzheimer's Disease and Mild Cognitive Impairment: A Systematic Review and a Meta-Analysis. *Journal of Integrative Neuroscience*, *22*(4), 85. https://doi.org/10.31083/j.jin2204085

Farias, S. T., Mungas, D., Reed, B. R., Harvey, D., & DeCarli, C. (2009). Progression of mild cognitive impairment to dementia in clinic- vs

community-based cohorts. *Arch Neurol*, *66*(9), 1151–1157. https://doi.org/10.1001/archneurol.2009.106

Fathy, Y. Y., Hoogers, S. E., Berendse, H. W., van der Werf, Y. D., Visser, P. J., de Jong, F. J., & van de Berg, W. D. J. (2020). Differential insular cortex sub-regional atrophy in neurodegenerative diseases: A systematic review and meta-analysis. *Brain Imaging and Behavior*, *14*(6), 2799–2816. https://doi.org/10.1007/s11682-019-00099-3

Femminella, G. D., Thayanandan, T., Calsolaro, V., Komici, K., Rengo, G., Corbi, G., & Ferrara, N. (2018). Imaging and Molecular Mechanisms of Alzheimer's Disease: A Review. *Int J Mol Sci*, *19*(12). https://doi.org/10.3390/ijms19123702

Ferreira, L. K., Diniz, B. S., Forlenza, O. V., Busatto, G. F., & Zanetti, M. V. (2011). Neurostructural predictors of Alzheimer's disease: A meta-analysis of VBM studies. *Neurobiology of Aging*, *32*(10), 1733–1741. https://doi.org/10.1016/j.neurobiolaging.2009.11.008

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. https://doi.org/10.1016/0022-3956(75)90026-6

Forbes-McKay, K. E., Ellis, A. W., Shanks, M. F., & Venneri, A. (2005). The age of acquisition of words produced in a semantic fluency task can reliably differentiate normal from pathological age related cognitive decline.

Neuropsychologia, *43*(11), 1625–1632. https://doi.org/10.1016/j.neuropsychologia.2005.01.008

Frisoni, G., Testa, C., Zorzan, A., Sabattoli, F., Beltramello, A., Soininen, H., & Laakso, M. (2002). Detection of grey matter loss in mild Alzheimer's disease with voxel based morphometry. *Journal of Neurology, Neurosurgery, and Psychiatry*, *73*(6), 657–664. https://doi.org/10.1136/jnnp.73.6.657

Gabrieli, J. D. E., Cohen, N. J., & Corkin, S. (1988). The impaired learning of semantic knowledge following bilateral medial temporal-lobe resection. *Brain and Cognition*, *7*(2), 157–177. https://doi.org/10.1016/0278-2626(88)90027-9

Garrard, P., Maloney, L. M., Hodges, J. R., & Patterson, K. (2005). The effects of very early Alzheimer's disease on the characteristics of writing by a renowned author. *Brain: A Journal of Neurology*, *128*(Pt 2), 250–260. https://doi.org/10.1093/brain/awh341

Gayraud, F., Thibert, C., & Barkat-Defradas, M. (2012, June 21). *Age of Acquisition Affects Word Retrieval in Spontaneous Speech produced by Patients with Alzheimer's Disease*. https://www.semanticscholar.org/paper/Age-of-Acquisition-Affects-Word-Retrieval-in-Speech-Gayraud-Thibert/162688545d49699fc3079163136a6eb57ac65c0c

Giffard, B., Laisney, M., Desgranges, B., & Eustache, F. (2015). An exploration of the semantic network in Alzheimer's disease: Influence of emotion and concreteness of concepts. *Cortex*, *69*, 201–211. https://doi.org/10.1016/j.cortex.2015.05.020

Glisky, E. L., Schacter, D. L., & Tulving, E. (1986). Computer learning by memory-impaired patients: Acquisition and retention of complex knowledge. *Neuropsychologia*, *24*(3), 313–328. https://doi.org/10.1016/0028-3932(86)90017-5

Gonzalez-Recober, C., Nevler, N., Shellikeri, S., Cousins, K. A. Q., Rhodes, E., Liberman, M., Grossman, M., Irwin, D., & Cho, S. (2023). Comparison of category and letter fluency tasks through automated analysis. *Frontiers in Psychology*, *14*, 1212793. https://doi.org/10.3389/fpsyg.2023.1212793

Goto, M., Abe, O., Hagiwara, A., Fujita, S., Kamagata, K., Hori, M., Aoki, S., Osada, T., Konishi, S., Masutani, Y., Sakamoto, H., Sakano, Y., Kyogoku, S., & Daida, H. (2022). Advantages of Using Both Voxel- and Surface-based Morphometry in Cortical Morphology Analysis: A Review of Various Applications. *Magnetic Resonance in Medical Sciences*, *21*(1), 41–57. https://doi.org/10.2463/mrms.rev.2021-0096

Graham, K. S., & Hodges, J. R. (1997). Differentiating the roles of the hippocampal complex and the neocortex in long-term memory storage: Evidence from the study of semantic dementia and Alzheimer's disease. *Neuropsychology*, *11*(1), 77–89.

Grande, G., Valletta, M., Rizzuto, D., Xia, X., Qiu, C., Orsini, N., Dale, M., Andersson, S., Fredolini, C., Winblad, B., Laukka, E. J., Fratiglioni, L., & Vetrano, D. L. (2025). Blood-based biomarkers of Alzheimer's disease and incident dementia in the community. *Nature Medicine*, 1–9. https://doi.org/10.1038/s41591-025-03605-x

Gray, K. R., Aljabar, P., Heckemann, R. A., Hammers, A., & Rueckert, D. (2013). Random forest-based similarity measures for multi-modal classification of Alzheimer's disease. *NeuroImage*, *65C*, 167–175. https://doi.org/10.1016/j.neuroimage.2012.09.065

Greenberg, S. M. (2024). CAA and ARIA: Emerging Concepts and Opportunities for Translation. *Alzheimer's & Dementia*, *20*(S1), e085576. https://doi.org/10.1002/alz.085576

Grossman, M., Anderson, C., Khan, A., Avants, B., Elman, L., & McCluskey, L. (2008). Impaired action knowledge in amyotrophic lateral sclerosis. *Neurology*, *71*(18), 1396–1401. https://doi.org/10.1212/01.wnl.0000319701.50168.8c

Gruenewald, P. J., & Lockhead, G. R. (1980). The free recall of category examples. *Journal of Experimental Psychology: Human Learning and Memory*, *6*(3), 225–240. https://doi.org/10.1037/0278-7393.6.3.225

Haass, C., & Selkoe, D. J. (2007). Soluble protein oligomers in neurodegeneration: Lessons from the Alzheimer's amyloid beta-peptide.

