

Novel precision biomarker models applied to Alzheimer's disease

Isaac Llorente Sager

A dissertation submitted in partial fulfilment
of the requirements for the degree of
Doctor of Philosophy
of
University College London.

Department of Med Phys and Biomedical Eng
University College London

10 September 2025

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

Abstract

Biomarkers are essential for diagnosing, monitoring, and developing treatments for Alzheimer's disease (AD). However, conventional biomarkers for tau pathology (PET), brain atrophy (MRI), and amyloid-beta ($A\beta$) peptides often rely on simple, pre-defined ratios that limit precision, stability, and statistical power, hindering clinical trial efficiency and research progress.

This thesis addresses these challenges by introducing and validating BioDisCVR, a novel, modality-agnostic framework for the data-driven discovery of optimised biomarkers. The work first establishes the methodological fragility of the gold-standard Standardised Uptake Value Ratio (SUVR) in tau PET, exposing the instability of commonly used reference regions and the detrimental effects of certain processing techniques. It further reveals the suboptimal performance of traditional volumetric measures for tracking longitudinal change.

One central innovation is the Composite Value Ratio (CVR), a data-driven biomarker construct where both numerator and denominator are optimised to maximise statistical power. Applied to tau PET and structural MRI, CVR demonstrates transformative improvements over established methods, with the potential to reduce clinical trial sample sizes by over 79%, and improve detection of pathological changes. The framework is extended to proteomics with a weighted CVR (wCVR), providing novel insights into the pathogenic contributions of different $A\beta$ peptides. This culminates in theta, a parameter-free multidimensional ratio that achieved outstanding classification of familial AD mutations ($AUC > 0.99$), dramatically outperforming all conventional peptide ratios.

Overall, this thesis delivers a new paradigm for biomarker discovery. It provides a validated framework and a suite of statistically superior biomarkers to accelerate clinical trials, enhance disease monitoring, and advance the fundamental under-

standing of Alzheimer's disease.

Impact Statement

The findings and contributions of this PhD thesis could have significant implications across academia, industry, and clinical practice:

1. **Academic Impact:** This thesis provides the field with three distinct contributions. First, it exposes fundamental vulnerabilities in established neuroimaging methods, demonstrating the instability of conventional SUVR reference regions and the detrimental effects of certain partial-volume correction techniques in tau PET. Second, it delivers the BioDisCVR framework, a novel, open-source toolkit for the data-driven discovery of optimised biomarkers. Finally, it introduces theta, a parameter-free multidimensional ratio that generalises the ratio concept to high-dimensional data. Together, these contributions provide researchers with not only superior biomarkers (CVR, wCVR, theta) but also the foundational tools and critical insights needed to improve the rigour and reproducibility of neurodegeneration research across modalities.
2. **Biopharmaceutical and Medical Imaging Industry:** These enhanced biomarkers could be pivotal for biopharmaceutical companies and medical imaging firms, especially in optimising clinical trials for neurodegenerative diseases. By reducing the sample size estimates required (by over 79% as demonstrated in tau PET trials), they can significantly lower costs and accelerate timelines for drug development, enabling faster progression of promising therapies to market. Alternatively, for studies where larger samples are necessary, the increased sensitivity of these biomarkers enables more efficient effect detection, benefiting the development and assessment of new treatments. Furthermore, medical imaging companies could integrate these novel biomarker methodologies into their software platforms, offering advanced analytical tools for researchers and clinicians.

3. **Clinical Applications:** While the tools developed are primarily for research, they chart a clear path toward future clinical translation. The superior stability and precision of CVR for tracking longitudinal change in both MRI and PET offer the potential for more reliable monitoring of individual disease progression than current clinical measures. For diagnostics, the theta biomarker's ability to perfectly distinguish pathogenic mutations from controls in iPSC models represents a breakthrough in summarising complex molecular data, providing a powerful new method for developing highly accurate fluid-based diagnostic tests that go beyond two measures.

Acknowledgements

I can't possibly list everyone and everything that contributed one way or another to the existence of this thesis. That being said, this would not have been possible without my grandma, Àvia Carme, who nurtured us and allowed us to accompany her during her last journey, while she was living with Alzheimer's disease. While I didn't answer all the questions I had at that moment, I surely learned a lot during this PhD.

I am very grateful to have had the emotional support of my parents, Elisabet and Ramon, who watched my every step of this trip into the unknown, and endured some of my talks. I want to thank my supervisor and colleague Neil, from whom I learned a lot, and who has been key to everything presented here, as he was always open and constructively critical to my wild ideas and exploration, and whose awareness of past and current research in AD never ceases to amaze me. My gratitude extends to the nice environment in the POND research group, with friendly moments and talks with the many integrants (Zeena, Sonja, Alex, Ellie, Anna, Lawrence, Maitrei, Gonzalo, Tiantian, Liza, An, Bárbara...), and many 90HH colleagues (Stefano, Gabriel, Marta, Fran, Athena...). The feedback and talks I had from all colleagues (Neil, Marc, Alex, David, Sonja, Gabriel, Charlie, Rebecca, Jonathan, Dave, Frederik, Gabriel, Pawel...) were like alchemy to my work, undoubtedly increasing its worth. Everything is a notch easier with nice collaborators (Neil, Antoine, Pierrick, Charlie, Rebecca). I'm grateful to my thesis panel, Professor Gary Zhang and Dr Timothy Rittman, for their valuable feedback and insightful discussion.

Participating in the Alzheimer's Association International Conference was truly inspiring every single time. And the many interactions I had there, exposing my work, were always a nice boost of confidence. So much work presented there was only possible thanks to the Alzheimer's Disease Neuroimaging Initiative, so thank you!

Institutionally, I am thankful to have had the chance to be in the i4health CDT, the Centre for Medical Image Computing (CMIC), now Hawkes Institute (together with the Wellcome / EPSRC Centre for Interventional and Surgical Sciences -WEISS), under the impressive umbrella institution that is University College London (UCL). The seminars and interest groups are a delicious hot pot of constant innovation and mind-teasers, so it is easy to get into a good research mindset. My application here was possible thanks to having had the two-year intensive MSc of Medical Imaging and Applications (Girona, Le Creusot and Cassino), so thank you, professors (Xavi, Arnau, Robert, Xevi, Edu, Desiré...) and colleagues.

I am also thankful for the social network of family (Marina, germans, cosins, tietes i tiets, àvies i avis...) and friends (Martí, Marc, Albert, Marta, Lúdia, Raquel, Laura, Òscar, Aina, Èlia, Ester, Èlia, NMP...), and the healthy interactions with strangers that makes one feel alive. Gràcies, Aleix, pels ànims i recolzament continuat en aquesta aventura! Gràcies, Nu :) I gràcies, natura, per compensar les masses hores que passo davant la pantalla.

UCL Research Paper Declaration Form: referencing the doctoral candidate's own published work(s)

Work 1/3: A data-driven framework for biomarker discovery

1. For a research manuscript that has already been published:

(a) What is the title of the manuscript? A data-driven framework for biomarker discovery applied to optimising modern clinical and preclinical trials on Alzheimer's disease.

(b) Please include a link to or doi for the work:

<https://doi.org/10.1093/braincomms/fcae438>

(c) Where was the work published? Brain Communications journal.

(d) Who published the work? Oxford Academic.

(e) When was the work published? 09 December 2024.

(f) List the manuscript's authors in the order they appear on the publication: Isaac Llorente-Saguer, Neil P Oxtoby.

(g) Was the work peer reviewed? Yes

(h) Have you retained the copyright? Yes. And the work is under Creative Commons Attribution (CC BY) license.

(i) Was an earlier form of the manuscript uploaded to a preprint server (e.g. medRxiv)? No.

- If 'Yes', please give a link or doi:

- If 'No', please seek permission from the relevant publisher and check the box next to the below statement:

☒ *I acknowledge permission of the publisher named under 1d to include in this thesis portions of the publication named as included in 1c.*

3. For multi-authored work, please give a statement of contribution covering all authors (if single-author, please skip to section 4): Isaac Llorente-Saguer: Conceptualisation, analysis, figures, original draft. Neil P. Oxtoby: analysis and manuscript feedback and editing.

4. In which chapter(s) of your thesis can this material be found? Chapter 4.

e-Signatures confirming that the information above is accurate (this form should be co-signed by the supervisor/ senior author unless this is not appropriate, e.g. if the paper was a single-author work):

ILLS

Candidate: _____

Date: _____ 13 March 2025 _____

NPO

Supervisor/Senior Author signature (where appropriate): _____

Date: _____ 13 March 2025 _____

Work 2/3: Composite Value Ratios for atrophy monitoring in Alzheimer's Disease

2. For a research manuscript prepared for publication but that has not yet been published (if already published, please skip to section 3):

(a) What is the current title of the manuscript? Enhanced Monitoring of Alzheimer's Disease Brain Atrophy using Composite Value Ratios of Volumes

(b) Has the manuscript been uploaded to a preprint server (e.g. medRxiv)? - If 'Yes', please give a link or doi: No

(c) Where is the work intended to be published? Brain Communications journal

(d) List the manuscript's authors in the intended authorship order: Isaac Llorente-Saguer, Neil P. Oxtoby

(e) Stage of publication: Under review

3. For multi-authored work, please give a statement of contribution covering all authors (if single-author, please skip to section 4): Isaac Llorente-Saguer: Conceptualisation, analysis, figures, original draft. Neil P. Oxtoby: analysis and manuscript feedback and editing.

4. In which chapter(s) of your thesis can this material be found? Chapter 5

e-Signatures confirming that the information above is accurate (this form should be co-signed by the supervisor/ senior author unless this is not appropriate, e.g. if the paper was a single-author work):

Candidate: ILLS

Date: 13 March 2025

Supervisor/Senior Author signature (where appropriate): NPO

Date: 13 March 2025

Work 3/3: Composite Value Ratios applied to amyloid-beta peptides

2. For a research manuscript prepared for publication but that has not yet been published (if already published, please skip to section 3):

(a) What is the current title of the manuscript? Unbiased data-driven analysis of five amyloid-beta peptides for biomarker investigations in familial Alzheimer's disease

(b) Has the manuscript been uploaded to a preprint server (e.g. medRxiv)? - If 'Yes', please give a link or doi: <https://doi.org/10.1101/2024.11.23.624811>

(c) Where is the work intended to be published? Alzheimer's Research & Therapy

(d) List the manuscript's authors in the intended authorship order: Isaac Llorente-Saguer, Rebecca Gabriele, Teisha Bradshaw, Claire A. Leckey, Christopher R.S. Belder, Lucía Chávez-Gutiérrez, Rohan de Silva, Nick C. Fox, Selina Wray, Neil P. Oxtoby and Charles Arber.

(e) Stage of publication: Editor sent it for peer reviewing.

3. For multi-authored work, please give a statement of contribution covering all authors (if single-author, please skip to section 4): Isaac Llorente-Saguer: data analysis, figures, discussion, original draft. Other coauthors were involved in the creation of new data (one out of the four datasets used, the iPSC one), facilitated funding, gave feedback on the manuscript, and put the work in the context of the literature (Rebecca Gabriele, Teisha Bradshaw, Claire Leckey, Rohan de Silva, Nick C. Fox, Selina Wray, Neil Oxtoby and Charles Arber)

4. In which chapter(s) of your thesis can this material be found? Chapter 6: Weighted CVR, a case with amyloid peptides.

e-Signatures confirming that the information above is accurate (this form should be co-signed by the supervisor/ senior author unless this is not appropriate, e.g. if the paper was a single-author work):

Candidate: ILLS

Date: 13 March 2025

Supervisor/Senior Author signature (where appropriate): _____

Date: 13 March 2025

NPO

Contents

1	PhD Summary	24
1.1	Introduction	24
1.2	Statement of problem	25
1.3	Thesis contributions	25
2	Background	28
2.1	The Alzheimer's Disease "Pandemic": A Growing Health Crisis	28
2.2	Introduction to biomarkers for Alzheimer's disease	30
2.3	Neuroimaging biomarkers: Visualising Brain Changes	32
2.3.1	Computed Tomography (CT) scan.	32
2.3.2	Magnetic Resonance Imaging (MRI).	32
2.3.3	Positron-Emission Tomography (PET).	37
2.4	Fluid biomarkers	40
2.4.1	Cerebrospinal Fluid (CSF)	40
2.4.2	Blood-Based Biomarkers (BBMs)	41
2.4.3	In Vitro Cell Models	41
2.4.4	Implications	41
2.5	Cognitive measures	42
2.5.1	Conventional cognitive instruments	42
2.5.2	Composite endpoints: PACC and related composites	43
2.5.3	Digital cognitive biomarkers: active and passive approaches	44
2.5.4	Psychometric and practical considerations	44
2.5.5	Limitations, equity and ethical concerns	45
2.6	Clinical Trials in Alzheimer's Disease	45
2.6.1	Introduction	45

2.6.2	The 2025 Trial Landscape	46
2.6.3	Therapeutic Targets and Landmark Trials	46
2.6.4	Imaging Biomarkers in Alzheimer’s Disease Clinical Trials . .	47
2.7	A Unifying Framework: The Evolving Biological Definition of AD . . .	48
2.8	Summary and motivation for this thesis	49
3	The effect of the reference region in standardised uptake value ratios in tau PET	51
3.1	Introduction	51
3.2	SUVR inconsistencies	52
3.2.1	Introduction	52
3.2.2	Materials	52
3.2.3	Methods	53
3.2.4	Results	54
3.2.5	Discussion.	55
3.3	Reference Region effect on test-retest metrics	56
3.3.1	Introduction	56
3.3.2	Methods	56
3.3.3	Results	58
3.3.4	Discussion	58
3.4	Combination of multiple references for SUVR quantification	61
3.4.1	Introduction	61
3.4.2	Methods	61
3.4.3	Discussion	62
3.5	Conclusions.	63
4	Composite Value Ratio (CVR)	66
4.1	Introduction	66
4.2	Materials and Methods	67
4.2.1	Study design	67
4.2.2	Data	69
4.2.3	BioDisCVR Framework	71
4.2.4	Baselines	73

4.2.5 Statistical analysis	73
4.3 Results	75
4.4 Discussion	81
4.5 Conclusions	85
5 CVR applied to MRI to track volumetric changes	86
5.1 Introduction	86
5.2 Materials and Methods	87
5.3 Results	92
5.4 Discussion	99
5.5 Conclusions	103
6 Weighted Composite Value Ratio, a case with amyloid peptides	105
6.1 Introduction	105
6.2 Methods	106
6.2.1 Data	106
6.2.2 Modelling and analysis	107
6.3 Results	109
6.4 Discussion	117
7 Theta, a multidimensional-input, one-dimensional output ratio-based biomarker	121
7.1 Introduction	121
7.2 Material and Methods	121
7.3 Results	124
7.4 Discussion	124
7.5 Conclusions	126
8 Discussion	128
8.1 Chapter 3: Reference regions in SUVR	128
8.2 Chapter 4: Composite Value Ratio	130
8.3 Chapter 5: CVR applied to MRI	131
8.4 Chapter 6: Weighted CVR	132
8.5 Chapter 7: Theta, a multidimensional ratio	133

8.6 Synthesis of thesis Contributions and their Implications	134
9 Conclusions	136
Appendices	137
A BioDisCVR Supplementary Material	137
B CVR in MRI for atrophy monitoring Supplementary Material	147
C Weighted CVR Supplementary Material	150
D Theta Supplementary Material	154
E Colophon	155