Nature Reviews. Molecular Cell Biology, *8*(2), 101–112. https://doi.org/10.1038/nrm2101

Hähnel, T., Feige, T., Kunze, J., Epler, A., Frank, A., Bendig, J., Schnalke, N., Wolz, M., Themann, P., & Falkenburger, B. (2023). A Semantic Relatedness Model for the Automatic Cluster Analysis of Phonematic and Semantic Verbal Fluency Tasks Performed by People With Parkinson Disease: Prospective Multicenter Study. *JMIR Neurotechnology*, *2*(1), e46021. https://doi.org/10.2196/46021

Hansson, O., Zetterberg, H., Buchhave, P., Londos, E., Blennow, K., & Minthon, L. (2006). Association between CSF biomarkers and incipient Alzheimer's disease in patients with mild cognitive impairment: A follow-up study. *Lancet Neurol*, *5*(3), 228–234. https://doi.org/10.1016/s1474-4422(06)70355-6

Henry, J. D., & Crawford, J. R. (2004). A meta-analytic review of verbal fluency performance following focal cortical lesions. *Neuropsychology*, *18*(2), 284–295. https://doi.org/10.1037/0894-4105.18.2.284

Hernández-Domínguez, L., Ratté, S., Sierra-Martínez, G., & Roche-Bergua, A. (2018). Computer-based evaluation of Alzheimer's disease and mild cognitive impairment patients during a picture description task. *Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring*, *10*, 260–268. https://doi.org/10.1016/j.dadm.2018.02.004

Hirshorn, E. A., & Thompson-Schill, S. L. (2006). Role of the left inferior frontal gyrus in covert word retrieval: Neural correlates of switching during verbal fluency. *Neuropsychologia*, *44*(12), 2547–2557. https://doi.org/10.1016/j.neuropsychologia.2006.03.035

Hodges, J. R., & Graham, K. S. (2001). Episodic memory: Insights from semantic dementia. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, *356*(1413), 1423–1434. https://doi.org/10.1098/rstb.2001.0943

Høilund-Carlsen, P., Revheim, M.-E., Costa, T., Kepp, K., Castellani, R., Perry, G., Alavi, A., & Barrio, J. (2023). FDG-PET versus Amyloid-PET Imaging for Diagnosis and Response Evaluation in Alzheimer's Disease: Benefits and Pitfalls. *Diagnostics*, *13*(13), 2254. https://doi.org/10.3390/diagnostics13132254

Hsieh, S., Schubert, S., Hoon, C., Mioshi, E., & Hodges, J. R. (2013). Validation of the Addenbrooke's Cognitive Examination III in frontotemporal dementia and Alzheimer's disease. *Dementia and Geriatric Cognitive Disorders*, *36*(3–4), 242–250. https://doi.org/10.1159/000351671

Hughes, B. (1970). MISSILE WOUNDS OF THE BRAIN A Study of Psychological Deficits. *Journal of Neurology, Neurosurgery, and Psychiatry*, *33*(4), 551.

Irish, M., Addis, D. R., Hodges, J. R., & Piguet, O. (2012). Considering the role of semantic memory in episodic future thinking: Evidence from semantic dementia. *Brain*, *135*(7), 2178–2191. https://doi.org/10.1093/brain/aws119

Irish, M., Bunk, S., Tu, S., Kamminga, J., Hodges, J. R., Hornberger, M., & Piguet, O. (2016). Preservation of episodic memory in semantic dementia: The importance of regions beyond the medial temporal lobes. *Neuropsychologia*, *81*, 50–60. https://doi.org/10.1016/j.neuropsychologia.2015.12.005

Jack, C., Bennett, D., Blennow, K., Dunn, B., Elliott, C., Haeberlein, S., Holtzman, D., Jagust, M., Jessen, F., Karlawish, J., Liu, E., Molinuevo, J., Montine, T., Phelps, C., Rankin, K., Rowe, C., Laurie, Ryan, Scheltens, P., … Sperling, R. (2017). *Framework: Towards a Biological Definition of Alzheimer ' s Disease 1 2 draft 9-19-17 3 4 5*. https://www.semanticscholar.org/paper/Framework-%3A-Towards-a-Biological-Definition-of-%E2%80%99-s-Jack-Bennett/4e78b56c0abd466b9090fef005c09c7baf0870c9

Jack, C. R., Albert, M., Knopman, D. S., McKhann, G. M., Sperling, R. A., Carillo, M., Thies, W., & Phelps, C. H. (2011). Introduction to Revised Criteria for the Diagnosis of Alzheimer's Disease: National Institute on Aging and the Alzheimer Association Workgroups. *Alzheimers Dement*, *7*(3), 257–262. https://doi.org/10.1016/j.jalz.2011.03.004

Jack, C. R., Bennett, D. A., Blennow, K., Carrillo, M. C., Dunn, B., Haeberlein, S. B., Holtzman, D. M., Jagust, W., Jessen, F., Karlawish, J., Liu, E., Molinuevo, J. L., Montine, T., Phelps, C., Rankin, K. P., Rowe, C. C., Scheltens, P., Siemers, E., Snyder, H. M., & Sperling, R. (2018). NIA-AA Research Framework: Toward a biological definition of Alzheimer's disease. *Alzheimers Dement*, *14*(4), 535–562. https://doi.org/10.1016/j.jalz.2018.02.018

Jack Jr., C. R., Andrews, J. S., Beach, T. G., Buracchio, T., Dunn, B., Graf, A., Hansson, O., Ho, C., Jagust, W., McDade, E., Molinuevo, J. L., Okonkwo, O. C., Pani, L., Rafii, M. S., Scheltens, P., Siemers, E., Snyder, H. M., Sperling, R., Teunissen, C. E., & Carrillo, M. C. (2024). Revised criteria for diagnosis and staging of Alzheimer's disease: Alzheimer's Association Workgroup. *Alzheimer's & Dementia*, *20*(8), 5143–5169. https://doi.org/10.1002/alz.13859

Janelidze, S., Bali, D., Ashton, N. J., Barthélemy, N. R., Vanbrabant, J., Stoops, E., Vanmechelen, E., He, Y., Dolado, A. O., Triana-Baltzer, G., Pontecorvo, M. J., Zetterberg, H., Kolb, H., Vandijck, M., Blennow, K., Bateman, R. J., & Hansson, O. (2023). Head-to-head comparison of 10 plasma phospho-tau assays in prodromal Alzheimer's disease. *Brain: A Journal of Neurology*, *146*(4), 1592–1601. https://doi.org/10.1093/brain/awac333

Janelidze, S., Stomrud, E., Smith, R., Palmqvist, S., Mattsson, N., Airey, D. C., Proctor, N. K., Chai, X., Shcherbinin, S., Sims, J. R., Dage, J. L., & Hansson, O. (2020). *Cerebrospinal fluid p-tau217 performs better than p-tau181 as a biomarker of Alzheimer's disease*. Neurology. https://doi.org/10.1101/2020.01.15.20017236

Janelidze, S., Teunissen, C. E., Zetterberg, H., Allué, J. A., Sarasa, L., Eichenlaub, U., Bittner, T., Ovod, V., Verberk, I. M. W., Toba, K., Nakamura, A., Bateman, R. J., Blennow, K., & Hansson, O. (2021). Head-to-Head Comparison of 8 Plasma Amyloid-β 42/40 Assays in Alzheimer Disease. *JAMA Neurology*, *78*(11), 1375–1382. https://doi.org/10.1001/jamaneurol.2021.3180

Jeong, S. Y., Suh, C. H., Lim, J.-S., Shim, W. H., Heo, H., Choi, Y., Kim, H. S., Kim, S. J., & Lee, J.-H. (2025). Incidence of Amyloid-Related Imaging Abnormalities in Phase III Clinical Trials of Anti-Amyloid-β Immunotherapy. *Neurology*, *104*(8), e213483. https://doi.org/10.1212/WNL.0000000000213483

Jeremic, D., Navarro-López, J. D., & Jiménez-Díaz, L. (2024). *Donanemab outperformed Aducanumab and Lecanemab on cognitive, but not on biomarker and safety outcomes: Systematic review, frequentist and Bayesian network meta-analyses*. Pharmacology and Therapeutics. https://doi.org/10.1101/2024.03.31.24305134

Johnson, K. A., Fox, N. C., Sperling, R. A., & Klunk, W. E. (2012). Brain imaging in Alzheimer disease. *Cold Spring Harb Perspect Med*, *2*(4), a006213. https://doi.org/10.1101/cshperspect.a006213

Johnson, K. A., Minoshima, S., Bohnen, N. I., Donohoe, K. J., Foster, N. L., Herscovitch, P., Karlawish, J. H., Rowe, C. C., Carrillo, M. C., Hartley, D. M., Hedrick, S., Pappas, V., & Thies, W. H. (2013). Appropriate use criteria for amyloid PET: a report of the Amyloid Imaging Task Force, the Society of Nuclear Medicine and Molecular Imaging, and the Alzheimer's Association. *Alzheimers Dement*, *9*(1), e-1-16. https://doi.org/10.1016/j.jalz.2013.01.002

Jones, D., Drew, P., Elsey, C., Blackburn, D., Wakefield, S., Harkness, K., & Reuber, M. (2016). Conversational assessment in memory clinic encounters: Interactional profiling for differentiating dementia from functional memory disorders. *Aging & Mental Health*, *20*(5), 500–509. https://doi.org/10.1080/13607863.2015.1021753

Jorge Cardoso, M., Leung, K., Modat, M., Keihaninejad, S., Cash, D., Barnes, J., Fox, N. C., & Ourselin, S. (2013). STEPS: Similarity and Truth Estimation for Propagated Segmentations and its application to hippocampal segmentation and brain parcelation. *Medical Image Analysis*, *17*(6), 671–684. https://doi.org/10.1016/j.media.2013.02.006

Joubert, S., Brambati, S. M., Ansado, J., Barbeau, E. J., Felician, O., Didic, M., Lacombe, J., Goldstein, R., Chayer, C., & Kergoat, M.-J. (2010). The

cognitive and neural expression of semantic memory impairment in mild cognitive impairment and early Alzheimer's disease. *Neuropsychologia*, *48*(4), 978–988. https://doi.org/10.1016/j.neuropsychologia.2009.11.019

Karve, S. J., Ringman, J. M., Lee, A. S., Juarez, K. O., & Mendez, M. F. (2012). Comparison of clinical characteristics between familial and non-familial early onset Alzheimer's disease. *Journal of Neurology*, *259*(10), 2182–2188. https://doi.org/10.1007/s00415-012-6481-y

Kim, H.-R., Lee, P., Seo, S. W., Roh, J. H., Oh, M., Oh, J. S., Oh, S. J., Kim, J. S., & Jeong, Y. (2019). Comparison of Amyloid β and Tau Spread Models in Alzheimer's Disease. *Cerebral Cortex*, *29*(10), 4291–4302. https://doi.org/10.1093/cercor/bhy311

Kim, J., Chakrabarty, P., Hanna, A., March, A., Dickson, D. W., Borchelt, D. R., Golde, T., & Janus, C. (2013). Normal cognition in transgenic BRI2-Aβ mice. *Molecular Neurodegeneration*, *8*(1), 15. https://doi.org/10.1186/1750-1326-8-15

Kim, N., Kim, J.-H., Wolters, M. K., MacPherson, S. E., & Park, J. C. (2019). Automatic Scoring of Semantic Fluency. *Frontiers in Psychology*, *10*, 1020. https://doi.org/10.3389/fpsyg.2019.01020

Klein, G., Delmar, P., Voyle, N., Rehal, S., Hofmann, C., Abi-Saab, D., Andjelkovic, M., Ristic, S., Wang, G., Bateman, R., Kerchner, G. A., Baudler, M., Fontoura, P., & Doody, R. (2019). Gantenerumab reduces amyloid-β plaques in patients with prodromal to moderate Alzheimer's

disease: A PET substudy interim analysis. *Alzheimers Res Ther*, *11*(1), 101. https://doi.org/10.1186/s13195-019-0559-z

Klöppel, S., Stonnington, C. M., Barnes, J., Chen, F., Chu, C., Good, C. D., Mader, I., Mitchell, L. A., Patel, A. C., Roberts, C. C., Fox, N. C., Jack, C. R., Ashburner, J., & Frackowiak, R. S. J. (2008). Accuracy of dementia diagnosis: A direct comparison between radiologists and a computerized method. *Brain: A Journal of Neurology*, *131*(Pt 11), 2969–2974. https://doi.org/10.1093/brain/awn239

Knopman, D. S., & Roberts, R. (2010). Vascular Risk Factors: Imaging and Neuropathologic Correlates. *Journal of Alzheimer's Disease*, *20*(3), 699–709. https://doi.org/10.3233/JAD-2010-091555

Kolte, P., Rabra, N., Shrivastava, A., Khadatkar, A., Choudhary, H., & Shrivastava, D. (2023). Early Alzheimer's Detection Using Random Forest Algorithm. *2023 International Conference on Signal Processing, Computation, Electronics, Power and Telecommunication (IConSCEPT)*, 1–5. https://doi.org/10.1109/IConSCEPT57958.2023.10170234

Kroth, H., Oden, F., Molette, J., Schieferstein, H., Capotosti, F., Mueller, A., Berndt, M., Schmitt-Willich, H., Darmency, V., Gabellieri, E., Boudou, C., Juergens, T., Varisco, Y., Vokali, E., Hickman, D. T., Tamagnan, G., Pfeifer, A., Dinkelborg, L., Muhs, A., & Stephens, A. (2019). Discovery and preclinical characterization of [18F]PI-2620, a next-generation tau PET tracer for the assessment of tau pathology in Alzheimer's disease and

other tauopathies. *European Journal of Nuclear Medicine and Molecular Imaging*, *46*(10), 2178–2189. https://doi.org/10.1007/s00259-019-04397-2

Kuller, L. H., & Lopez, O. L. (2021). ENGAGE and EMERGE: Truth and consequences? *Alzheimer's & Dementia*, *17*(4), 692–695. https://doi.org/10.1002/alz.12286

Lai, R., Li, B., & Bishnoi, R. (2024). P-tau217 as a Reliable Blood-Based Marker of Alzheimer's Disease. *Biomedicines*, *12*(8), Article 8. https://doi.org/10.3390/biomedicines12081836

Langa, K. M., & Levine, D. A. (2014). The diagnosis and management of mild cognitive impairment: A clinical review. *Jama*, *312*(23), 2551–2561. https://doi.org/10.1001/jama.2014.13806

Le, X., Lancashire, I., Hirst, G., & Jokel, R. (2011). Longitudinal detection of dementia through lexical and syntactic changes in writing: A case study of three British novelists. *Literary and Linguistic Computing*, *26*(4), 435–461. https://doi.org/10.1093/llc/fqr013

Li, X., Morgan, P. S., Ashburner, J., Smith, J., & Rorden, C. (2016). The first step for neuroimaging data analysis: DICOM to NIfTI conversion. *Journal of Neuroscience Methods*, *264*, 47–56. https://doi.org/10.1016/j.jneumeth.2016.03.001