List of Figures

3.1	Inferior cerebellar grey matter versus eroded subcortical white matter SUV	54
3.2	Future/baseline brainstem SUVR(gm) versus future/baseline eroded subcorical white matter SUV	55
3.3	“Future SUVRgm divided by past SUVRgm” against “future wm SUV divided by past wm SUV”, for all regions	55
3.4	Slope of linear regression fit to previous plots	56
3.5	Geometric representation of SUVR and its generalisation	62
4.1	Conceptual Representation of the BioDisCVR Framework	68
4.2	Biomarker performance for Experiment 1.	75
4.3	Biomarker performance for Experiment 2.	78
5.1	Conceptual abstract for the composite value ratio applied to MRI for atrophy monitoring.	88
5.2	Summary metrics report	97
5.3	Predicted vs observed values	98
6.1	Literature biomarkers applied to the newly-generated data from UCL.	110
6.2	Peptide distribution relative to the controls.	111
6.3	Example of the need for harmonisation, and the result of it.	112
6.4	Heatmap of results per biomarker, per site.	113
6.5	Visual plot of evaluation metrics for all biomarkers.	113
6.6	Heatmap of evaluation metrics.	117
7.1	Distribution of the three merged cell datasets in three dimensions, corresponding to A β 37, 40 and 42.	123

7.2	Precision-recall curves for biomarkers A β 42/40, A β 37/42, and Theta.	126
A.1	Repeatability (error) and SSE performance of selected biomarkers .	137
A.2	Amyloid distribution of the validation set	138
A.3	Visualization of numerator (blue) and denominator (red) regions for our biomarker CVR-mSUV-B	140
B.1	Predicted vs observed values, for subset 1	148
B.2	Predicted vs observed values, for subset 2	149
C.1	Relative abundance of controls, by dataset.	150
C.2	Distribution of peptide relative weights of 200 bootstrapped samples, optimized for R² vs age-at-onset	151
C.3	Distribution of peptide relative weights of 200 bootstrapped samples, optimized for precision-recall of controls	151
C.4	Distribution of peptide relative weights of 200 bootstrapped samples, optimized for R² vs age-at-onset and precision-recall area-under- the-curve of controls	152
C.5	Evaluation metrics versus peptide weights	153
D.1	Strip plot of theta values per site	154

List of Tables

3.1	Distribution of individuals and their respective clinical status in the different visits.	53
3.2	Available data for the pseudo-test-retest analysis	57
3.3	Comparison of WSCV, ICC, and PTRT for different reference regions in 18F-AV1451 tau PET studies.	59
4.1	Proportion of individuals per interval time between visits.	70
4.2	Results of Experiment 1	76
4.3	Results of Experiment 2	79
4.4	Results of Experiment 3	80
5.1	Data characteristics	89
5.2	Sample size estimate results	93
5.3	Percentage error results	94
5.4	Group separation results	95
5.5	Regions of the biomarkers discovered for the cognitively unimpaired group	96
5.6	Regions of the biomarkers discovered for the cognitively impaired group	96
5.7	Detectable effect size per example clinical trial	104
6.1	Data count per site.	106
6.2	Correlation of peptides and age-at-onset.	114
6.3	Quantification of biomarker performance, with 95% confidence intervals.	115
6.4	Peptide relative weights	115
6.5	Cross-validated performance for wCVR and other models	118
7.1	Classification metrics per dataset, independently and merged. . . .	125

A.1 BioDisCVR fixing either the numerator or denominator as known literature ROIs.	139
B.1 Solanezumab treatment effect on the change of atrophy in the A4 trial cohort.	147
C.1 Permutation feature importance.	152

Acronyms

A β Amyloid-beta.

A4 Study Anti-Amyloid Treatment in Asymptomatic Alzheimer's Disease study.

AAIC Alzheimer's Association International Conference.

AAO Age At Onset.

AD Alzheimer's disease.

ADNI Alzheimer's Disease Neuroimaging Initiative.

ADRC Alzheimer's Disease Research Center.

AI Artificial Intelligence.

AIBL Australian Imaging Biomarkers and Lifestyle Study of Ageing.

ANOVA Analysis of Variance.

APOE4 Apolipoprotein E epsilon 4 allele.

APP Amyloid Precursor Protein.

ARIA Amyloid-Related Imaging Abnormalities.

AUC Area Under the receiver operating Curve.

BCE Before Common Era.

BioDisCVR Biomarker Discovery using Composite Value Ratio framework.

CDR Clinical Dementia Rating.

CE Common Era.

CI Cognitively impaired.

CPU Central Processing Unit.

CSF Cerebrospinal Fluid.

CSV Comma-separated values.

CU Cognitively unimpaired.

CVR Composite Value Ratio.

DDS Data-Driven Stage.

DX Diagnosis / Cognitive Status.

EOD Early-Onset Alzheimer's Disease.

EOAD Early Onset Familial Alzheimer's Disease.

FN False Negative.

FP False Positive.

GFAP Glial Fibrillary Acidic Protein.

GPU Graphics Processing Unit.

ICC Intraclass Correlation Coefficient.

ICV Intracranial Volume.

iPSC induced Pluripotent Stem Cell.

LOAD Late-Onset Alzheimer's Disease.

LONI Laboratory of Neuro Imaging.

MCI Mild Cognitive Impairment.

MCSA Mayo Clinic Study of Aging.

ML Machine Learning.

- MMSE** Mini-Mental State Examination.
- MRI** Magnetic resonance imaging.
- NfL** Neurofilament light chain.
- PET** Positron emission tomography.
- PR-AUC** Precision-Recall Area Under the Curve.
- PSEN1** Presenilin 1.
- PSEN2** Presenilin 2.
- PTRT** Percent Test-Retest.
- PVC** Partial Volume Correction.
- R²** Coefficient of Determination.
- ROI** Region of Interest.
- SSE** Sample Size Estimate.
- SUIT** Spatially Unbiased Infra-tentorial Template.
- SUV** Standardised Uptake Value.
- SUVR** Standardised Uptake Value Ratio.
- TN** True Negative.
- TP** True Positive.
- UCL** University College London.
- wCVR** Weighted composite value ratio.
- WSCV** Within-Subject Coefficient of Variation.

Chapter 1

PhD Summary

1.1 Introduction

With a growing global population and improvements in addressing other health challenges, Alzheimer's Disease (AD) continues to affect an increasing number of individuals each year [1, 2], with enormous physical, emotional, social and economic impact. Without a known cure, understanding biological changes due to AD can be key to advancing in various aspects such as a better diagnosis, a better disease progression model, and eventually, more and better disease-modifying treatments and a cure. To change the trajectory of the disease, we must first understand its path. And for this, we need in vivo biomarkers [3, 4, 5].

Two main characteristics of AD are the accumulation of proteins amyloid-beta peptides (plaques) and hyperphosphorylated tau (tangles). These pathological proteins start to aggregate long before the onset of noticeable cognitive symptoms, offering a potential window for early detection and intervention. Recently, new guidelines for the diagnosis of AD were published, where imaging-based biomarkers (targeting amyloid plaques and tau tangles) are used to detect and define the disease, before cognitive symptoms manifest [4], highlighting the importance of such biomarkers. Later in the neurodegenerative cascade of events, brain atrophy, another key feature of AD, can be effectively quantified using structural magnetic resonance imaging (MRI) techniques (e.g., the hippocampus shrinks, and the ventricles expand). This non-invasive imaging modality provides valuable insights into the progressive loss of brain volume, aiding in diagnosis confirmation and disease monitoring.

1.2 Statement of problem

The development of robust and precise biomarkers for Alzheimer's disease is crucial for early diagnosis, effective disease monitoring, and the evaluation of novel therapies. However, identifying biomarkers that accurately track neurodegenerative progression in AD can be challenging due to various factors, such as population heterogeneity, measurement noise, and biological changes unrelated to the disease. These challenges are compounded by limitations in current methodologies, where, for example, biologically relevant brain regions may exhibit low signal-to-noise ratios, making them impractical for reliable disease monitoring and delaying progress in the fight against AD. In contrast, less commonly studied but methodologically stable regions (e.g., easier to segment, or with a lower within-subject test-retest coefficient of variance) may offer untapped potential for improved disease progression modelling. These obstacles listed above can impede timely diagnosis, treatment monitoring, and therapeutic innovation, thereby complicating patient management and stalling clinical advancements.

This thesis aims to overcome these barriers by first, identifying limitations of current methods, and second, by proposing new methods and biomarkers to improve disease progression modelling. The focus is specifically on the following hallmark pathologies of AD: neurofibrillary tau tangles (using positron emission tomography), atrophy (using magnetic resonance imaging), and amyloid-beta peptide signatures (relative proportion of peptides, as measured from lab-grown cell cultures).

1.3 Thesis contributions

The general aim of this thesis is to leverage available data to obtain more precise biomarkers of AD. Two fields are explored: imaging (tau PET and structural MRI), and amyloid-beta peptide counts of lab-grown cell cultures.

These are the contributions:

1. **Chapter 3 - part 1** I identified and quantified flaws in the current method for PET signal transformation, the standardised uptake value ratio (SUVR): two reference regions (used in multiple studies) are surprisingly not correlated. Various approaches are proposed to improve it: the usage of multiple reference regions, and a call for a data-driven search of composite reference

regions (executed in Chapter 4). This was presented at the New Horizons in Alzheimer's Disease conference in Leuven, Belgium, in October 2021.

2. **Chapter 3 - part 2** Common reference regions used in tau PET were assessed, highlighting the superiority of one of them (average of whole cerebellum, brainstem, and eroded subcortical white matter). The worst-performing one, the inferior cerebellum grey matter, is still widely used. Additionally, the metrics highlighted the poor performance of partial-volume-corrected PET data. This work was presented at the Alzheimer's Association International Conference in 2024.
3. **Chapter 4** I hypothesised that a better biomarker could come from a data-driven evaluation of brain regions, and then developed a method to explore the space of ratios of composite brain regions and found a solution that outperformed the current literature methods. I called the biomarker a composite value ratio (CVR), and the framework for biomarker discovery BioDisCVR. This work was presented at the Alzheimer's Association International Conference in 2022. Upon obtaining new data for validation, this was later improved and accepted for publication in the journal Brain Communications.
4. **Chapter 5** I applied the BioDisCVR framework with the same conceptual biomarker (CVR) to regional volumes from MRI T1 scans, improving the longitudinal precision, cognitive group separation (disease signal) and statistical power compared to volumetric regions of interest commonly used. This work was presented at the Alzheimer's Association International Conference in 2023, and an extension of this work is currently under revision at the journal Brain Communications (27 February 2025).
5. **Chapter 6** BioDisCVR was applied to the quantification of amyloid-beta peptides from iPSC-derived neurons. Data were from controls (cognitively unimpaired individuals without any family history of early-onset Alzheimer's disease) and individuals with PSEN1 mutations associated with early-onset Alzheimer's disease. Allowing the algorithm to optimise the weights of the different available peptides allowed a further analysis of peptide importance and roles. Results improved on the current state-of-the-art in classification (controls ver-

sus mutation-carriers) and regression with age-at-onset. This work is available at bioRxiv.

6. **Chapter 7** The initial analysis of reference regions in PET scans led to a mathematical model that allowed the multidimensional equivalent of the common, simple, ratio-based biomarker. I further expanded this and applied it to amyloid-beta peptides, obtaining a perfect classification of pathogenic mutations versus controls. This work was presented at the Alzheimer's Association International Conference in 2024. A manuscript showcasing this novel biomarker is available at medRxiv.

Chapter 2

Background

This chapter aims to provide a comprehensive overview of Alzheimer's disease (AD) and the indispensable role of biomarkers in understanding, diagnosing, and ultimately treating this complex neurodegenerative disorder. Beginning with a broad perspective on the global health crisis posed by AD and its historical understanding, the chapter progressively narrows its focus to the biological hallmarks of the disease. It then delves into the multifaceted utility of biomarkers, with particular emphasis on their crucial role in powering the design and interpretation of clinical trials. Detailed discussions on major biomarker modalities, including neuroimaging (PET, MRI) and fluid-based assays (cerebrospinal fluid, blood, and neuronal cell cultures), are presented, establishing the context for the novel contributions of this thesis.

2.1 The Alzheimer's Disease "Pandemic": A Growing Health Crisis

Global impact. Alzheimer's disease is a devastating neurodegenerative disease well-known for its hallmark symptoms such as memory loss, eventually dementia, and ultimately leading to profound disability and death due to body functions shutting down. However, the impact of AD extends far beyond the individuals directly afflicted. The disease casts a long shadow on countless families, friends and caregivers who endure the agonising experience of witnessing a loved one slowly lose their cognitive abilities and sense of self. This profound human toll is compounded by the immense economic burden AD places on families and healthcare systems globally. Since dementia is the most visible part of the disease, it is easier to find global metrics on dementia. It is worth noting that although there are multiple causes

of dementia (e.g., vascular, frontotemporal and Lewy body dementia), AD is the most common one, accounting for 60-70% of cases [6, 2]. A 2024 report by the Alzheimer's Society [7] estimated that 982,000 people were living with dementia in the UK, forecasting 1.4 million in 2040. Similar alarming figures were estimated by the World Health Organization, where in 2023 they estimated that more than 55 million people had dementia worldwide [6]. These escalating figures threaten to overwhelm healthcare and social infrastructures globally, underscoring the urgent need for effective solutions.

Historical perspective. For much of history, only the terminal stage of dementia was recognised, and even then poorly understood. It was often labelled senile dementia, as if advanced age itself inevitably entailed profound cognitive decline. References to cognitive impairment appear sporadically in early medical and cultural records: the Yellow Emperor's Inner Classic (Huangdi Neijing, estimated 475–221 BCE) describes a patient becoming forgetful and "losing their mind" [8]; Greek and Roman sources (8th century BCE–3rd century CE) contain rare mentions of mild cognitive impairment [9]; and an Ancient Egyptian papyrus from around 1500 BCE describes a "heavy heart" that "remembers not yesterday" [10]. Modern research has since demonstrated that dementia is not an inevitable consequence of ageing—although age remains the principal risk factor [11]—but rather the result of underlying pathologies, the most common being Alzheimer's disease [6]. The term Alzheimer's disease was introduced by psychiatrist Emil Kraepelin, based on Alois Alzheimer's 1907 report [12] describing Auguste Deter, a 51-year-old woman with progressive cognitive decline.

AD is broadly classified into two main groups: Early-Onset Alzheimer's Disease (EOAD, representing fewer than 10% of cases [13]), and sporadic Late-Onset Alzheimer's Disease (LOAD). The general rule of thumb is that LOAD occurs after 65 years [13]. A small subset of EOAD cases are known to be caused by mutations in the *APP*, *PSEN1*, and *PSEN2* genes, with the outcome being early onset familial Alzheimer's disease (EOFAD); this is reported to occur in fewer than 1.5% of total cases of AD [13]. There is an ongoing study in Colombia that traced back EOAD families (with a specific mutation related to AD) to the time of the Spanish Conquistadores who began colonising Colombia during the early 16th century [14].