Li, Y., Rinne, J. O., Mosconi, L., Pirraglia, E., Rusinek, H., DeSanti, S., Kemppainen, N., Någren, K., Kim, B.-C., Tsui, W., & de Leon, M. J. (2008). Regional analysis of FDG and PIB-PET images in normal aging,

mild cognitive impairment, and Alzheimer's disease. *European Journal of Nuclear Medicine and Molecular Imaging*, *35*(12), 2169–2181. https://doi.org/10.1007/s00259-008-0833-y

Lipton, S. A. (2005). The molecular basis of memantine action in Alzheimer's disease and other neurologic disorders: Low-affinity, uncompetitive antagonism. *Current Alzheimer Research*, *2*(2), 155–165. https://doi.org/10.2174/1567205053585846

Lipton, Z. C. (2018). The mythos of model interpretability. *Commun. ACM*, *61*(10), 36–43. https://doi.org/10.1145/3233231

Lobo, A., Launer, L., Fratiglioni, L., Andersen, K., Carlo, A., Breteler, M., Copeland, J., Dartigues, J., Jagger, C., Martínez-Lage, J., Soininen, H., & Hofman, A. (2000). Prevalence of dementia and major subtypes in Europe: A collaborative study of population-based cohorts. Neurologic Diseases in the Elderly Research Group. *Neurology*. https://www.semanticscholar.org/paper/Prevalence-of-dementia-and-major-subtypes-in-A-of-Lobo-Launer/3289f03869a1798c837225e958d9c619a5509dd9

Luo, J., Agboola, F., Grant, E., Masters, C. L., Albert, M. S., Johnson, S. C., McDade, E. M., Vöglein, J., Fagan, A. M., Benzinger, T., Massoumzadeh, P., Hassenstab, J., Bateman, R. J., Morris, J. C., Perrin, R. J., Chhatwal, J., Jucker, M., Ghetti, B., Cruchaga, C., … Xiong, C. (2020). Sequence of Alzheimer disease biomarker changes in cognitively normal adults: A

cross-sectional study. *Neurology*, *95*(23). https://doi.org/10.1212/WNL.0000000000010747

Mandelkow, E.-M. (2008). S2-03–06: The role of tau in axonal transport and neurodegeneration. *Alzheimer's & Dementia*, *4*(4S_Part_4). https://doi.org/10.1016/j.jalz.2008.05.288

Mandera, P., Keuleers, E., & Brysbaert, M. (2020). Recognition times for 62 thousand English words: Data from the English Crowdsourcing Project. *Behavior Research Methods*, *52*(2), 741–760. https://doi.org/10.3758/s13428-019-01272-8

Mantellatto Grigoli, M., Pelegrini, L. N. C., Whelan, R., & Cominetti, M. R. (2024). Present and Future of Blood-Based Biomarkers of Alzheimer's Disease: Beyond the Classics. *Brain Research*, *1830*, 148812. https://doi.org/10.1016/j.brainres.2024.148812

Marcus, D. S., Olsen, T. R., Ramaratnam, M., & Buckner, R. L. (2007). The Extensible Neuroimaging Archive Toolkit: An informatics platform for managing, exploring, and sharing neuroimaging data. *Neuroinformatics*, *5*(1), 11–34. https://doi.org/10.1385/ni:5:1:11

Marra, C., Piccininni, C., Masone Iacobucci, G., Caprara, A., Gainotti, G., Costantini, E. M., Callea, A., Venneri, A., & Quaranta, D. (2021). Semantic Memory as an Early Cognitive Marker of Alzheimer's Disease: Role of Category and Phonological Verbal Fluency Tasks. *Journal of Alzheimer's Disease*, *81*(2), 619–627. https://doi.org/10.3233/JAD-201452

Masters, C. L., Bateman, R., Blennow, K., Rowe, C. C., Sperling, R. A., & Cummings, J. L. (2015). Alzheimer's disease. *Nature Reviews Disease Primers*, *1*(1), 15056. https://doi.org/10.1038/nrdp.2015.56

Matsuda, H. (2013). Voxel-based Morphometry of Brain MRI in Normal Aging and Alzheimer's Disease. *Aging and Disease*. https://www.semanticscholar.org/paper/Voxel-based-Morphometry-of-Brain-MRI-in-Normal-and-Matsuda/be351303c6bfa3086bc48db7a08582c98871a04c

Mattke, S., Cho, S. K., Bittner, T., Hlávka, J., & Hanson, M. (2020). Blood-based biomarkers for Alzheimer's pathology and the diagnostic process for a disease-modifying treatment: Projecting the impact on the cost and wait times. *Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring*, *12*(1). https://doi.org/10.1002/dad2.12081

Mattsson, N., Andreasson, U., Zetterberg, H., Blennow, K., & for the Alzheimer's Disease Neuroimaging Initiative. (2017). Association of Plasma Neurofilament Light With Neurodegeneration in Patients With Alzheimer Disease. *JAMA Neurology*, *74*(5), 557. https://doi.org/10.1001/jamaneurol.2016.6117

McDonnell, M., Dill, L., Panos, S., Amano, S., Brown, W., Giurgius, S., Small, G., & Miller, K. (2020). Verbal Fluency as a Screening Tool for Mild Cognitive Impairment. *International Psychogeriatrics*, *32*(9), 1055–1062. https://doi.org/10.1017/S1041610219000644

McKhann, G. M., Knopman, D. S., Chertkow, H., Hyman, B. T., Jack, C. R., Kawas, C. H., Klunk, W. E., Koroshetz, W. J., Manly, J. J., Mayeux, R., Mohs, R. C., Morris, J. C., Rossor, M. N., Scheltens, P., Carrillo, M. C., Thies, B., Weintraub, S., & Phelps, C. H. (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. *Alzheimers Dement*, *7*(3), 263–269. https://doi.org/10.1016/j.jalz.2011.03.005

Meng, X. F., Yu, J. T., Wang, H. F., Tan, M. S., Wang, C., Tan, C. C., & Tan, L. (2014). Midlife vascular risk factors and the risk of Alzheimer's disease: A systematic review and meta-analysis. *J Alzheimers Dis*, *42*(4), 1295–1310. https://doi.org/10.3233/jad-140954

Messina, D., Cerasa, A., Condino, F., Arabia, G., Novellino, F., Nicoletti, G., Salsone, M., Morelli, M., Lanza, P. L., & Quattrone, A. (2011). Patterns of brain atrophy in Parkinson's disease, progressive supranuclear palsy and multiple system atrophy. *Parkinsonism & Related Disorders*, *17*(3), 172–176. https://doi.org/10.1016/j.parkreldis.2010.12.010

Mignon, L., Kordasiewicz, H., Lane, R., Smith, A., Miller, T., Narayanan, P., Swayze, E., Norris, D., Fitzsimmons, B., & Bennett, F. (2018). Design of the First-in-Human Study of IONIS-MAPTRx, a Tau-lowering Antisense Oligonucleotide, in Patients With Alzheimer Disease (S2.006). *Neurology*, *90*(15 Supplement), S2.006.