2.2 Introduction to biomarkers for Alzheimer's disease

Below the tip of the iceberg: biological measurable manifestations before dementia onset. Fast forward to today (early 2025), with the advance of biomarkers (informative measures or metrics on the current state of an individual), we can now recognise AD years before cognitive decline is first perceived. This is thanks to the understanding of key elements in AD: brain atrophy (loss of neuronal tissue), and the pathological accumulation of amyloid-beta and tau proteins in the brain. Initially, they were detected by performing a histopathological examination of the brain, post-mortem. Nowadays, we have different ways to detect these, in vivo [3]. As a matter of fact, in 2024 the Alzheimer's Association Workgroup proposed new guidelines for an early diagnosis of AD [4], thanks to using biomarkers, years before cognitive symptoms arise.

The Crucial Role of Biomarkers in Alzheimer's Disease Research and Clinical Practice. Biomarkers have become indispensable tools in the fight against Alzheimer's disease, serving as objective and quantifiable measures that reflect underlying biological processes related to disease pathology [15, 4]. In the context of AD, biomarkers play a multifaceted role, spanning from fundamental research to clinical applications [15, 16, 4]. They are essential for:

- **Early and Accurate Diagnosis:** Distinguishing AD from other forms of dementia and identifying individuals in the early, pre-symptomatic or prodromal stages, where therapeutic interventions may be most effective.
- **Monitoring Disease Progression:** Tracking the trajectory of AD pathology over time, providing crucial insights into disease dynamics and heterogeneity.
- **Predicting Clinical Outcomes:** Identifying individuals at higher risk of developing cognitive symptoms due to AD or progressing more rapidly, enabling personalised risk assessment and management strategies.
- **Evaluating Therapeutic Efficacy:** Biomarkers can serve as surrogate endpoints in clinical trials, enabling objective assessment of how novel therapies influence underlying disease processes. On a personal note – paraphrasing Mercè Boada's reflections at EuroPAD 2025 – I believe this is valuable, but not

sufficient; ultimately, we must demonstrate meaningful, tangible improvements in patients' quality of life.

- **Advancing Pathophysiological Understanding:** By revealing molecular and cellular changes, biomarkers provide invaluable insights into the complex mechanisms driving AD, guiding the development of targeted therapeutic strategies.
- **Clinical Trial Design:** In clinical trials, biomarkers are essential for screening and enrichment (selecting cohorts based on biomarker positivity), stratification (e.g., by *APOE4* status), and safety monitoring (e.g., detecting amyloid-related imaging abnormalities, ARIA). Regulatory qualification of biomarkers can depend on a precisely defined context of use.

Quoting Jack Jr. et al. [4], "*AD is defined by its unique neuropathologic findings; therefore, detection of AD neuropathologic change by biomarkers is equivalent to diagnosing the disease*". The pursuit of robust and reliable biomarkers is thus central to accelerating progress across the entire spectrum of AD research, from basic discovery science to the development and implementation of effective clinical interventions.

Heterogeneity and Subtyping. AD is not a homogeneous disease among individuals; it exhibits considerable biological and clinical heterogeneity [17, 18, 19, 20, 21, 22]. Biomarkers reveal distinct subtypes (e.g., limbic-predominant, hippocampal-sparing) with varying progression rates and clinical presentations. Data-driven computational models, such as event-based models [23] (EBMs) and Subtype and Stage Inference [17] (SuStaln), are increasingly used to parse this subtype and stage structure, integrating multimodal biomarker data to better characterise individual disease trajectories. This heterogeneity underscores the need for precise biomarkers that can accurately capture these diverse profiles.

Biomarkers for Alzheimer's disease Alzheimer's disease has multiple identified characteristics, as we have stated: accumulation of amyloid-beta (plaques), tau (tangles), atrophy (brain tissue loss)... and this is gradual. Imaging and fluid biomarkers enable measurement of these as they accumulate over time. At the far end of the biomarkers, cognitive tests allow for the assessment of different cognitive

abilities in individuals, which become affected after years of the neurodegenerative process. On the early side of the process, before amyloid plaques are detectable, we can observe different amyloid-beta peptides (chains of amino acids) in various fluids (CSF, blood, and cell-based cultures in vitro). All of these are windows that allow us to examine and assess the state of an individual, and have a better understanding of the mechanisms and chain of events going on as the disease advances.

2.3 Neuroimaging biomarkers: Visualising Brain Changes

There are various technologies that enable us to peek inside the brain without needing to open it up. A characteristic property of neuroimaging is that it allows us to have spatial information (e.g., assessing the volume of the hippocampus, or the severity of tau accumulation in a specific region of interest). This has undoubtedly revolutionised how we understand neuropathological disorders [24, 25, 4].

Note on Scope. While the current section outlines key neuroimaging techniques relevant to Alzheimer's disease, this thesis's primary focus lies in the data science methodologies applied to the resulting quantitative outputs. Accordingly, detailed discussions of image acquisition protocols, scanner hardware specifications, and preprocessing pipelines fall outside the scope of this work. Instead, the core chapters focus on the interpretation, modelling, and integration of neuroimaging-derived metrics within a broader analytical framework designed for downstream applications.

2.3.1 Computed Tomography (CT) scan.

A CT scan allows for the evaluation of vascular brain injuries [4]. Not as popular for AD as the imaging modalities below, and not directly used in this thesis.

2.3.2 Magnetic Resonance Imaging (MRI).

Magnetic Resonance Imaging has become a fundamental tool in the search for reliable neuroimaging biomarkers in Alzheimer's disease research and clinical practice. The ability of MRI to non-invasively probe the brain's structure, function, and microenvironment with exceptional spatial and tissue contrast makes it useful for diagnosis, prognosis, and therapeutic monitoring of AD. This subsection offers an overview of the principles behind MRI, with a focused discussion on image formation and tissue contrast, followed by a detailed examination of the MRI modalities most relevant to Alzheimer's disease, including their applications and limitations.

2.3.2.1 Fundamentals of MRI Signal Generation

Principle of Nuclear Magnetic Resonance. At the core of MRI is the phenomenon of nuclear magnetic resonance (NMR). NMR leverages the property of certain atomic nuclei to behave like tiny bar magnets due to their intrinsic quantum mechanical spin. The hydrogen nucleus is of particular interest in biological tissues because of its abundance in water and fat, two major constituents of the human brain. When an external static magnetic field (B_0) is applied, hydrogen nuclei align either parallel or anti-parallel to the field, with a slight excess in the lower-energy parallel state. This creates a measurable but small net magnetisation in the direction of B_0 , known as the equilibrium magnetisation. In the absence of a magnetic field, nuclear spins are randomly oriented, resulting in zero net magnetisation. Under B_0 , the net magnetisation vector also precesses around the field axis at a frequency determined by the gyromagnetic ratio of the nucleus and the field strength, known as the Larmor frequency. For hydrogen, this falls within the radiofrequency (RF) range.

Role of Static Magnetic Fields: Image Quality and Homogeneity. The strength of the applied magnetic field (measured in Tesla, T) is a central determinant of MRI image quality. Higher field strengths (e.g., 7T vs. 1.5T) produce a greater net magnetisation, thereby increasing the signal-to-noise ratio (SNR) and enabling higher spatial resolution. However, susceptibility to artefacts, chemical shift differences, and safety considerations such as tissue heating also increase with field strength [26]. Magnetic field homogeneity is equally critical for accurate signal localisation and artefact-free images. Inhomogeneities degrade SNR, introduce geometric distortions (especially near air–tissue interfaces such as the frontal sinuses and ear canals) [27], and impair reliable tissue contrast. In practice, careful magnet design and active shimming routines are employed to optimise homogeneity.

Radiofrequency Excitation and Signal Induction. Once the system reaches equilibrium, an RF pulse at the Larmor frequency is applied orthogonally to B_0 . This RF pulse tips the net magnetisation away from the longitudinal (z) axis into the transverse (xy) plane by a specific flip angle (typically 90°). At this point, the excited protons precess in phase, maximising their transverse magnetisation. When the RF pulse is switched off, the protons relax back to equilibrium, releasing energy. During this process, the transverse component of magnetisation, oscillating at the Larmor

frequency, induces a small voltage in a surrounding receiver coil: the measurable NMR signal. Two key timing parameters govern MRI signal acquisition [28]. The *repetition time* (TR) is the interval between consecutive RF excitations of the same slice, which determines the degree of longitudinal relaxation allowed before the next excitation, strongly influencing image contrast. The *echo time* (TE) is the interval between the RF pulse and the peak of the received signal, reflecting the amount of transverse relaxation that has occurred. Together, TR and TE modulate the relaxation dynamics encoded in the signal, shaping both tissue contrast and diagnostic sensitivity. A third parameter, the *inversion time* (TI), is used in inversion recovery sequences (e.g., FLAIR) to selectively null specific tissues such as CSF.

Relaxation Mechanisms: T1 and T2. Relaxation describes how protons return to equilibrium after excitation, governing both the duration and intensity of the MR signal:

- **T1 Relaxation (Spin-Lattice/Longitudinal):** The process by which longitudinal magnetisation realigns with B_0 as protons transfer energy to their molecular environment. T1 times are tissue-dependent and form the basis of T1-weighted imaging.
- **T2 Relaxation (Spin-Spin/Transverse):** The progressive loss of phase coherence among precessing protons due to spin–spin interactions, causing decay of transverse magnetisation. T2 times are tissue-specific and underpin T2-weighted imaging.

A related parameter, **T2*** (T2-star), incorporates both intrinsic T2 decay and additional dephasing from magnetic field inhomogeneities. T2* is particularly relevant for detecting microbleeds and mineralisation.

Tissue Contrast Generation and Signal Weighting. A powerful aspect of MRI is its ability to generate tissue contrast by exploiting differences in relaxation times, water content, and microenvironmental factors. Contrast is primarily manipulated through adjustment of TR, TE, and, in some sequences, TI:

- **T1-Weighted Imaging:** Achieved with short TR and short TE. The short TR prevents full longitudinal recovery, thereby accentuating differences in T1

relaxation. Fat appears bright, whereas water-rich tissues (CSF, oedema) are dark. T1-weighting is the standard for anatomical imaging.

- **T2-Weighted Imaging:** Achieved with long TR and long TE. The long TE allows transverse relaxation differences to dominate. Fluids thus appear bright, making T2-weighted sequences sensitive to pathology such as oedema and gliosis.
- **Proton Density (PD) Imaging:** Uses intermediate TR and short TE to minimise T1 and T2 weighting, yielding images proportional to proton concentration.

Contrast can be further refined with specialised sequences, suppression techniques (e.g., Fluid-Attenuated Inversion Recovery (FLAIR) to null CSF, or fat suppression), or gadolinium contrast agents. These principles form the foundation for the diverse range of MRI modalities described below.

2.3.2.2 MRI Modalities and Their Applications in Alzheimer's Disease

MRI encompasses a wide spectrum of pulse sequences, each providing unique and sometimes complementary information about the living brain. In AD research, several modalities serve as core biomarkers for diagnosis, differential diagnosis, monitoring progression, and evaluating treatment effects.

Structural MRI (sMRI): Volumetric Imaging and Cortical Thickness. Structural MRI, typically acquired with high-resolution T1-weighted sequences, has been instrumental in neurodegenerative research. It enables sensitive quantification of regional and global brain atrophy, cortical thinning, and other morphometric changes that characterise AD pathology [29]. Key sMRI quantitative biomarkers include: volumetric measures (e.g., whole brain [30], hippocampus [31], ventricles [32], medial temporal lobe [31, 4]); cortical thickness (e.g., [33]); and atrophy proxies such as the *Boundary Shift Integral* (BSI) [34] and its generalised version [35]. Automated volumetric analysis tools, along with standardisation initiatives such as the Alzheimer's Disease Neuroimaging Initiative (ADNI), have harmonised imaging protocols, reduced scanner-dependent variability, and enabled the creation of large-scale datasets that have advanced our understanding of AD. Visual rating of medial temporal lobe [36] or ventricular enlargement [37] also remains useful in clinical

staging when atrophy is pronounced.

Functional MRI (fMRI): Neuronal Activation and Connectivity. Functional MRI leverages the blood-oxygen-level-dependent (BOLD) signal to map regional changes in neuronal activity during cognitive tasks and at rest (resting-state fMRI). In AD and late mild cognitive impairment (MCI), fMRI demonstrates reduced hippocampal/medial temporal activation [38] and altered Default Mode Network (DMN) connectivity [38, 39]. Resting-state paradigms are particularly valuable in prodromal AD, where patient compliance with tasks is limited. Although promising for diagnosis and monitoring, fMRI is subject to signal variability and is not included in the revised diagnostic criteria for AD from the Alzheimer's Association Workgroup [4].

Diffusion MRI: White Matter Integrity and Connectivity. Diffusion MRI assesses white matter microstructure by quantifying the directional diffusion of water molecules. Diffusion Tensor Imaging (DTI) provides metrics such as fractional anisotropy (FA) and mean diffusivity (MD), which reflect axonal integrity and myelination. In AD, reduced FA and increased MD have been consistently reported in association tracts, indicating early microstructural degeneration [40, 41]. Advanced models, such as Neurite Orientation Dispersion and Density Imaging (NODDI), offer improved specificity by disentangling neurite density and orientation dispersion [42], while free-water imaging has recently emerged as another sensitive marker of neuroinflammation and neurodegeneration. Connectomics integrates diffusion metrics with graph theory to map network-level disruptions, offering promising biomarkers for early diagnosis and staging [43]. Nonetheless, diffusion abnormalities are not disease-specific and are also observed in other neurodegenerative and psychiatric conditions.

Additional MRI Modalities. Beyond sMRI, fMRI, and diffusion MRI, several specialised modalities provide complementary insights: **Susceptibility-Weighted Imaging (SWI)** enhances contrast based on magnetic susceptibility, enabling sensitive detection of cerebral microhaemorrhages, iron deposition, and vascular abnormalities associated with cerebral amyloid angiopathy or mixed AD pathology [44]. **Arterial Spin Labelling (ASL)** quantifies cerebral blood flow non-invasively by using magnetically labelled arterial blood water as an endogenous tracer, revealing hypoperfusion in AD-affected regions [45]. **Magnetic Resonance Spectroscopy (MRS)** measures brain metabolites by suppressing the dominant water signal. Altered

metabolite ratios in AD reflect neuronal loss, membrane turnover, and glial activation [46], though MRS remains less commonly used.

While none of these modalities is specific to AD, they provide valuable physiological and biochemical context that enhances the utility of structural and functional MRI. When combined, multimodal MRI approaches can contribute to comprehensive biomarker strategies alongside PET and fluid-based markers.

2.3.3 Positron-Emission Tomography (PET).

Positron Emission Tomography (PET) is a molecular imaging modality that enables the in vivo visualisation and quantification of biological processes at the cellular and molecular level. In Alzheimer's disease research, PET has become indispensable for detecting hallmark pathological processes such as amyloid- β deposition and tau neurofibrillary tangles, as well as for measuring cerebral metabolism and neuroinflammation [47, 48, 49, 5]. PET is central both to basic pathophysiological research and to clinical translation, where it informs diagnosis, prognosis, and patient stratification in therapeutic trials.

This subsection reviews the fundamental principles of PET imaging, the development and application of radiotracers relevant to AD, quantification approaches with emphasis on the Standardised Uptake Value Ratio (SUVR), and methodological considerations such as partial volume effects, reference region selection, and harmonisation efforts. The section also highlights limitations of current methods and emerging directions in PET biomarker research.

2.3.3.1 Principles of PET Imaging

Physics and Scanner Instrumentation. PET imaging is based on the detection of coincident gamma photons following positron emission. Radiotracers labelled with positron-emitting isotopes (e.g., ^{18}F , ^{11}C) are injected intravenously and taken up by specific molecular targets. Upon decay, the emitted positron annihilates with an electron, producing two 511 keV photons emitted nearly 180° apart. PET scanners, typically configured as detector rings, detect these coincident photons and reconstruct three-dimensional maps of tracer distribution [50]. Hybrid PET/CT and PET/MRI systems provide anatomical co-registration, attenuation correction, and multiparametric imaging capabilities, enhancing both sensitivity and interpretability [51].

Radiotracers and Labelling. PET relies on radiotracers labelled with short-lived positron emitters. Fluorine-18 ($t_{1/2} = 110$ min) is the most widely used isotope due to its relatively long half-life, allowing centralised production and distribution. Carbon-11 ($t_{1/2} = 20$ min) offers chemical versatility but requires onsite cyclotrons. Successful tracer design demands high affinity and selectivity for the molecular target, favourable blood–brain barrier penetration, and low nonspecific binding.

2.3.3.2 PET Radiotracers in Alzheimer’s Disease

Glucose Metabolism: FDG-PET. ^{18}F -fluorodeoxyglucose (FDG) PET measures cerebral glucose metabolism, serving as a proxy for synaptic activity and neuronal integrity. AD is characterised by a stereotyped pattern of hypometabolism in temporoparietal association cortices and posterior cingulate/precuneus, with relative sparing of primary sensory, motor and occipital cortices [52, 53]. FDG-PET has diagnostic and prognostic utility, often identifying dysfunction before atrophy is detectable by MRI [52, 53].

Amyloid PET Tracers. Amyloid PET was the first molecular imaging biomarker to achieve widespread clinical use in AD. The prototypical tracer [^{11}C]PiB demonstrated selective binding to fibrillar amyloid- β [54]. Subsequently developed ^{18}F -labelled tracers, including florbetapir, florbetaben, and flutemetamol, enabled broader clinical adoption due to their longer half-lives [55, 56]. Amyloid PET detects pathology years before symptom onset, is incorporated into diagnostic criteria, and is now used for therapy eligibility in anti-amyloid trials [47, 57]. However, amyloid positivity is not specific to AD, as deposits can be found in cognitively normal older adults and in other dementias [58].

Tau PET Tracers. Tau PET provides a closer correlate of clinical severity, as the regional distribution of tau tangles mirrors Braak staging and tracks cognitive decline [59, 60]. First-generation tau tracers, such as [^{18}F]flortaucipir (AV-1451), demonstrated feasibility but were limited by off-target binding to monoamine oxidase, basal ganglia, and choroid plexus [61, 62]. Second-generation tracers, including [^{18}F]MK-6240, [^{18}F]RO948, and [^{18}F]PI-2620, offer higher affinity, improved signal-to-noise, and reduced off-target binding [63, 64]. These advances have positioned tau PET as a leading biomarker for staging and monitoring therapeutic response in clinical trials.

Emerging Radiotracers. Beyond amyloid and tau, new PET tracers target complementary aspects of AD biology. Neuroinflammation can be assessed using TSPO ligands, though specificity and sensitivity remain under debate [65]. Synaptic density tracers such as [^{11}C]UCB-J and [^{18}F]SynVesT-1 bind to synaptic vesicle glycoprotein 2A (SV2A), providing sensitive markers of synaptic loss [66, 67]. Other targets under development include mitochondrial function [68], cholinergic neurotransmission [69], and myelin integrity [70], broadening the scope of PET as a multimodal biomarker platform.

2.3.3.3 Quantification and Interpretation

Quantitative Approaches. PET allows both absolute and relative quantification of tracer binding. Full kinetic modelling with arterial blood sampling and compartmental analysis (1- or 2-tissue models) provides gold-standard parameters such as binding potential (BP_{ND}) or distribution volume (V_T) [71]. However, these methods are resource-intensive and impractical for large-scale or clinical use.

SUV and SUVR. In practice, static imaging protocols and semi-quantitative measures are widely used. The Standardised Uptake Value (SUV) normalises tracer uptake to injected dose and body weight, while the Standardised Uptake Value Ratio (SUVR) compares target regions to a reference region presumed to be free of specific binding [72]. Cerebellar cortex, subcortical white matter, and pons are commonly used references [73, 74]. SUVR is robust for group-level analyses and is used extensively in multicentre studies and clinical trials [75, 57], but it remains a relative measure sensitive to blood flow, reference region choice, and partial volume effects [76].

Partial Volume Effects (PVE). The limited spatial resolution of PET leads to spillover between regions, exacerbated by cortical thinning in AD. Partial volume correction (PVC) methods, such as region-based voxel-wise correction or geometric transfer matrix approaches, are meant to improve SNR, but add complexity and variability [77, 78, 76]

2.3.3.4 Clinical and Research Applications

PET has transformed the AD research landscape. Amyloid and tau PET are now embedded in diagnostic frameworks [47, 4, 5], provide enrichment strategies for clinical trials, and serve as pharmacodynamic readouts in anti-amyloid and anti-tau

therapies [79, 75]. FDG-PET remains valuable for differential diagnosis, distinguishing AD from frontotemporal dementia or dementia with Lewy bodies [53]. Ongoing harmonisation efforts, such as Centiloid scaling for amyloid PET [80] and analogous initiatives for tau PET, such as CenTauRz [81], aim to improve cross-trial comparability and accelerate regulatory and clinical acceptance. Nevertheless, PET is resource-intensive, involves radiation exposure, and is not universally accessible, highlighting the need for complementary, scalable biomarkers such as blood-based assays.

2.4 Fluid biomarkers

Fluid biomarkers provide direct molecular evidence of Alzheimer's disease pathology, complementing the spatial resolution of imaging. They can be sampled from **cerebrospinal fluid (CSF)**, **blood**, or **in vitro models** such as patient-derived cell cultures, each contributing distinct advantages for research and clinical applications.

2.4.1 Cerebrospinal Fluid (CSF)

For over two decades, CSF has been the reference standard for in vivo quantification of molecular hallmarks of Alzheimer's disease [82, 83, 84]. The most established biomarkers include:

- **Amyloid- β peptides:** Multiple isoforms ($A\beta_{37}$, $A\beta_{38}$, $A\beta_{40}$, $A\beta_{42}$, $A\beta_{43}$) can be quantified. $A\beta_{42}$ is selectively reduced in AD due to deposition in plaques, and the $A\beta_{42}/A\beta_{40}$ ratio corrects for inter-individual variability in peptide production [82].
- **Phosphorylated tau (p-tau):** Hyperphosphorylated tau is a marker of neurofibrillary pathology. Sites such as threonine-181, -217 and -231 (p-tau181, p-tau217, p-tau231) show strong specificity for AD and close correspondence with tau PET imaging [85, 86].
- **Total tau (t-tau):** Reflects overall axonal damage, not specific to AD, but informative for disease intensity [87].

Despite their sensitivity, the invasiveness of lumbar puncture limits CSF assays in routine practice and large trials.

2.4.2 Blood-Based Biomarkers (BBMs)

Technological advances in ultra-sensitive assays (e.g., Simoa, Elecsys, mass spectrometry) have made peripheral detection of AD biomarkers reliable [88]. Key plasma biomarkers include:

- **A β 42/A β 40 ratio:** A low ratio indicates cerebral amyloidosis and correlates with amyloid PET positivity .
- **Plasma p-tau:** Especially p-tau217 and p-tau181, which strongly predict amyloid and tau PET status, discriminate AD from other dementias, and change in response to anti-amyloid therapy.
- **Neurofilament light chain (NfL):** A non-specific marker of neuronal injury, useful for tracking progression but elevated in many neurodegenerative conditions.
- **Glial fibrillary acidic protein (GFAP):** Reflects astrocytic activation; elevated early in the AD continuum and linked to amyloid burden.

The translational impact of BBMs is underscored by a landmark regulatory milestone: in 2025 the U.S. Food and Drug Administration (FDA) approved the first blood-based AD biomarker assay for clinical use: the ratio between pTau17 and beta-amyloid 1-42 [89]. This positions plasma biomarkers as not only research tools but emerging components of diagnostic workflows and trial endpoints.

2.4.3 In Vitro Cell Models

Induced pluripotent stem cell (iPSC)-derived neurons and other cell-based systems provide a complementary in vitro source of fluid biomarkers. Particularly in familial AD, where APP, PSEN1, or PSEN2 mutations alter amyloid processing, iPSC-derived neurons faithfully recapitulate individual-specific A β peptide profiles [90]. These models enable the dissection of peptide heterogeneity and phosphorylation patterns under controlled conditions, and they serve as translational platforms for therapeutic screening and biomarker discovery.

2.4.4 Implications

Together, CSF, blood, and in vitro fluid biomarkers form a multi-layered molecular window into AD pathology. Their integration into clinical trials can aid diagnos-

tic precision for cheap and fast stratification, enable early detection, and provide pharmacodynamic readouts. However, harmonisation of assays, context-specific cut-points, and careful consideration of specificity remain essential for their widespread adoption in both research and clinical settings.

2.5 Cognitive measures

Eventually, after pathological change has affected particular brain regions, cognition becomes measurably altered, affecting the individual's interaction with the world, and subsequently, their quality of life. Cognitive assessment is therefore central to the (late, symptomatic) diagnosis, staging, monitoring, and trial-based evaluation of Alzheimer's disease. Cognitive tests (1) quantify impairment across memory, language, executive function and other domains, (2) support differential diagnosis, (3) provide outcomes for clinical trials, and (4) serve as anchors for biomarker correlations. In addition to traditional pen-and-paper instruments, the last decade has seen a rapid development of *digital cognitive biomarkers* (active and passive), which aim to increase sensitivity, scalability and temporal resolution of cognitive measurement. This section summarises established instruments, composite endpoints used in AD trials, and recent advances (and limits) in digital cognitive assessment.

Note on Cognitive Assessment Scope. This thesis is based on secondary analysis of previously collected data. All cognitive assessments were conducted prior to the present work, using standardised protocols defined by the original study design. As such, no new cognitive testing was performed by the author. Nevertheless, given the central role of cognitive profiling in Alzheimer's disease research and diagnosis, we provide a brief overview of commonly used assessment tools and their relevance to biomarker interpretation and disease staging.

2.5.1 Conventional cognitive instruments

A number of well-established clinical instruments remain widely used in clinical practice and trials:

- **Mini-Mental State Examination (MMSE).** Brief 30-point (more is better) screening test covering orientation, recall, attention, language and construction; widely used as a global index of cognition. It is quick to administer but has ceiling effects in early disease and is influenced by education and language

[91].

- **Clinical Dementia Rating (CDR) and CDR–Sum of Boxes (CDR-SB).** Clinician/informant interview yielding global staging (0, 0.5, 1, 2 . . .) and a sum-of-boxes score used as an outcome in many trials and epidemiologic studies. The CDR integrates functional information and is useful for staging across AD severity [92, 93].
- **ADAS-Cog (and variants).** The Alzheimer’s Disease Assessment Scale cognitive subscale (ADAS-Cog) is a multi-item instrument designed for clinical trials of symptomatic drugs; several modified/extended versions (including 13-item variants) have been proposed to improve sensitivity in early stages [94, 95, 96].
- **Domain-specific neuropsychological tests.** Episodic memory tests (e.g., Logical Memory, Rey Auditory Verbal Learning Test (RAVLT), Free and Cued Selective Reminding Test (FCSRT)), processing-speed/executive tasks (Digit Symbol/Substitution, Trail Making), and language/fluency measures are used to characterise the cognitive profile and detect subtle changes.

While these instruments are well validated, their conventional administration (clinic visit, single time point) limits temporal resolution and may miss early, subtle longitudinal changes or produce practice effects when used repeatedly.

2.5.2 Composite endpoints: PACC and related composites

To increase sensitivity to early decline in preclinical and prodromal stages, trials have adopted composite endpoints that combine multiple domain measures. The *Preclinical Alzheimer’s Cognitive Composite* (PACC) was developed to detect amyloid-related cognitive decline in clinically normal older adults by combining multiple sensitive neuropsychological measures into a single endpoint [97]. The original PACC combined episodic memory, global cognition and processing speed / executive measures; later variants (e.g., PACC5) extended the composite with additional tests (semantic fluency, other memory indices) to increase sensitivity and coverage of early deficits [98]. Composites are attractive for prevention trials because they (i) aggregate signal across domains, (ii) reduce multiplicity when properly pre-specified, and (iii) can be calibrated for longitudinal change in enriched samples (e.g., amyloid

positive). Nevertheless, composite selection must balance sensitivity, interpretability, ceiling/floor effects, and clinical meaningfulness for regulators and clinicians.

2.5.3 Digital cognitive biomarkers: active and passive approaches

Definitions and taxonomy. *Digital cognitive biomarkers* are objective measures of cognition derived from digital devices. Broadly:

- *Active digital assessments* require the participant to perform structured cognitive tasks on a device (smartphone, tablet, web) – e.g., episodic memory tasks, reaction-time tests, gamified tasks. Examples include Cogstate [99], the Mezurio app [100], Altoida DNS [101], and numerous research batteries.
- *Passive / behavioural digital biomarkers* are derived from continuous or background data streams (speech during phone calls, keystroke and touchscreen dynamics, GPS / mobility traces, sleep and gait metrics from wearables, patterns of computer use). These do not require dedicated testing sessions and can capture real-world behaviour and variability. [102, 103]

2.5.4 Psychometric and practical considerations

Reliability, validity and practice effects. Digital tests must demonstrate test–retest reliability, construct and criterion validity (correlation with established neuropsychological measures and biomarkers), and sensitivity to clinically meaningful change. Frequent repeated measurement increases power but raises the problem of practice effects; clever task design (alternate forms, adaptive difficulty) and statistical modelling of practice trajectories are essential to mitigate bias.

Standardisation and cross-platform harmonisation. Device heterogeneity (screen size, input modality), OS variability and network differences can induce measurement noise. Harmonisation steps include device-agnostic task design, device metadata logging, pre-deployment calibration, and analytic pipelines that correct for platform effects.

Validation and regulatory expectations. Regulatory bodies now recognise the potential of digital endpoints but expect rigorous evidentiary standards. The FDA's guidance on *Digital Health Technologies for Remote Data Acquisition in Clinical Investigations* outlines expectations for selection, verification/validation, data management, usability, and risk assessment when DHTs are used to collect trial end-

points; early engagement with regulators and pre-specified validation plans are recommended [104].

2.5.5 Limitations, equity and ethical concerns

Digital approaches risk exacerbating health disparities unless plans explicitly address device ownership, digital literacy, language and cultural adaptation, and accessibility (sensory impairments). Privacy, data governance and informed consent are paramount: passive sensing in particular can reveal sensitive behavioural patterns and location/time data that require robust data minimisation, secure storage, and transparent participant information. Many digital biomarkers show good within-sample performance but lack widespread external validation and standardised normative frameworks – a strong limitation for regulatory qualification and clinical translation [105, 106].

In summary, established psychometric scales (MMSE, CDR, ADAS-Cog and domain tests) remain essential. Composite endpoints such as the PACC increase sensitivity in preclinical AD trials. Digital cognitive assessments – both active and passive – offer powerful new avenues for scalable, high-frequency, and ecologically valid measurement, but they require careful validation, harmonisation, and ethical/regulatory planning before replacement of conventional primary endpoints in pivotal trials is justified.