Mirheidari, B., Blackburn, D., Harkness, K., Walker, T., Venneri, A., Reuber, M., & Christensen, H. (2017). Toward the Automation of Diagnostic Conversation Analysis in Patients with Memory Complaints. *J Alzheimers Dis*, *58*(2), 373–387. https://doi.org/10.3233/JAD-160507

Mirheidari, B., Blackburn, D., Reuber, M., Walker, T., & Christensen, H. (2016, September 8). *Diagnosing people with dementia using automatic conversation analysis* [Proceedings Paper]. Interspeech. Proceedings of Interspeech; ISCA. http://dx.doi.org/10.21437/Interspeech.2016-857

Mirheidari, B., Blackburn, D., Walker, T., Reuber, M., & Christensen, H. (2019). Dementia detection using automatic analysis of conversations. *Computer Speech & Language*, *53*, 65–79. https://doi.org/10.1016/j.csl.2018.07.006

Mitchell, A. J., & Shiri-Feshki, M. (2009). Rate of progression of mild cognitive impairment to dementia—Meta-analysis of 41 robust inception cohort studies. *Acta Psychiatr Scand*, *119*(4), 252–265. https://doi.org/10.1111/j.1600-0447.2008.01326.x

Mohammad, S. (2018). Obtaining Reliable Human Ratings of Valence, Arousal, and Dominance for 20,000 English Words. *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 174–184. https://doi.org/10.18653/v1/P18-1017

Monsch, A. U., Seifritz, E., Taylor, K. I., Ermini-Fünfschilling, D., Stähelin, H. B., & Spiegel, R. (1997a). Category fluency is also predominantly affected in

Swiss Alzheimer's disease patients. *Acta Neurologica Scandinavica*, *95*(2), 81–84. https://doi.org/10.1111/j.1600-0404.1997.tb00073.x

Monsch, A. U., Seifritz, E., Taylor, K. I., Ermini-Fünfschilling, D., Stähelin, H. B., & Spiegel, R. (1997b). Category fluency is also predominantly affected in Swiss Alzheimer's disease patients. *Acta Neurologica Scandinavica*, *95*(2), 81–84. https://doi.org/10.1111/j.1600-0404.1997.tb00073.x

Mullane, K., & Williams, M. (2020). Alzheimer's disease beyond amyloid: Can the repetitive failures of amyloid-targeted therapeutics inform future approaches to dementia drug discovery? *Biochemical Pharmacology*, *177*, 113945. https://doi.org/10.1016/j.bcp.2020.113945

Muraki, E. J., Abdalla, S., Brysbaert, M., & Pexman, P. M. (2023). Concreteness ratings for 62,000 English multiword expressions. *Behavior Research Methods*, *55*(5), 2522–2531. https://doi.org/10.3758/s13428-022-01912-6

NICE rejects second Alzheimer's disease treatment for NHS use. (2024, October 24). *The Pharmaceutical Journal*. https://pharmaceutical-journal.com/article/news/nice-rejects-second-alzheimers-disease-treatment-for-nhs-use

Niidome, T., Ishikawa, Y., Ogawa, T., Nakagawa, M., & Nakamura, Y. (2024). Mechanism of action and clinical trial results of Lecanemab (Leqembi® 200 mg, 500 mg for Intravenous Infusion), a novel treatment for Alzheimer's disease. *Folia Pharmacologica Japonica*, *159*(3), 173–181. https://doi.org/10.1254/fpj.24005

Ohman, A., Sheppard, C., Monetta, L., & Taler, V. (2022). Assessment of semantic memory in mild cognitive impairment: The psychometric properties of a novel semantic battery. *Applied Neuropsychology. Adult*, *29*(4), 492–498. https://doi.org/10.1080/23279095.2020.1774885

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. https://doi.org/10.1007/s40336-018-0290-y

Olsson, B., Lautner, R., Andreasson, U., Öhrfelt, A., Portelius, E., Bjerke, M., Hölttä, M., Rosén, C., Olsson, C., Strobel, G., Wu, E., Dakin, K., Petzold, M., Blennow, K., & Zetterberg, H. (2016). CSF and blood biomarkers for the diagnosis of Alzheimer's disease: A systematic review and meta-analysis. *Lancet Neurol*, *15*(7), 673–684. https://doi.org/10.1016/s1474-4422(16)00070-3

O'Malley, R. P. D., Mirheidari, B., Harkness, K., Reuber, M., Venneri, A., Walker, T., Christensen, H., & Blackburn, D. (2020). Fully automated cognitive screening tool based on assessment of speech and language. *Journal of Neurology, Neurosurgery, and Psychiatry*, jnnp-2019-322517. https://doi.org/10.1136/jnnp-2019-322517

Ossenkoppele, R., Schonhaut, D. R., Schöll, M., Lockhart, S. N., Ayakta, N., Baker, S. L., O'Neil, J. P., Janabi, M., Lazaris, A., Cantwell, A., Vogel, J., Santos, M., Miller, Z. A., Bettcher, B. M., Vossel, K. A., Kramer, J. H.,

Gorno-Tempini, M. L., Miller, B. L., Jagust, W. J., & Rabinovici, G. D. (2016). Tau PET patterns mirror clinical and neuroanatomical variability in Alzheimer's disease. *Brain*, *139*(5), 1551–1567. https://doi.org/10.1093/brain/aww027

Ozcelikay-Akyildiz, G., Karadurmus, L., Cetinkaya, A., Uludag, İ., Ozcan, B., Unal, M. A., Sezginturk, M. K., & Ozkan, S. A. (2024). The Evaluation of Clinical Applications for the Detection of the Alzheimer's Disease Biomarker GFAP. *Critical Reviews in Analytical Chemistry*, 1–12. https://doi.org/10.1080/10408347.2024.2393874

Paek, E. J. (2021a). Emotional Valence Affects Word Retrieval During Verb Fluency Tasks in Alzheimer's Dementia. *Frontiers in Psychology*, *12*, 777116. https://doi.org/10.3389/fpsyg.2021.777116

Paek, E. J. (2021b). Emotional Valence Affects Word Retrieval During Verb Fluency Tasks in Alzheimer's Dementia. *Frontiers in Psychology*, *12*. https://doi.org/10.3389/fpsyg.2021.777116

Pakhomov, S. V. S., & Hemmy, L. S. (2014). A computational linguistic measure of clustering behavior on semantic verbal fluency task predicts risk of future dementia in the nun study. *Cortex; a Journal Devoted to the Study of the Nervous System and Behavior*, *55*, 97–106. https://doi.org/10.1016/j.cortex.2013.05.009

Pakhomov, S. V. S., Marino, S. E., Banks, S., & Bernick, C. (2015). Using automatic speech recognition to assess spoken responses to cognitive

tests of semantic verbal fluency. *Speech Communication*, *75*, 14–26. https://doi.org/10.1016/j.specom.2015.09.010

Palmer, A. M. (2003). Cholinergic therapies for Alzheimer's disease: Progress and prospects. *Current Opinion in Investigational Drugs (London, England: 2000)*, *4*(7), 820–825.

Palmqvist, S., Janelidze, S., Quiroz, Y. T., Zetterberg, H., Lopera, F., Stomrud, E., Su, Y., Chen, Y., Serrano, G. E., Leuzy, A., Mattsson-Carlgren, N., Strandberg, O., Smith, R., Villegas, A., Sepulveda-Falla, D., Chai, X., Proctor, N. K., Beach, T. G., Blennow, K., … Hansson, O. (2020). Discriminative Accuracy of Plasma Phospho-tau217 for Alzheimer Disease vs Other Neurodegenerative Disorders. *JAMA*, *324*(8), 772–781. https://doi.org/10.1001/jama.2020.12134

Palmqvist, S., Mattsson, N., & Hansson, O. (2016). Cerebrospinal fluid analysis detects cerebral amyloid-β accumulation earlier than positron emission tomography. *Brain*, *139*(Pt 4), 1226–1236. https://doi.org/10.1093/brain/aww015

Palop, J. J., & Mucke, L. (2010). Amyloid-β–induced neuronal dysfunction in Alzheimer's disease: From synapses toward neural networks. *Nature Neuroscience*, *13*(7), 812–818. https://doi.org/10.1038/nn.2583

Papp, K. V., Mormino, E. C., Amariglio, R. E., Munro, C., Dagley, A., Schultz, A. P., Johnson, K. A., Sperling, R. A., & Rentz, D. M. (2016). Biomarker validation of a decline in semantic processing in preclinical Alzheimer's

disease. *Neuropsychology*, *30*(5), 624–630. https://doi.org/10.1037/neu0000246

Parnetti, L., & Eusebi, P. (2018). Cerebrospinal Fluid Biomarkers in Alzheimer's Disease: An Invaluable Tool for Clinical Diagnosis and Trial Enrichment. *Journal of Alzheimer's Disease*, *64*(s1), S281–S287. https://doi.org/10.3233/JAD-179910

Pekkala, S., Albert, M. L., Spiro, A., & Erkinjuntti, T. (2008). Perseveration in Alzheimer's disease. *Dementia and Geriatric Cognitive Disorders*, *25*(2), 109–114. https://doi.org/10.1159/000112476

Pérez, G. N., Ponferrada, J., Ferrante, F. J., Caro, I., Bisé, J. V., Migeot, J., Welford, A. S., Olavarria, L., Pelle, P., Chonchol, A. S., Ferrer, L., & Garcia, A. M. (2024). Automated linguistic and acoustic measures of verbal fluency as neurocognitive markers of mild cognitive impairment. *Alzheimer's & Dementia*, *20*(S2), e086150. https://doi.org/10.1002/alz.086150

Petersen, R. C. (2004). Mild cognitive impairment as a diagnostic entity. *J Intern Med*, *256*(3), 183–194. https://doi.org/10.1111/j.1365-2796.2004.01388.x

Petersen, R. C., Caracciolo, B., Brayne, C., Gauthier, S., Jelic, V., & Fratiglioni, L. (2014). Mild cognitive impairment: A concept in evolution. *J Intern Med*, *275*(3), 214–228.

Petersen, R. C. & Mayo Alzheimer's Disease Research Center. (2018). FTS3-01-03: U.S. PERSPECTIVE ON CLINICAL AMYLOID IMAGING. *Alzheimer's & Dementia*, *14*(7S_Part_18). https://doi.org/10.1016/j.jalz.2018.06.2758

Pexman, P. M., Muraki, E., Sidhu, D. M., Siakaluk, P. D., & Yap, M. J. (2019). Quantifying sensorimotor experience: Body–object interaction ratings for more than 9,000 English words. *Behavior Research Methods*, *51*(2), 453–466. https://doi.org/10.3758/s13428-018-1171-z

Plassman, B. L., Langa, K. M., Fisher, G. G., Heeringa, S. G., Weir, D. R., Ofstedal, M. B., Burke, J. R., Hurd, M. D., Potter, G. G., Rodgers, W. L., Steffens, D. C., McArdle, J. J., Willis, R. J., & Wallace, R. B. (2008). Prevalence of cognitive impairment without dementia in the United States. *Annals of Internal Medicine*, *148*(6), 427–434.

Pobric, G., Jefferies, E., & Lambon Ralph, M. A. (2010). Category-specific versus category-general semantic impairment induced by transcranial magnetic stimulation. *Current Biology: CB*, *20*(10), 964–968. https://doi.org/10.1016/j.cub.2010.03.070

Poewe, W., Gauthier, S., Aarsland, D., Leverenz, J. B., Barone, P., Weintraub, D., Tolosa, E., & Dubois, B. (2008). Diagnosis and management of Parkinson's disease dementia. *International Journal of Clinical Practice*, *62*(10), 1581–1587. https://doi.org/10.1111/j.1742-1241.2008.01869.x

Prados Carrasco, F., Cardoso, M. J., Burgos, N., Wheeler-Kingshott, C. A. M., & Ourselin, S. (2016, May 13). *NiftyWeb: Web based platform for image*

processing on the cloud [Proceedings paper]. 24th Scientific Meeting and Exhibition of the International Society for Magnetic Resonance in Medicine (ISMRM). In: Proceedings of the 24th Scientific Meeting and Exhibition of the International Society for Magnetic Resonance in Medicine (ISMRM). International Society for Magnetic Resonance in Medicine (ISMRM): Singapore. (2016); International Society for Magnetic Resonance in Medicine (ISMRM). https://doi.org/10/1/Submit-niftyweb.pdf

Prince, M., Wimo, A., Guerchet, M., Ali, G.-C., Wu, Y.-T., & Prina, M. (2015). *World Alzheimer Report 2015. The Global Impact of Dementia. An Analysis of Prevalence, Incidence, Cost and Trends*.

Radua, J., Vieta, E., Shinohara, R., Kochunov, P., Quidé, Y., Green, M. J., Weickert, C. S., Weickert, T., Bruggemann, J., Kircher, T., Nenadić, I., Cairns, M. J., Seal, M., Schall, U., Henskens, F., Fullerton, J. M., Mowry, B., Pantelis, C., Lenroot, R., … Pineda-Zapata, J. (2020). Increased power by harmonizing structural MRI site differences with the ComBat batch adjustment method in ENIGMA. *NeuroImage*, *218*, 116956. https://doi.org/10.1016/j.neuroimage.2020.116956

Raj, A., Tosun, D., & Weiner, M. W. (2016). O4-07-02: NETWORK TRANSMISSION MODEL RECAPITULATES AMYLOID AND TAU SPREAD AND PREDICTS IMAGING FINDINGS. *Alzheimer's & Dementia*, *12*(7S_Part_7). https://doi.org/10.1016/j.jalz.2016.06.644

Reitz, C. (2012). Alzheimer's Disease and the Amyloid Cascade Hypothesis: A Critical Review. *International Journal of Alzheimer's Disease*, *2012*, 1–11. https://doi.org/10.1155/2012/369808

Reuber, M., Blackburn, D. J., Elsey, C., Wakefield, S., Ardern, K. A., Harkness, K., Venneri, A., Jones, D., Shaw, C., & Drew, P. (2018). An Interactional Profile to Assist the Differential Diagnosis of Neurodegenerative and Functional Memory Disorders. *Alzheimer Disease and Associated Disorders*, *32*(3), 197–206. https://doi.org/10.1097/WAD.0000000000000231

Rice, D. P., Fox, P. J., Max, W., Webber, P. A., Hauck, W. W., Lindeman, D. A., & Segura, E. (1993). The Economic Burden of Alzheimer's Disease Care. *Health Affairs*, *12*(2), 164–176. https://doi.org/10.1377/hlthaff.12.2.164

Robinson, J. L., Yan, N., Caswell, C., Xie, S. X., Suh, E., Van Deerlin, V. M., Gibbons, G., Irwin, D. J., Grossman, M., Lee, E. B., Lee, V. M.-Y., Miller, B., & Trojanowski, J. Q. (2020). Primary Tau Pathology, Not Copathology, Correlates With Clinical Symptoms in PSP and CBD. *Journal of Neuropathology & Experimental Neurology*, *79*(3), 296–304. https://doi.org/10.1093/jnen/nlz141

Rocher, A. B., Chapon, F., Blaizot, X., Baron, J. C., & Chavoix, C. (2003). Resting-state brain glucose utilization as measured by PET is directly related to regional synaptophysin levels: A study in baboons. *Neuroimage*, *20*(3), 1894–1898. https://doi.org/10.1016/j.neuroimage.2003.07.002

Rogers, S. L., Doody, R. S., Pratt, R. D., & Ieni, J. R. (2000). Long-term efficacy and safety of donepezil in the treatment of Alzheimer's disease: Final analysis of a US multicentre open-label study. *European Neuropsychopharmacology*, *10*(3), 195–203. https://doi.org/10.1016/S0924-977X(00)00067-5

Rong, X.-F., & Wang, X.-L. (2012). [An overview of biomarkers in Alzheimer's disease]. *Yao xue xue bao = Acta pharmaceutica Sinica*, *47*(5), 551–557.