2.6 Clinical Trials in Alzheimer's Disease

2.6.1 Introduction

As we mentioned earlier, Alzheimer's disease remains one of the greatest challenges in clinical drug development, with a global burden that continues to rise in parallel with population ageing. Despite decades of research, only a small number of disease-modifying treatments have reached approval, and their clinical impact remains modest. This has catalysed a transformation of the clinical trial landscape, characterised by diversification of therapeutic targets, the adoption of novel biomarkers, innovative trial designs, and regulatory frameworks that increasingly accommodate biomarker-driven decision making.

This section provides an overview of AD clinical trials in 2025, with emphasis on therapeutic categories, the role of biomarkers (particularly imaging and fluid

markers), regulatory considerations, methodological challenges, and emerging trends. Particular attention is given to biomarker-driven trial designs, as these underpin the methodological contributions of this thesis. The 2025 "Alzheimer's disease drug development pipeline" [107] is a highly recommended read, and contributes to a significant portion of this section.

2.6.2 The 2025 Trial Landscape

The AD clinical trial pipeline in 2025 is larger and more diverse than in previous years. According to the most recent systematic review, there are **182 active clinical trials** evaluating **138 unique therapeutic agents**, compared with 164 trials and 127 drugs in 2024 [107]. These span all stages of development:

- **Phase 1:** 48 trials (45 drugs)
- **Phase 2:** 86 trials (75 drugs)
- **Phase 3:** 48 trials (31 drugs)

Roughly one-third of these agents are **repurposed drugs**, highlighting the strategic use of existing safety profiles to accelerate progress. Importantly, trial populations now extend across the disease continuum, from cognitively unimpaired at-risk individuals to patients with late-stage dementia, reflecting a paradigm shift towards both prevention and treatment.

2.6.3 Therapeutic Targets and Landmark Trials

The historical dominance of amyloid as a therapeutic target has given way to a more balanced portfolio encompassing tau, neuroinflammation, synaptic dysfunction, and neuroprotection, alongside symptomatic and device-based interventions.

2.6.3.1 Amyloid

Monoclonal antibodies remain the most advanced amyloid-directed therapies. Leqembi (lecanemab) received full FDA approval in 2023 and is being evaluated in prevention-focused studies such as *AHEAD 3-45*, enrolling asymptomatic but biomarker-positive individuals [79]. Donanemab (Eli Lilly) has completed phase 3 trials demonstrating modest slowing of cognitive decline [75]. Oral anti-amyloid agents such as ALZ-801 (valiltramiprosate) have targeted high-risk APOE4/4 carriers, though with mixed efficacy [108, 109]. Currently, amyloid is targeted by 11, 12

and 7 agents in Phase 1, 2 and 3 clinical trials, respectively [107].

2.6.3.2 Tau

Tau has emerged as a key therapeutic focus, given its stronger correlation with neurodegeneration and clinical symptoms. The antisense oligonucleotide BIIB080 (Biogen/Ionis) has shown promising dose-dependent reductions in CSF and PET measures of tau and was granted FDA Fast Track status in 2025 [110]. Currently, tau is targeted by 7, 7 and 1 agents in Phase 1, 2 and 3 clinical trials, respectively [107].

2.6.3.3 Neuroinflammation and Metabolic Pathways

Inflammatory cascades are increasingly recognised as critical in AD progression [49]. The glucagon-like peptide-1 (GLP-1) receptor agonist semaglutide, already approved for diabetes and obesity, is under evaluation in two large phase 3 trials (*EVOKE/EVOKE+*), designed with extensive biomarker substudies [111]. Novel NLRP3 inflammasome inhibitors are also progressing through early development [112].

2.6.3.4 Synaptic Function and Neuroprotection

Synaptic loss is the pathological feature most closely associated with cognitive impairment in AD [113]. Several agents are in clinical testing, including SPG302 (Spinogenix) [114], a first-in-class synaptic regenerative drug now in phase 2, and buntanetap (ANVS401) [115], a multi-target oral compound in phase 3. Sigma-1 receptor agonists such as blarcamesine (Anavex 2-73) [116] exemplify efforts to combine neuroprotective and synaptogenic effects.

2.6.4 Imaging Biomarkers in Alzheimer's Disease Clinical Trials

Neuroimaging plays a central role in modern Alzheimer's disease clinical trials, both for patient selection and as quantitative outcome measures. The two principal modalities are volumetric magnetic resonance imaging (MRI), which captures neurodegeneration, and positron emission tomography (PET), which visualises amyloid and tau pathology. Their applications in trials can be broadly grouped into screening/enrichment, longitudinal monitoring, safety evaluation, and primary or secondary endpoints.

Screening and Enrichment. Amyloid PET is now a standard inclusion criterion in most phase II–III trials [75, 79, 57] to ensure pathological confirmation and reduce heterogeneity, while tau PET is increasingly used to stage disease severity [75]. Structural MRI is used as a tool to detect abnormalities, which can lead to exclusion from studies [75].

Monitoring Disease Progression and Endpoints. In phase-3 randomised clinical trials targeting Alzheimer’s disease progression, MRI-based volumetric measures of the whole brain, the ventricles and the hippocampus have been used as primary [117] and secondary outcome measures [118, 57, 75, 119]. Likewise, both amyloid and tau PET SUVR measures have been used as secondary outcomes [75]. Compared with clinical endpoints (e.g., CDR-SB, ADAS-Cog), imaging measures typically show greater effect sizes and lower variability, thereby reducing required trial sample sizes [3].

Safety Monitoring. With the advent of monoclonal antibody therapies, MRI has become essential for detecting amyloid-related imaging abnormalities (ARIA), including vasogenic oedema (ARIA-E) and microhemorrhages (ARIA-H). Thus, standardised protocols are embedded in anti-amyloid trial designs [79, 120].

Outlook. The AD clinical trial ecosystem in 2025 is marked by robust growth, diversification of therapeutic targets, and increasing reliance on biomarkers. In particular, imaging biomarkers are indispensable tools in AD clinical trials: they enrich trial populations, monitor disease progression, ensure patient safety, and provide regulatory-grade endpoints. Ultimately, though, having improved biomarkers is not enough (paraphrasing Mercè Boada’s reflections at EuroPAD 2025), and we need to see practical, tangible gains in the quality of life of the treated individuals; in fact, *they*, and their close relatives and friends need to notice the improvement.

2.7 A Unifying Framework: The Evolving Biological Definition of AD

The proliferation of imaging and fluid biomarkers has catalysed a paradigm shift in Alzheimer’s disease research, moving from symptom-based diagnostic definitions toward biology-based ones. The 2018 NIA-AA research framework was a major step, introducing the A/T/(N) system to classify individuals based on biomarker evidence

of Amyloid (A), Tau (T), and Neurodegeneration (N) [47]. While transformative for standardising research, its categorical (positive/negative) nature limited its ability to capture the vast heterogeneity and predict individual clinical trajectories.

To address this, Jack et al. proposed revised criteria in 2024 [4], designed to provide greater prognostic granularity and better align with the continuous nature of the disease. The key evolutions include:

- **Integration of Blood-Based Biomarkers:** Core diagnostic biomarkers (Core 1) now formally include validated blood-based assays, notably plasma p-tau217. This acknowledges that certain plasma markers have achieved sufficient accuracy for screening and diagnostic purposes, offering a scalable and less invasive alternative to PET or CSF.
- **Clarification of Co-Pathologies:** The framework explicitly separates the core AD definition from common co-pathologies. Biomarkers for Vascular brain injury [V], Inflammatory mechanisms [I], and alpha-synucleinopathy [S] are designated as important prognostic modifiers but are not part of the primary AD definition.
- **Emphasis on Quantitative Staging:** A new focus is placed on staging (Core 2), driven primarily by the amount and spatial extent of tau pathology on PET scans. This is a critical shift, as tau burden is a much stronger correlate of neurodegeneration and cognitive symptoms than amyloid burden, thus providing a more dynamic and clinically relevant measure of disease severity.

These revisions represent an important evolution: they preserve the basic A/T/(N) idea, but shift toward incorporating quantitation, fluid biomarkers, and staging, thereby helping to bridge research and eventual clinical translation.

2.8 Summary and motivation for this thesis

While the pathological hallmarks of AD are well-established, the disease itself remains remarkably heterogeneous [17, 18, 19, 20, 21]. Variation in the amount, distribution, and evolution of pathology leads to diverse clinical outcomes that cannot be fully explained by current categorical frameworks, and current biomarkers often fail to capture the full spectrum of pathological variation.

This thesis is motivated by the need for more precise, quantitative biomarkers that better reflect this heterogeneity. The following chapters present a series of models and data-driven biomarkers designed to address this challenge.

Chapter 3

The effect of the reference region in standardised uptake value ratios in tau PET

The initial work that follows (section 3.2) led to the presentation of a poster at the New Horizons in Alzheimer's Disease conference (NHAD) in Leuven, Belgium, in November 2021. A further analysis (section 3.3) was presented at the Alzheimer's Association International Conference in Philadelphia, July 2024.

3.1 Introduction

As stated in Chapter 2, positron emission tomography (PET) enables precise spatial and temporal measurement of Alzheimer's disease pathology in the brain. The default quantification measure is the standardised uptake value ratio (SUVR) — radiotracer uptake in a target region of interest normalised by uptake in a reference region.

In this chapter, we examine how the choice of reference region influences SUVR consistency and explore whether combining multiple reference regions can improve signal stability. We first conduct exploratory analyses to identify inconsistencies in SUVR calculations across commonly used reference regions (Section 3.2). We then perform an updated quantification analysis using test–retest metrics to assess reliability (Section 3.3). Finally, we propose a method for combining reference regions to mitigate these inconsistencies and enhance stability (Section 3.4).

3.2 SUVr inconsistencies

3.2.1 Introduction

The quantification of regional PET measures is commonly the standardised uptake value ratio (SUVr), where a region of interest (ROI) standardised uptake value (SUV) is divided by the SUV of a region of reference. The region of reference is supposed to be free of pathology, and provides an intra-scan reference measure that can cancel out factors affecting the measurements (e.g., radiodecay of the tracer). The cerebellar grey matter is the most commonly used reference region for tau PET[62, 121]. There is an alternate proposal of a reference region, which is a subject-specific subset of voxels of the subcortical white matter (PERSI[122]), where they fit a bimodal Gaussian distribution and discard the component with the highest mean intensity. Later, subsets of the anatomical reference regions were proposed, to avoid tracer binding contamination: an inferior part of the cerebellar grey matter was proposed as a better ROI[62], recently proposed for cross-sectional studies, while an eroded subcortical white matter was proposed for longitudinal studies[123], similar to PERSI, but using erosion as a morphological operation.

Here, we examine basic assumptions regarding the two reference regions proposed by the ADNI tau PET curators: the inferior cerebellar grey matter, and an eroded white matter, as per 2021. Since the goal of a reference region is to obtain a quantitative measure comparable within scans of the same radiotracer, and reference regions are pathology-free, 1) all of the reference regions should be correlated, and 2) Once a reference region is used, the other should not explain measurement changes in paired longitudinal scans.

3.2.2 Materials

The data in this study consists of regional standardised uptake values (SUV) from 1153 [^{18}F]-AV-1451 tau PET scans from 770 different individuals from the ADNI database (adni.loni.usc.edu), with the cognitive distribution depicted in Table 3.1. Labels are cognitively unimpaired (CU), mild cognitive impairment (MCI) and Alzheimer's disease dementia (AD dementia).

Averaged Standardised Uptake Values (SUV) from subcortical regions are taken from the partial volume corrected tau-PET summary files. Clinical diagnosis is taken from the diagnosis summary table. Off-target binding regions [62] are those in

	CU	MCI	AD dementia
CU	417	14	1
MCI		234	16
AD dementia			88

Table 3.1: Distribution of individuals and their respective clinical status in the different visits. A total of 417 individuals were assessed as CU for all their visits, while 14 had CU and MCI labels throughout the different visits. None of them was assigned the three clinical diagnosis labels within the scope of all visits.

which the radiotracer accumulates in places other than the target tau tangles. They (choroid plexus, caudate, putamen, pallidum, thalamus and hippocampus -due to signal bleeding from the choroid plexus) are removed from this study, as is common practice in all disease studies involving PET. We will consider two reference regions: the inferior cerebellar grey matter (gm) and an eroded subcortical white matter (wm)[123].

3.2.3 Methods

We assessed longitudinal SUVR stability (consistency) for each reference region and examined whether incorporating a second reference region improved stability. For each participant with at least two scans, we calculated the ratio of SUVR values between the follow-up and baseline scans for regions with no expected pathology change. Ratios close to unity (1) indicate high stability, with deviations attributed to measurement noise.

To test whether the second reference region could explain residual variability, we computed the Spearman rank correlation between the SUVR ratio from one reference region and the SUV ratio from the other. Only cognitively unimpaired individuals were included to minimise the influence of disease progression.

As an illustrative example, we used the brainstem, a region expected to remain pathology-free, and plotted the follow-up/baseline SUVR (gm) against the follow-up/baseline SUV (wm). We repeated this analysis for all regions, estimating 95% confidence intervals for the correlations using the Fisher transformation.

Finally, we explored a simple correction method by fitting linear regression models to the paired ratios for each region, using the slope as an indicator of the relationship between the two reference region measures.

3.2.4 Results

Figure 3.1 is a scatterplot of SUV values for the two reference regions considered. The Spearman correlation is 0.072 ($t = 2.48$, $df = 1182$, $p\text{-value} = 0.013$), indicating extremely low correlation. The R^2 value is 0.005, so a negligible part of the variance is explained by this relationship.

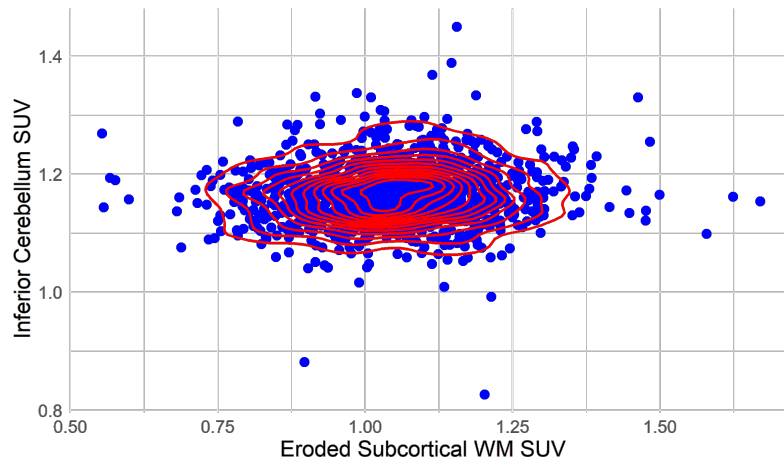


Figure 3.1: SUV of two commonly used reference regions: inferior cerebellar grey matter and eroded subcortical white matter. Pearson correlation 0.072.

Figure 3.2 shows the brainstem SUVR (using the inferior cerebellum grey matter as reference), part of a composite reference region in both amyloid PET [124] and tau PET [78]); no change of AD pathology is expected. We observe a noticeable correlation with respect to the subcortical white matter (Spearman: 0.70, $p < 10^{-5}$). A simple regression line is fitted as a visual aid. This suggests that a significant portion of the inconsistency can be resolved by utilising the second reference region (wm).

Figure 3.3 extends the analysis for the rest of the regions, and shows the Spearman correlation (“future SUVR_{gm} divided by past SUVR_{gm}” against “future wm SUV divided by past wm SUV”), along with their 95% CI (using the Fisher transformation).

Linear regression slopes fitted to the paired ratios were significantly different from zero for most regions, again consistent with the correlation results (Figure 3.4). The magnitude and direction of slopes varied across regions, suggesting that a single global correction factor would not be optimal.

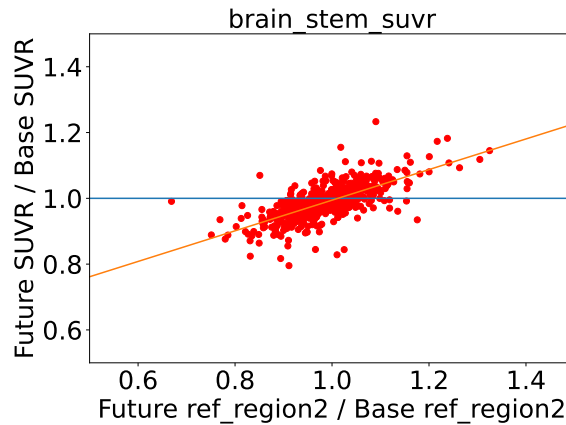


Figure 3.2: Future/baseline brainstem SUVR(gm) versus future/baseline eroded sub-cortical white matter SUV. Ideally, the pattern left after a signal correction, when plotted against another variable should be zero rank-correlated. Otherwise, like in this example, we know that there is a part of the signal variation that can still be accounted for.

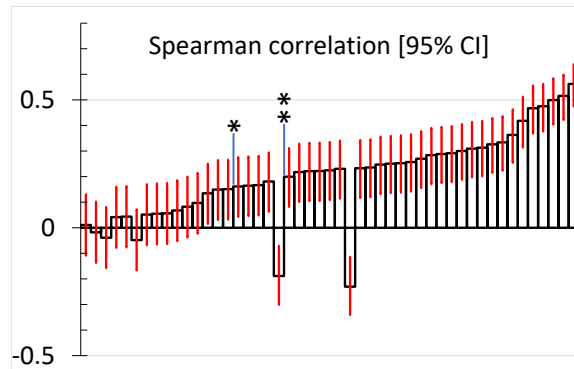


Figure 3.3: Spearman correlation of all regions (“future SUVRgm divided by past SUVRgm” against “future wm SUV divided by past wm SUV”), along with their 95% CI (using the Fisher transformation). Regions to the right of (*) have a Spearman correlation p-value smaller than 0.01. Regions to the right of (**) have a Spearman correlation p-value smaller than 0.001.

3.2.5 Discussion.

A shocking first finding is that the SUVs of two reference regions are not well-correlated (Figure 3.3). This alone should encourage further examination of the SUVR principle, as it suggests that at least one of the reference regions is not fit for the purpose of transforming the SUV into a more stable SUVR measure. The widely used method of normalising the $[^{18}\text{F}]$ -AV-1451 tau-PET signal with the inferior cerebellum grey matter is shown to lack consistency in paired scans of cognitively unimpaired that we do not attribute to random noise (Figures 3.2, 3.3, 3.4), consistent with [123]. The same Figures showed how using both reference

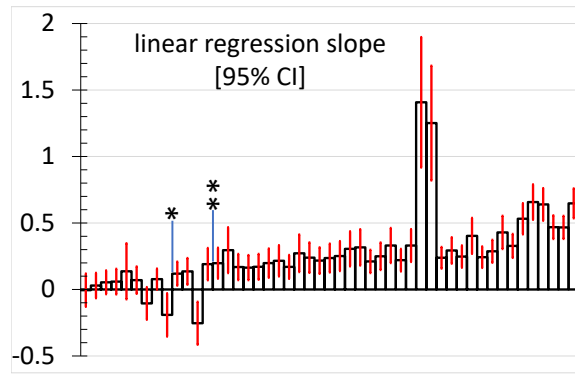


Figure 3.4: Slope of linear regression fit to plots as in Figure 3.2, along with their 95% CI. Regions to the right of (*) have a p-value smaller than 0.01 on the hypothesis that the slope is 0. Regions to the right of (**) have a p-value smaller than 0.001.

regions simultaneously can achieve better consistency between scans, as opposed to using the inferior cerebellar grey matter. Additionally, due to the different behaviour of regions in the brain, the new method could see some improvement by adjusting to each region. Further analysis with more datasets could serve to validate the findings. Furthermore, the different regional behaviour could be analysed to better understand the physical and biological causes, especially with different radiotracers.

3.3 Reference Region effect on test-retest metrics

3.3.1 Introduction

Evaluations of new radiotracers commonly assess test-retest consistency in scans with no pathological changes, but rarely question the SUVR paradigm. Notably, even the Centiloid [80] and CenTauR [81] scales (PET normalisation methods to compare measures among radiotracers) depend on SUVR, and thus, the choice of reference region. This section aims to quantitatively assess the impact of different reference regions on PET scans. Additionally, we will examine the impact of the partial volume corrected (PVC) data provided by the ADNI study. PVC is used in imaging analysis to address the partial volume effect, where individual image voxels contain a mixture of signals from different tissues.

3.3.2 Methods

Data. Data used in the preparation of this study were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). Additionally, we used data from the A05 Study [125]. Both datasets used the ^{18}F -AV-1451

radiotracer for tau PET brain scans. Data from ADNI was examined with, and without partial volume correction (using the Geometric Transfer Matrix approach [126]).

Data study	PVC	CU		MCI	
		Individuals	Scans	Individuals	Scans
ADNI	Yes	206	529	140	369
ADNI	No	210	537	141	372
A05	No	18	36	17	34

Table 3.2: Number of available unique individuals and scans per study. PVC indicates partial-volume-corrected data. CU = cognitively unimpaired. MCI = mild cognitive impairment.

Evaluation. We analyse how stable the repeated measure of tau PET is, with various target composite regions (Braak Stages 5 and 6 [127], the latest of the regions to accumulate tau), in cognitively unimpaired (CU) and mild cognitive impairment (MCI) individuals. In such a cohort, we expect minimal pathology in these regions, and therefore, expect high test-retest consistency. We evaluate the whole cerebellum, the inferior cerebellar grey matter[62], and a composite reference region (average of whole cerebellum, eroded subcortical white matter and brainstem)[124].

Evaluation metrics follow test-retest literature [128]: percent test-retest (PTRT, equation 3.1), intraclass correlation coefficient (ICC, equation 3.2), and within-subject coefficient of variation (WSCV, equation 3.3). If repeated measures of regions that should not see noticeable pathological changes are taken, then ideally we would expect a PTRT of 0% (lowest possible value), an ICC of 1 (highest possible value), and a WSCV of 0 (lowest possible value).

Formulas of each metric are listed below, reproduced from Baumgartner et al. [128]:

$$PTRT = \frac{1}{2} \sum_{i=1}^n \left| 2 \frac{y_{i2} - y_{i1}}{y_{i2} + y_{i1}} \right| \cdot 100\% \quad (3.1)$$

Where n is the number of subjects in the study and y_{i1} and y_{i2} are the estimated PET outcome measures obtained for the i -th subject in a given region in the baseline and future scan, respectively.

$$ICC = \frac{\sigma_S^2}{\sigma_S^2 + \sigma_e^2} \quad (3.2)$$

$$WSCV = \frac{\sigma_e}{\mu} \quad (3.3)$$

Where σ_s and σ_e are the between- and within-subject standard deviations, respectively, and μ is the mean.

3.3.3 Results

Table 3.3 shows the test-retest metrics (WSCV, ICC and PTRT) for Braak stage 5 and 6 composite regions, calculated using various reference regions in the A05 and ADNI datasets. The ADNI data was analyzed both with and without partial volume correction (PVC).

Partial volume correction. The partial volume corrected data dramatically altered the test-retest metrics, as we see that all metrics, for all reference regions, worsen considerably: WSCV and PTRT worsen by a factor of about 2, and ICC drops between 3% and 15%.

Within-subject coefficient of variation. The composite reference performs best, being 7-8% better in Braak 5 and 4-10% better in Braak 6 compared to the whole cerebellum. Compared to the inferior cerebellum gray matter, the composite reference is 13% and 12% better in Braak 5 and 6, respectively.

Intraclass correlation coefficient. The composite and the whole cerebellum perform similarly, with the composite having an edge (6 and 9% higher ICC in Braak 5 and 6, respectively) in the biggest dataset (ADNI). The inferior cerebellum grey matter had the worst performance, with over 16% lower ICC in ADNI and 12-20% lower ICC in A05, compared to the composite reference region.

Percent test-retest. The inferior cerebellum gray matter performs worse than the other two, with 7-13% higher PTRT. The composite reference region has the best performance here, with the lowest metrics in all Braak stages and datasets, except for Braak 5 in A05, where the whole cerebellum has a 3% lower PTRT.

3.3.4 Discussion

The findings of this study highlight the considerable influence of reference region selection and image processing techniques, specifically partial volume correction (PVC), on the test-retest reliability of 18F-AV1451 tau PET measurements.

The relatively good test-retest performance observed in the A05 study, using

18F-AV1451 tau PET (A05 study)						
	WSCV (x100)		ICC		PTRT (median)	
Reference used	Braak5	Braak6	Braak5	Braak6	Braak5	Braak6
composite.ref	2.61	2.19	0.96	0.92	2.75	2.37
whole.cerebellum	2.84	2.44	0.97	0.94	2.67	2.49

18F-AV1451 tau PET (ADNI study)						
	WSCV (x100)		ICC		PTRT (median)	
Reference used	Braak5	Braak6	Braak5	Braak6	Braak5	Braak6
composite.ref	3.78	3.02	0.86	0.83	2.88	2.68
inferior.cerebellum	4.35	3.41	0.87	0.86	3.22	2.85
whole.cerebellum	4.08	3.13	0.87	0.86	2.88	2.53

18F-AV1451 tau PET (partial volume corrected, ADNI study)						
	WSCV (x100)		ICC		PTRT (median)	
Reference used	Braak5	Braak6	Braak5	Braak6	Braak5	Braak6
composite.ref	7.45	7.50	0.82	0.72	5.88	6.67
inferior.cerebellum	7.50	7.01	0.84	0.77	5.23	5.71
whole.cerebellum	7.62	7.23	0.84	0.75	5.40	5.92

Table 3.3: Comparison of WSCV, ICC, and PTRT for different reference regions in 18F-AV1451 tau PET studies. Individuals with dementia were excluded.

both the whole cerebellum and a composite reference region, aligns with previous research suggesting the suitability of the cerebellum as a reference for tau PET in populations without extensive tau pathology [129, 125]. The slightly better performance of the composite region in the A05 data is likely due to the inclusion of white matter, potentially reducing noise and variability. However, the sample size of this cohort (after inclusion criteria) was rather small, with 35 total individuals, as compared to the 351 in ADNI.

Within the non-PVC ADNI data, the composite reference region consistently outperformed the other two, suggesting that averaging signal across multiple regions might provide some robustness against noise, even in a larger and potentially more heterogeneous dataset. The inferior cerebellum, while advocated for in the literature [62] and data curators of the ADNI PET data, did not show superior performance compared to the whole cerebellum in this specific dataset.

The most visible finding was the detrimental effect of PVC on test-retest reliability. While PVC is intended to improve quantitative accuracy by correcting for signal blurring between adjacent regions, our results suggest that it can introduce significant variability in repeated measures, at least for 18F-AV1451 and the specific

implementation used in the ADNI dataset. This finding contrasts with the intended purpose of PVC and raises concerns about its application in test-retest studies of tau PET SUVR, particularly when using this tracer. It's possible that the PVC algorithm, while correcting for partial volume effects, amplified noise or introduced artifacts that disproportionately affected the test-retest metrics. This result underscores the need for careful validation of image processing techniques in the context of longitudinal studies. A previous study on longitudinal amyloid PET [130] found that this PVC technique (GTM) had significantly worse relative precision compared to other methods, when measuring within-person change over time. Given this poor performance of PVC data in ADNI for 18F-AV-1451 tau PET, I revisited the analysis from the previous section (3.2), and found out that using non-PVC data, the coefficient of determination between inferior cerebellum SUV and eroded subcortical white matter was $R^2 = 0.12$; while an improvement over the 0.005 of PVC data, it is still a low percentage of explained variance.

Several limitations should be considered. First, the specific PVC method (Geometric Transfer Matrix) may not represent all PVC approaches, potentially affecting generalisation. Second, only a single radiotracer was evaluated (18F-AV-1451), so a multitracer study would be needed to confirm a common, well-performing, reference choice. Third, we didn't have actual test-retest data; however, we did filter the analysis to include only late Braak stage regions, and individuals with no- or mild cognitive impairment.

Future research could focus on:

1. Investigating the impact of different PVC algorithms and parameters on tau PET test-retest reliability.
2. Exploring and validating data-driven approaches for reference region selection that are less susceptible to noise.
3. Conducting test-retest studies with larger and more diverse populations, including individuals with varying degrees of tau pathology.
4. Extending this investigation to other tau PET radiotracers.

In conclusion, this study demonstrates that the choice of reference region and the application of GTM PVC can considerably impact the test-retest reliability of

18F-AV1451 tau PET measurements. The dramatic negative effect of PVC raises significant concerns about its use in longitudinal tau PET studies. Surprisingly, the widely used inferior cerebellum grey matter did not perform as well as the other considered reference regions. The composite reference region (whole cerebellum, eroded white subcortical matter and brainstem) had the best performance, motivating a data-driven search for optimal region compositions.

3.4 Combination of multiple references for SUVR quantification

3.4.1 Introduction

In Section 3.2 we demonstrated the potential of combining two reference regions to enhance the robustness of PET measurements. This section extends the concept of multi-region referencing to an arbitrary number of reference regions, providing a generalised mathematical framework for SUVR calculation.

3.4.2 Methods

The traditional SUVR calculation can be viewed geometrically. The numerator (target region SUV) and the denominator (reference region SUV) can be represented as components of a two-dimensional vector. The SUVR is then equivalent to the tangent of the angle (α) between this vector and the reference region axis (Figure 3.5-A). This geometric interpretation is not limited to two dimensions. We can extend this concept to a higher-dimensional space, incorporating multiple reference regions. Figure 3.5-B illustrates this for three dimensions (one target and two reference regions). In an n -dimensional space, we have one target region SUV and $(n-1)$ reference region SUVs. The general formula is derived below.

$$\cos\alpha = \frac{\vec{j} \cdot \vec{r}}{\|\vec{j}\| \|\vec{r}\|} = \frac{\sum r_i^2}{\sqrt{t^2 + \sum r_i^2} \cdot \sqrt{\sum r_i^2}} = \frac{\sqrt{\sum r_i^2}}{\sqrt{t^2 + \sum r_i^2}} = \frac{A}{B} \quad (3.4)$$

Where t is the SUV of the target region of interest. The vector $\vec{j}$ is $(t, r_1, \dots, r_n)$, and the vector $\vec{r}$ is $(0, r_1, \dots, r_n)$, where n is the total number of reference values, each with its own SUV value of r_i . The angle α is the angle between the vectors $\vec{j}$ and $\vec{r}$. As depicted in Figure 3.5-C, we visualise this concept using a generic right triangle with one of the acute angles being α , where the hypotenuse is B , the

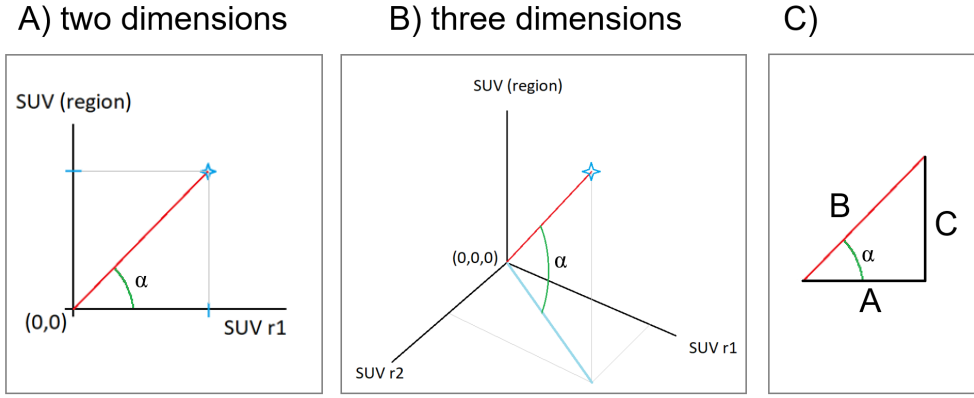


Figure 3.5: Geometric representation of SUVR and its generalisation. (A) Traditional SUVR in 2-dimensions, where the ratio is equivalent to the tangent of the angle α . (B) Extension to 3-dimensions with two reference regions. (C) generalised right triangle representation, depicting the numerator as the opposite cathetus and the denominator as the adjacent cathetus to α

adjacent cathetus is A , and the opposite cathetus is C .