Rosenmann, H., Blum, D., Kayed, R., & Ittner, L. M. (2012). Tau Protein: Function and Pathology. *International Journal of Alzheimer's Disease*, *2012*, 1–2. https://doi.org/10.1155/2012/707482

Roytman, M., Mashriqi, F., Al-Tawil, K., Schulz, P. E., Zaharchuk, G., Benzinger, T. L. S., & Franceschi, A. M. (2023). Amyloid-Related Imaging Abnormalities: An Update. *American Journal of Roentgenology*, *220*(4), 562–574. https://doi.org/10.2214/AJR.22.28461

Sapir-Pichhadze, R., & Kaplan, B. (2020). Seeing the Forest for the Trees: Random Forest Models for Predicting Survival in Kidney Transplant Recipients. *Transplantation*, *104*(5), 905–906. https://doi.org/10.1097/TP.0000000000002923

Scheltens, P., Blennow, K., Breteler, M. M., de Strooper, B., Frisoni, G. B., Salloway, S., & Van der Flier, W. M. (2016). Alzheimer's disease. *Lancet*, *388*(10043), 505–517. https://doi.org/10.1016/s0140-6736(15)01124-1

Selkoe, D. J., & Hardy, J. (2016). The amyloid hypothesis of Alzheimer's disease at 25 years. *EMBO Molecular Medicine*, *8*(6), 595–608. https://doi.org/10.15252/emmm.201606210

Siakaluk, P. D., Pexman, P. M., Sears, C. R., Wilson, K., Locheed, K., & Owen, W. J. (2008). The benefits of sensorimotor knowledge: Body-object interaction facilitates semantic processing. *Cognitive Science*, *32*(3), 591–605. https://doi.org/10.1080/03640210802035399

Sims, J. R., Zimmer, J. A., Evans, C. D., Lu, M., Ardayfio, P., Sparks, J., Wessels, A. M., Shcherbinin, S., Wang, H., Monkul Nery, E. S., Collins, E. C., Solomon, P., Salloway, S., Apostolova, L. G., Hansson, O., Ritchie, C., Brooks, D. A., Mintun, M., Skovronsky, D. M., & TRAILBLAZER-ALZ 2 Investigators. (2023). Donanemab in Early Symptomatic Alzheimer Disease: The TRAILBLAZER-ALZ 2 Randomized Clinical Trial. *JAMA*, *330*(6), 512–527. https://doi.org/10.1001/jama.2023.13239

Slooter, A. J. C., Cruts, M., Kalmijn, S., Hofman, A., Breteler, M. M. B., Van Broeckhoven, C., & Van Duijn, C. M. (1998). Risk Estimates of Dementia by Apolipoprotein E Genotypes From a Population-Based Incidence Study: The Rotterdam Study. *Archives of Neurology*, *55*(7), 964. https://doi.org/10.1001/archneur.55.7.964

Snowdon, D. A., Kemper, S. J., Mortimer, J. A., Greiner, L. H., Wekstein, D. R., & Markesbery, W. R. (1996). Linguistic ability in early life and cognitive

function and Alzheimer's disease in late life. Findings from the Nun Study. *Jama*, *275*(7), 528–532.

Snowdon, D. A. & Nun Study. (2003). Healthy aging and dementia: Findings from the Nun Study. *Annals of Internal Medicine*, *139*(5 Pt 2), 450–454. https://doi.org/10.7326/0003-4819-139-5_part_2-200309021-00014

Sperling, R. A., Aisen, P. S., Beckett, L. A., Bennett, D. A., Craft, S., Fagan, A. M., Iwatsubo, T., Jack, C. R., Kaye, J., Montine, T. J., Park, D. C., Reiman, E. M., Rowe, C. C., Siemers, E., Stern, Y., Yaffe, K., Carrillo, M. C., Thies, B., Morrison-Bogorad, M., … Phelps, C. H. (2011). Toward defining the preclinical stages of Alzheimer's disease: Recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimers Dement*, *7*(3), 280–292. https://doi.org/10.1016/j.jalz.2011.03.003

Squire, L. R. (1992). Declarative and Nondeclarative Memory: Multiple Brain Systems Supporting Learning and Memory. *Journal of Cognitive Neuroscience*, *4*(3), 232–243. https://doi.org/10.1162/jocn.1992.4.3.232

Sutton, T. M., Herbert, A. M., & Clark, D. Q. (2019). Valence, arousal, and dominance ratings for facial stimuli. *Quarterly Journal of Experimental Psychology*, *72*(8), 2046–2055. https://doi.org/10.1177/1747021819829012

Tanabe, J., Lim, M. F., Dash, S., Pattee, J., Steach, B., Pressman, P., Bettcher, B. M., Honce, J. M., Potigailo, V. A., Colantoni, W., Zander, D., & Thaker,

A. A. (2024). Automated Volumetric Software in Dementia: Help or Hindrance to the Neuroradiologist? *American Journal of Neuroradiology*, *45*(11), 1737–1744. https://doi.org/10.3174/ajnr.A8406

Tay, L. X., Ong, S. C., Tay, L. J., Ng, T., & Parumasivam, T. (2024). Economic Burden of Alzheimer's Disease: A Systematic Review. *Value in Health Regional Issues*, *40*, 1–12. https://doi.org/10.1016/j.vhri.2023.09.008

Tetzloff, K. A., Graff-Radford, J., Martin, P. R., Tosakulwong, N., Machulda, M. M., Duffy, J. R., Clark, H. M., Senjem, M. L., Schwarz, C. G., Spychalla, A. J., Drubach, D. A., Jack, C. R., Lowe, V. J., Josephs, K. A., & Whitwell, J. L. (2018). Regional Distribution, Asymmetry, and Clinical Correlates of Tau Uptake on [^{18} F]AV-1451 PET in Atypical Alzheimer's Disease. *Journal of Alzheimer's Disease*, *62*(4), 1713–1724. https://doi.org/10.3233/JAD-170740

Thal, D. R., & Tomé, S. O. (2022). The central role of tau in Alzheimer's disease: From neurofibrillary tangle maturation to the induction of cell death. *Brain Research Bulletin*, *190*, 204–217. https://doi.org/10.1016/j.brainresbull.2022.10.006