Given A and B we can derive C using Pythagoras' theorem:

$$C = \sqrt{B^2 - A^2} \quad (3.5)$$

Finally, we can define the general formula for the ratio biomarker:

$$\text{biomarker} = \tan \alpha = \frac{C}{A} = \frac{\sqrt{B^2 - A^2}}{A} = \frac{\sqrt{t^2 + \sum r_i^2 - \sum r_i^2}}{\sqrt{\sum r_i^2}} = \frac{t}{\sqrt{\sum r_i^2}} \quad (3.6)$$

Finally, we introduce a scaling factor of $\sqrt{n}$. This normalisation ensures that if we were to subdivide a single reference region (with $\text{SUV}=1$) into any number of subregions, the resulting ratio would remain unchanged. As it is a factor, it does not introduce noise or affect the signal information: a classifier performance would remain unchanged. So the final, general formula is given below.

$$\text{biomarker} = \frac{t \cdot \sqrt{n}}{\sqrt{\sum r_i^2}} \quad (3.7)$$

3.4.3 Discussion

A mathematical model was derived to accommodate the usage of multiple reference regions. The usefulness of the model depends on how much information the

reference regions have. Although we derived a universal ratio, we still have not explored which regions are useful as references... or targets.

This section presents a mathematical framework for incorporating multiple reference regions into PET quantification, extending the traditional SUVR approach. The derivation ensures that the new model is mathematically consistent with SUVR in the special case of a single reference region, and it maintains invariance when a reference region is subdivided into subregions. The potential advantage lies in its ability to leverage information from multiple reference regions, potentially increasing the stability of the resulting biomarker. By combining information from multiple regions, the new composite may be less susceptible to noise or artifacts affecting individual reference regions. On the other hand, the practical utility of using multiple regions as reference depends on each one contributing independent and meaningful information. If a chosen region contains non-pathological binding of the tracer, combining them may not improve, and could even worsen, the accuracy and reliability of the quantification. Therefore, careful selection and validation of reference regions remains an important task. Further analysis is needed to empirically evaluate the performance of this new conceptual merging of reference regions compared to traditional SUVR and other normalisation methods, and to identify optimal combinations of reference regions.

3.5 Conclusions.

This chapter investigated the critical role of the reference region in tau PET quantification using SUVR, challenging common assumptions and proposing a new approach for combining multiple reference regions. The work presented across three sections reveals inconsistencies in current practices, demonstrates the impact of reference region choice and image processing on test-retest reliability metrics, and introduces a generalised mathematical framework for multi-reference region normalisation.

Key Findings and Contributions:

1. Inconsistency of Standard Reference Regions (Section 3.2): The widely used inferior cerebellar grey matter and eroded subcortical white matter, commonly regarded and employed as reference regions for 18F-AV1451 tau PET, showed surprisingly poor correlation in the ADNI dataset. This finding directly chal-

lenges the assumption that these regions are interchangeable and highlights a fundamental problem with the current approach for normalisation. Furthermore, it was shown how measurements taken using one reference region can be improved using information from the other, proving that a single reference region might not be optimal.

2. Impact of Reference Region Choice on Test-Retest Reliability (Section 3.3): The choice of reference region vastly affected the test-retest metrics (WSCV, ICC, PTRT) of 18F-AV1451 tau PET. While a composite reference region (combining whole cerebellum, eroded subcortical white matter, and brainstem) generally performed best in the non-PVC ADNI data, the commonly advocated inferior cerebellar grey matter surprisingly did not consistently outperform the whole cerebellum.
3. Detrimental effect of partial volume correction (Section 3.3): The application of partial volume correction (PVC) using the Geometric Transfer Matrix method in the ADNI data substantially worsened test-retest reliability across all tested reference regions. This unexpected result raises serious concerns about the routine use of PVC in longitudinal tau PET studies, particularly with 18F-AV1451, and highlights the need for rigorous validation of image processing techniques.
4. A multi-reference framework (Section 3.4): A mathematical framework was derived to generalise the SUVR concept to an arbitrary number of reference regions. This framework provides a consistent and scalable approach for combining information from multiple reference regions, potentially improving the robustness of PET quantification. The formula ensures invariance when a reference region is subdivided.

In summary, this chapter provides compelling evidence for the need to move beyond simplistic assumptions about reference regions in tau PET, and possibly in PET in general. By highlighting inconsistencies, demonstrating the impact of methodological choices, and proposing a new conceptual multi-reference framework, this work contributes to a more rigorous approach to tau PET quantification, ultimately advancing our ability to study and understand Alzheimer's disease and

related imaging-based analysis.

Chapter 4

Composite Value Ratio (CVR)

The work described in this chapter was first introduced at the Alzheimer's Association International Conference in 2022, and later expanded to a paper, peer-reviewed and published in Brain Communications [December 2024, doi.org/10.1093/braincomms/fcae438]. Supplementary files mentioned throughout the chapter can be accessed online: code, and a spreadsheet with all results (Results.xlsx). Appendix A includes supplementary tables and figures. Neil Oxtoby, as my supervisor, contributed to the submitted work. An updated code (May 2025, from a work in progress) is available on GitHub (<https://github.com/isaac-6/biodiscvr>), with the possibility to have multiple cohorts for federated discovery, and with added regularisation terms to avoid convergence to a Pareto frontier.

4.1 Introduction

Chapter 3 highlighted notable differences depending on the reference region used, and showed that combining multiple regions (as with the composite reference region proposed by Landau et al. [124]) could improve signal correction in tau PET studies of neurodegenerative disease. For the purpose of understanding the progression or existence of a neurodegenerative disease such as Alzheimer's, a target region of interest with measurable pathology needs to be chosen. That's how the SUVR is defined: the value of a target region with respect to the value of a reference region (see Background Chapter 2). What would happen if we let an algorithm freely choose which regions go in the numerator and which in the denominator? Now that we know something is at fault, and we know how we can improve it, we need a specific, quantitative measure, to evaluate the biomarkers. This choice will be primarily the sample size estimate for a hypothetical clinical trial, since clinical

trials are the primary use case for tau PET in Alzheimer’s disease. The better the longitudinal trajectory of a biomarker is (similar trajectories for all individuals, with low residual errors fitting a model, or variance), the lower the sample size estimate needed to detect trajectory changes (e.g., in a clinical trial).

4.2 Materials and Methods

To avoid confusion with the current understanding of SUVR, we call the new concept of biomarker a composite value ratio (CVR, pronounced “cover”), which is the ratio of two regions of interest in the brain. The main difference is that traditional SUVR relies on biological definitions of the numerator (commonly referred to as target) and the denominator (commonly referred to as reference), while our concept of CVR is a data-driven ratio designed for a specific purpose (e.g., for a particular group and clinical trial), and is reference-agnostic (see 4.2.3). We call the framework for biomarker discovery *BioDisCVR* (“bio-discover”), and we describe it in section 4.2.3. To test the usefulness of CVR as a biomarker for tau PET modelling over time, we set up three experiments. The experiments described below are designed to discover CVRs that minimise the sample size estimate in modern randomised controlled clinical trials on Alzheimer’s disease. CVR biomarkers are compared against relevant benchmarks from the literature, within and across multiple datasets.

4.2.1 Study design

Figure 4.1 conceptualises the *BioDisCVR* framework applied to reducing clinical trial sample sizes by finding an appropriate CVR for the trial design. The algorithm is described in *BioDisCVR* Framework and Benchmarks 4.2.3. Input features are regional brain volumes and regional PET SUV values (described in Data 4.2.2). The algorithm searches through the combinatorial space (numerator and denominator simultaneously) to find a ratio of composite regions driven by an application-specific fitness function. Here, we demonstrate *BioDisCVR* using two scenarios relevant to modern clinical trials on Alzheimer’s disease:

- Experiment 1: 54-month secondary prevention trial in amyloid-positive, cognitively unimpaired individuals at risk of Alzheimer’s disease (cf., the A4 Study [132]).
- Experiment 2: 18-month treatment trial in amyloid-positive, cognitively impaired

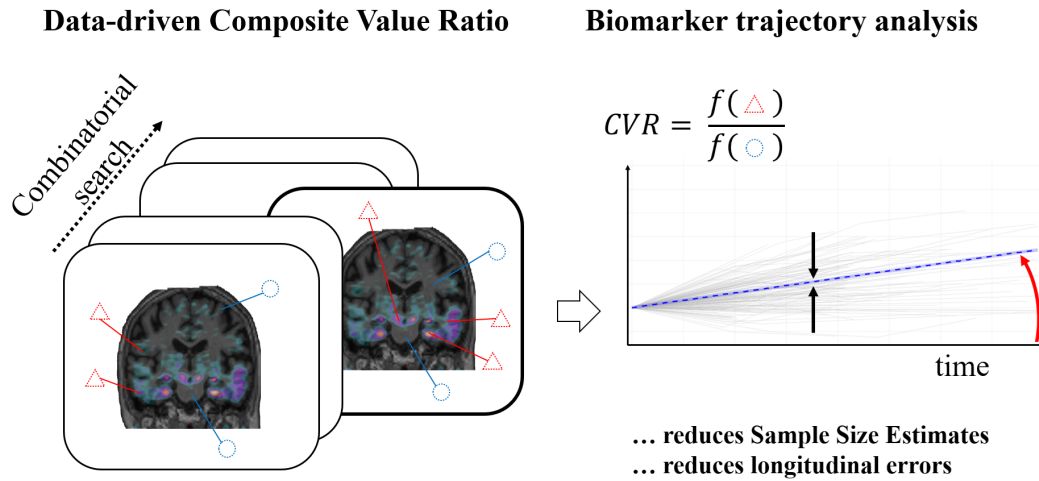


Figure 4.1: Conceptual Representation of the BioDisCVR Framework. A search algorithm is deployed to find a well-performing biomarker, defined as the ratio of signals $f(\text{triangles})/f(\text{circles})$ in two composite brain regions ($f(\text{triangles})$ and $f(\text{circles})$). A lower variance of trajectories (black arrows) and a higher gradient (red arrow) of the biomarker over time will contribute to lowering the sample size estimate for a hypothetical clinical trial. Figure reproduced from Llorente-Saguer and Oxtoby [131].

individuals –cf., recent prominent trials of monoclonal antibodies:

- Aducanumab (ENGAGE/EMERGE [57]),
 - Lecanemab (Clarity-AD [79]), and
 - Donanemab (TRAILBLAZER-ALZ [75]).
- Experiment 3: The same as Experiment 2, but with a subset of regions common to multiple datasets, to facilitate external validation. Here we use the Alzheimer’s Disease Neuroimaging Initiative (ADNI) dataset for discovery and two datasets from the Mayo Clinic for validation (ADRC and MCSA). The datasets are described below in Data 4.2.2.

The fitness function drives the algorithm to simultaneously reduce the sample size estimate (SSE) for a given clinical trial design, while maximising cognitive group (biomarker) gradient separation (see Statistical Analysis 4.2.5). We report the SSE for 20% effect size and 80% power. Sample size estimate is an inverse square function of the effect size, so changing the effect size of the clinical trial design will not change the performance ranking of the biomarkers. Extensive supplementary experiments are performed with multiple random initialisations, alternative evaluation

metrics, restricted search spaces (Appendix A Table A.1) and regional ablation (Supplementary Figure A.1) to aid interpretability and comparability with the literature, while also demonstrating robustness and flexibility of the BioDisCVR framework.

4.2.2 Data

This study uses a discovery dataset to find and evaluate biomarkers, and an additional two cohorts for validation.

Discovery dataset: the Alzheimer's Disease Neuroimaging Initiative (ADNI). Data analysed is regional brain tau PET signal and MRI volumes from the ADNI database (downloaded on 12 August 2024 from adni.loni.usc.edu), using participants with available processed [^{18}F]-AV-1451 tau PET scans. Our inclusion criteria consist of being amyloid-positive at any visit (composite amyloid-PET SUVR above or equal to 0.78 for Florbetapir F18 and 0.74 for Florbetaben F18)[133] and having at least two visits with concurrent tau-PET and T1w MRI scans. Although the linear mixed model we use needs three longitudinal measures per individual to return a measure of the residuals, we can (and do) include data from individuals having only two scans, as this data contributes to the intercept and gradient, forcing the model to fit to a larger sample. Diagnosis is defined by the individual's maximum level of cognitive impairment across available visits. Of the 391 participants having two or more eligible visits, 198 were amyloid-positive (48.7% females). Among these, 76 were cognitively unimpaired (CU), 122 cognitively impaired (CI, which is the aggregation of individuals with either mild cognitive impairment or dementia due to probable Alzheimer's disease), of which 71 were labelled to have dementia due to probable Alzheimer's disease. The PET scans were acquired across 48 different sites, with voxel resolutions of 6mm and 8mm. The majority of participants (about 94%) identify as belonging to a white ethnicity (see Discussion).

Validation datasets: Mayo Clinic Alzheimer's Disease Research Center (ADRC) and Mayo Clinic Study of Aging (MCSA). Individuals with dementia due to aetiologies other than Alzheimer's disease were discarded for this study (N=5). The same criteria as with ADNI data were applied (three visits with [^{18}F]-AV-1451 tau PET, structural MRI scan and a measure of amyloid), with the exception of the amyloid cutoff: since the composite reference region is not available in these datasets, we used the available regions from the cortical summary region[133] (that is made

up of frontal, anterior/posterior cingulate, lateral parietal, lateral temporal regions: full list in Supplementary file Results.xlsx, tab “amyloid.target”) [131] and “inferior cerebellum grey matter” as the reference. Then, we fit a 2-component Gaussian mixture model [134] (see Supplementary Figure A.2) and derived the cutoff of 1.53 for the ^{11}C -Pittsburgh Compound B PET, to determine amyloid positivity. This reduced the available individuals to 43 (1 CU, 13 MCI and 29 AD). Because of the small sample size of the CU group, we only used the available 42 cognitively impaired individuals (33.3% females).