Thompson-Schill, S. L., D'Esposito, M., Aguirre, G. K., & Farah, M. J. (1997). Role of left inferior prefrontal cortex in retrieval of semantic knowledge: A reevaluation. *Proceedings of the National Academy of Sciences*, *94*(26), 14792–14797. https://doi.org/10.1073/pnas.94.26.14792

Troyer, A. K., Moscovitch, M., Winocur, G., Alexander, M. P., & Stuss, D. (1998). Clustering and switching on verbal fluency: The effects of focal frontal- and temporal-lobe lesions. *Neuropsychologia*, *36*(6), 499–504. https://doi.org/10.1016/s0028-3932(97)00152-8

Troyer, A. K., Moscovitch, M., Winocur, G., Leach, L., & Freedman, M. (1998a). Clustering and switching on verbal fluency tests in Alzheimer's and Parkinson's disease. *Journal of the International Neuropsychological Society: JINS*, *4*(2), 137–143. https://doi.org/10.1017/s1355617798001374

Troyer, A. K., Moscovitch, M., Winocur, G., Leach, L., & Freedman, M. (1998b). Clustering and switching on verbal fluency tests in Alzheimer's and Parkinson's disease. *Journal of the International Neuropsychological Society: JINS*, *4*(2), 137–143. https://doi.org/10.1017/s1355617798001374

Tulving, E. (1972). Episodic and semantic memory. In *Organization of memory* (pp. xiii, 423–xiii, 423). Academic Press.

Tzioras, M., Davies, C., Newman, A., Jackson, R., & Spires-Jones, T. (2019). Invited Review: APOE at the interface of inflammation, neurodegeneration and pathological protein spread in Alzheimer's disease. *Neuropathology and Applied Neurobiology*, *45*(4), 327–346. https://doi.org/10.1111/nan.12529

Van Dyck, C. H., Swanson, C. J., Aisen, P., Bateman, R. J., Chen, C., Gee, M., Kanekiyo, M., Li, D., Reyderman, L., Cohen, S., Froelich, L., Katayama, S., Sabbagh, M., Vellas, B., Watson, D., Dhadda, S., Irizarry, M., Kramer,

L. D., & Iwatsubo, T. (2023). Lecanemab in Early Alzheimer's Disease. *New England Journal of Medicine*, *388*(1), 9–21. https://doi.org/10.1056/NEJMoa2212948

van Heuven, W. J. B., Mandera, P., Keuleers, E., & Brysbaert, M. (2014). SUBTLEX-UK: A new and improved word frequency database for British English. *Quarterly Journal of Experimental Psychology (2006)*, *67*(6), 1176–1190. https://doi.org/10.1080/17470218.2013.850521

VandeVrede, L., & Schindler, S. E. (2024). Clinical use of biomarkers in the era of Alzheimer's disease treatments. *Alzheimer's & Dementia*, *21*(1), e14201. https://doi.org/10.1002/alz.14201

Vargha-Khadem, F., Gadian, D. G., Watkins, K. E., Connelly, A., Van Paesschen, W., & Mishkin, M. (1997). Differential Effects of Early Hippocampal Pathology on Episodic and Semantic Memory. *Science*, *277*(5324), 376–380. https://doi.org/10.1126/science.277.5324.376

Vemuri, P., & Jack, C. R. (2010). Role of structural MRI in Alzheimer's disease. *Alzheimer's Research & Therapy*, *2*(4), 23. https://doi.org/10.1186/alzrt47

Vogel, A., Gade, A., Stokholm, J., & Waldemar, G. (2005). Semantic Memory Impairment in the Earliest Phases of Alzheimer's Disease. *Dementia and Geriatric Cognitive Disorders*, *19*(2–3), 75–81. https://doi.org/10.1159/000082352

Vonk, J. M. J., Flores, R. J., Rosado, D., Qian, C., Cabo, R., Habegger, J., Louie, K., Allocco, E., Brickman, A. M., & Manly, J. J. (2019). Semantic network

function captured by word frequency in nondemented APOE ε4 carriers. *Neuropsychology*, *33*(2), 256–262. https://doi.org/10.1037/neu0000508

Walt, G. (2004). WHO's World Health Report 2003. *Bmj*, *328*(7430), 6. https://doi.org/10.1136/bmj.328.7430.6

Wang, X., Su, B., Perry, G., Smith, M. A., & Zhu, X. (2007). Insights into amyloid-beta-induced mitochondrial dysfunction in Alzheimer disease. *Free Radical Biology & Medicine*. https://www.semanticscholar.org/paper/Insights-into-amyloid-beta-induced-mitochondrial-in-Wang-Su/018620fafa7f50fe746571f6ba68d0c9ffc87966

Warriner, A. B., Kuperman, V., & Brysbaert, M. (2013). Norms of valence, arousal, and dominance for 13,915 English lemmas. *Behavior Research Methods*, *45*(4), 1191–1207. https://doi.org/10.3758/s13428-012-0314-x

Westman, E., Muehlboeck, J.-S., & Simmons, A. (2012). Combining MRI and CSF measures for classification of Alzheimer's disease and prediction of mild cognitive impairment conversion. *NeuroImage*, *62*(1), 229–238. https://doi.org/10.1016/j.neuroimage.2012.04.056

Whitwell, J. L. (2009). Voxel-Based Morphometry: An Automated Technique for Assessing Structural Changes in the Brain. *Journal of Neuroscience*, *29*(31), 9661–9664. https://doi.org/10.1523/JNEUROSCI.2160-09.2009

Willis, B. A., Bhagunde, P., Bell, R., Penner, N., Charil, A., Irizarry, M. C., Hersch, S., & Reyderman, L. (2024). Amyloid Plaque Reduction as a

Surrogate Marker of Efficacy: Assessment of Amyloid PET and Change in CDR-SB Utilizing Semi-Mechanistic Model. *Alzheimer's & Dementia*, *20*(S6), e091955. https://doi.org/10.1002/alz.091955

Wu, L., Rosa-Neto, P., Hsiung, G.-Y. R., Sadovnick, A. D., Masellis, M., Black, S. E., Jia, J., & Gauthier, S. (2012). Early-Onset Familial Alzheimer's Disease (EOFAD). *Canadian Journal of Neurological Sciences / Journal Canadien Des Sciences Neurologiques*, *39*(4), 436–445. https://doi.org/10.1017/S0317167100013949

Xu, W., Tan, L., Wang, H. F., Jiang, T., Tan, M. S., Tan, L., Zhao, Q. F., Li, J. Q., Wang, J., & Yu, J. T. (2015). Meta-analysis of modifiable risk factors for Alzheimer's disease. *J Neurol Neurosurg Psychiatry*, *86*(12), 1299–1306. https://doi.org/10.1136/jnnp-2015-310548

Zimmer, J. A., Ardayfio, P., Wang, H., Khanna, R., Evans, C. D., Lu, M., Sparks, J., Andersen, S., Lauzon, S., Nery, E. S. M., Battioui, C., Engle, S. E., Biffi, A., Svaldi, D., Salloway, S., Greenberg, S. M., Sperling, R. A., Mintun, M., Brooks, D. A., & Sims, J. R. (2025). Amyloid-Related Imaging Abnormalities With Donanemab in Early Symptomatic Alzheimer Disease: Secondary Analysis of the TRAILBLAZER-ALZ and ALZ 2 Randomized Clinical Trials. *JAMA Neurology*, *82*(5), 461–469. https://doi.org/10.1001/jamaneurol.2025.0065