Table 4.1: Proportion of individuals per interval time between visits. Most individuals had visits separated by 0.7 to 1.5 years. Data from the Alzheimer’s Disease Neuroimaging Initiative (ADNI) and from the Mayo Clinic Alzheimer’s Disease Research Center (ADRC) and Mayo Clinic Study of Aging (MCSA), the last two grouped in the Mayo column. Abbreviations: CU = cognitively unimpaired; CI = cognitively impaired. Table reproduced from Llorente-Saguer and Oxtoby [131].

Time between visits (years)	Discovery (ADNI)		Validation (Mayo)
	CU	CI	CI
$t \leq 1.5$	61%	72%	76%
$1.5 > t \leq 2.5$	22%	17%	18%
$2.5 > t \leq 3.5$	6%	6%	6%
$t > 3.5$	11%	5%	0%

Time between visits is shown for all experiments in Table 1. The majority of intervals are from 0.7 to 1.5 years, but some span 3 years. We did not use partial-volume-corrected SUVR data, as it has been shown to introduce longitudinal errors [78, 76]. BioDisCVR input features, as with classical machine learning models, can be the aggregation of any numerical variables. In our case, they consist of regional PET SUV and volumes for brain regions. In the Discovery dataset (ADNI), regions are defined by the Desikan-Killiany atlas [135], with the data provided on the Image & Data Archive (IDA) run by the University of Southern California Laboratory of Neuro Imaging (LONI) for download in CSV format. For details of the processing pipelines, see ADNI’s manuscript [136]; this pipeline depends on a contemporaneous native space MRI that is segmented and parcellated with Freesurfer v7.1.1 (surfer.nmr.mgh.harvard.edu) [137]. For the Validation datasets, volume measures were obtained by the Mayo Clinic using the second version of the SUI atlas [138]. Off-target binding regions were excluded to ensure accurate quantification of tau pathology by minimising interference from non-specific binding [136]. The complete

list of brain regions made available to our model is provided in the Supplementary File Results.xlsx (“regions.available” tab) of the online version [131].

4.2.3 BioDisCVR Framework

BioDisCVR is a data-driven optimisation framework designed to discover novel, ratio-based biomarkers that are tailored to a specific, user-defined objective (e.g., statistical power for a specific clinical trial scenario). It is built upon three core components: (1) a flexible biomarker model, (2) a search and optimisation engine, and (3) a customisable fitness function.

In this study, we define and use our framework BioDisCVR to discover biomarkers tailored for a given clinical trial design. Here we consider both preclinical and clinical trials (in CU/CI individuals, respectively) of fixed duration (see Study Design 4.2.1).

The Biomarker Model: Composite Value Ratio (CVR). The framework’s output is a Composite Value Ratio (CVR), which moves beyond the traditional, biologically pre-defined ‘target’ and ‘reference’ regions of SUVR. A CVR is a ratio of two composite regions, where each composite is a combination (e.g., simple or volume-weighted average) of signals from any number of input features (e.g., brain regions’ measures). This structure is inherently reference-agnostic and allows the data itself to determine the optimal combination of regions for both the numerator and denominator. As shown in Equation 4.1, this formulation ensures that if SUVR values are used as input, the arbitrary reference region cancels out, making the final CVR dependent only on the underlying SUV signals. Specifically,

$$\text{CVR} = \frac{\text{composite}_1 \text{ SUVR}}{\text{composite}_2 \text{ SUVR}} = \frac{\text{composite}_1 \text{ SUV} / \text{reference SUV}}{\text{composite}_2 \text{ SUV} / \text{reference SUV}} = \frac{\text{composite}_1 \text{ SUV}}{\text{composite}_2 \text{ SUV}} \quad (4.1)$$

Where composite_i is the signal from any number of regions, combined in any way (e.g., volume-weighted, averaged).

The Search and Optimisation Engine. To discover the optimal CVR from an intractably large combinatorial space (over 10^{40} permutations in this work), the framework employs an objective-driven genetic algorithm [139]. Potential biomarkers are encoded in a vector, where integer values map each input feature to the numerator,

the denominator, or exclusion from the ratio. The algorithm then iteratively evolves a population of these candidate biomarkers to find the one that best maximises the fitness function.

The Fitness Function. The guiding principle of the discovery process is a user-defined fitness function. For the application in this Chapter, the objective was to find biomarkers with high statistical power for monitoring longitudinal trajectories, and thus useful for a hypothetical clinical trial. Therefore, the fitness function was defined as the ratio of disease signal (trajectory separation between cognitive groups) to measurement noise (the square of the Sample Size Estimate, SSE). This design forces the algorithm to find CVRs that are highly sensitive to disease-related change while remaining longitudinally stable.

Practical Implementation in this Thesis. For the specific application to tau PET, the search space is reduced (by choice) by using two biologically motivated priors: 1) no BraakI-III regions [127] in the denominator; 2) no commonly used SUVR reference regions in the numerator. We report the best result found after 300 generations of the genetic algorithm for each of the constraints or scenarios. The top 5 biomarkers found, as well as the effect of different random initialisations, are shown in Supplementary File Results.xlsx, “Random.seeds_top.n” tab of the online version [131]. Our aim is not to show convergence or a focus on the search algorithm; rather, the evaluation of the biomarkers found by the algorithm. Thus, the choice of 300 generations is arbitrary, although in 20 random initialisation experiments with 500 generations each, 90% of the final performance was achieved within a mean of 140 generations (and below a standard deviation of 100), for both clinical and preclinical cases. Furthermore, the generations alone are not an indicator of the total number of combinations tested, as the genetic algorithm population size should also be taken into account.

Pipelines. Our experiments include several pipelines to investigate laterality (right-left hemisphere differences) and region size effects. Laterality is explored through averaged bilateral regions (indicated by the suffix -B in the biomarker nomenclature) vs left/right independent (indicated by the suffix -L in the biomarker nomenclature). Region size effects are explored through volume-weighted SUV vs simple mean SUV (mSUV). Supplementary Table A.1 contains additional experiments where

either the numerator or denominator is defined a priori.

In essence, the BioDisCVR framework is defined by these core components: a flexible biomarker model (the CVR), a powerful search engine to navigate the combinatorial space, and a user-defined fitness function to guide the discovery towards a specific, quantifiable objective.

4.2.4 Baselines

We compare the performance of CVR to SUVR biomarkers from the literature: meta-temporal [140] (entorhinal cortex, parahippocampal gyrus, amygdala, inferior temporal gyrus, fusiform gyrus, middle temporal gyrus) divided by a composite reference region (eroded subcortical white matter, whole cerebellum, and brainstem) [124]; highest tau-PET-positive data-driven stage (DDS [141], an adaptive individualised approach) divided by the inferior cerebellum grey matter. Additionally, we consider all combinations of composite target and reference regions we found in the literature: Braak stages [127], meta-temporal [140], mesial-temporal [142], temporoparietal [142], “rest” [142] as target/numerator regions; and a set of popular reference regions (whole cerebellum, inferior cerebellar grey matter, composite reference region, eroded subcortical white matter, cerebellum cortex, brainstem, eroded subcortical white matter + inferior cerebellar grey matter, eroded subcortical white matter + whole cerebellum, whole cerebellum + brainstem, inferior cerebellar grey matter + brainstem, eroded subcortical white matter + inferior cerebellar grey matter + brainstem). Including composites calculated using both mean SUV and volume-weighted SUV, we compared our results with a total of 220 literature-inspired baseline biomarkers.

4.2.5 Statistical analysis

All analyses are performed in R [143] (version 4.3.2). Our analysis involves calculating SSE, longitudinal group separation, and longitudinal precision (defined below) for clinical trials on Alzheimer’s disease in either cognitively unimpaired (CU) or impaired (CI) participants. The SSE and group separation measures drive the genetic algorithm. Longitudinal group separation (t-statistic) and longitudinal precision (standard deviation of model residuals) of tau PET biomarkers are quantified using a linear mixed effects model [144] fit to $\log(\text{SUVR})$. Covariates sex, APOE4, and education were not included because they did not significantly impact the intercept

or the slope of the biomarker, replicating the analysis of Schwarz et al. [78]. Our linear mixed effects model includes correlated random slopes and intercepts, since this makes it invariant to time-shifts [144], and we use a relative time point (the average of all visits per individual). Furthermore, in the discovery cohort data (ADNI), a Chi-square test indicated that the correlated random effects model fits the data better than the uncorrelated model, both for the cognitively unimpaired ($\chi^2=13.38$, Df = 1, $p < 0.001$) and cognitively impaired ($\chi^2=13.39$, Df = 1, $p < 0.001$). This is the formula used:

$$\text{SUVR}_{ij} = \beta_0 + \text{time}_{ij}(\beta_1 + \beta_3 \cdot \text{DX}_{ij} + b_{1j}) + \beta_2 \cdot \text{DX}_{ij} + b_{0j} + \varepsilon_{ij} \quad (4.2)$$

Where:

- i indexes the observation,
- j indicates the individual,
- DX is a binary variable indicating whether the individual j belongs to the cognitively impaired group (1) or not (0),
- β_{0-3} are fixed effects,
- b_{0j} b_{1j} are random effects (intercepts and slopes varying by individual),
- ε is the residual error.

Cognitive group separation (CU vs CI) is quantified by the t-statistic between fixed effect gradients (β_3 in equation 4.2). The measure of longitudinal repeatability is quantified using the standard deviation of the model residuals (ε in equation 4.2), which is a measure of the relative error with respect to the biomarker values. Finally, we use the longpower package [145] to calculate the sample size estimate (SSE) [146] for each of two hypothetical clinical trials designed for 80% power to reduce tau PET accumulation by 20% vs. placebo: CU individuals, 54 months (Experiment 1, c.f. the A4 Study [132]); and CI individuals, 18 months (Experiment 2, c.f. ENGAGE/EMERGE [57], CLARITY-AD [79], TRAILBLAZER-ALZ [75]).

All metrics are provided with 95% confidence intervals, calculated using the model-based (semi-)parametric bootstrap for mixed models function from the lme4 package [144], version 1.1.35.1.

4.3 Results

Computation. Using a 12th Gen Intel® Core™ i7-12700H, with 2.70 GHz and 20 cores, we ran the search algorithm for less than three minutes per experiment (Supplementary File Results.xlsx, tab Random.seeds-top.n, of the online paper [131] shows different random initialisations of BioDisCVR).

Experiment 1 (hypothetical 54-month preclinical AD trial).

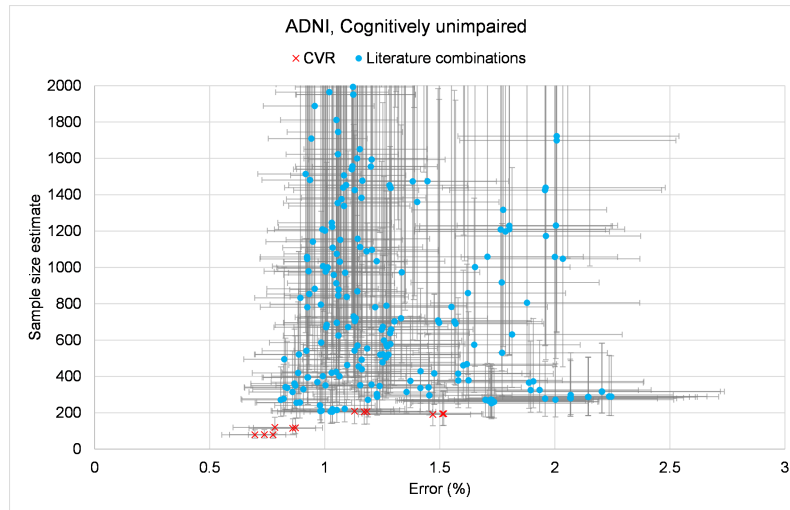


Figure 4.2: Biomarker performance for Experiment 1. The plot shows sample size estimate (SSE) against percentage error (longitudinal precision) for all biomarkers in Experiment 1, truncated at $SSE \leq 2000$. Composite value ratio (CVR) biomarkers (red crosses) showed superior performance to baseline comparators (blue dots), reducing both the SSE and the error. Error bars show 95% confidence intervals from bootstrapping. The SSE shown is the minimum number of individuals that would be needed per branch for a hypothetical preclinical trial (cognitively unimpaired individuals), designed for 80% power and 20% effect size, for a duration of 4.5 years. Figure reproduced from Llorente-Saguer and Oxtoby [131].

Figure 4.2 is a plot of sample size estimate (SSE) against percentage error (longitudinal precision) for all biomarkers in Experiment 1, truncated at $SSE \leq 2000$. CVR biomarkers (red crosses) showed superior performance to baseline comparators (blue dots). Error bars show 95% confidence intervals from bootstrapping (see Statistical analysis 4.2.5).

Table 4.2 highlights selected biomarkers for Experiment 1 (all results are included in Supplementary Material File Results.xlsx as a spreadsheet). Namely, specific literature biomarkers, plus the best-performing baseline (biomarker inspired by the literature) and our composite value ratio (CVR) biomarkers found by the

Table 4.2: Results of Experiment 1: Biomarker performance for a 54-month clinical trial in an amyloid-positive, cognitively unimpaired cohort. Numbers in parentheses correspond to the 95% confidence interval. Sample size estimate (SSE) is assessed with 80% power to detect a 20% effect size with respect to placebo, with two visits separated by 54 months. In bold, the best result per metric. Composite = average PET SUV of whole cerebellum, brainstem and eroded subcortical white matter. Inferior cerebellum gm = inferior cerebellum grey matter. “ewm+cerebellum” indicates the average of the eroded subcortical white matter and the whole cerebellum. Our biomarkers have the prefix “CVR”, short for composite value ratio, with mSUV for mean SUV or SUV for volume-weighted SUV, and a suffix L or B indicating a lateral or bilateral regional analysis. Table reproduced from Llorente-Saguer and Oxtoby [131].

Biomarker	SSE	Separation	Repeatability
meta-temp/composite	626 (353, 1401)	4.37 (2.46, 6.49)	1.06 (0.82, 1.31)
DDS/ inferior cerebellum gm (mSUV)	438 (259, 895)	3.48 (1.57, 5.7)	1.66 (1.32, 2.06)
mesial temporal/ (ewm+cerebellum)	206 (143, 325)	0.17 (-1.84, 2.09)	1.03 (0.82, 1.3)
CVR-SUV-L	195 (131, 321)	4.2 (2.24, 6.25)	1.52 (1.19, 1.68)
CVR-mSUV-L	80 (61, 108)	4.32 (2.31, 6.37)	0.78 (0.67, 0.88)
CVR-SUV-B	205 (138, 338)	4.43 (2.61, 6.56)	1.17 (0.91, 1.47)
CVR-mSUV-B	115 (84, 168)	3.3 (1.33, 5.43)	0.86 (0.67, 0.96)

algorithm. Overall, Table 4.2 shows that one of the CVR biomarkers always outperforms state-of-the-art biomarkers from the literature — for all metrics and, in particular, our primary goal of smaller SSE. The best results for SSE were obtained with CVR-mSUV-L which is a ratio of simple-mean SUV combinations and laterality (separate hemispheric regions). The numerator included grey-matter regions from both hemispheres (superior frontal gyrus, amygdala), from the left hemisphere (caudal middle frontal gyrus, isthmus of cingulate gyrus, lingual gyrus, paracentral lobule, parahippocampal gyrus, postcentral gyrus, superior temporal gyrus), from the right hemisphere (lateral orbitofrontal gyrus). The denominator included grey- and white-matter regions from both hemispheres (precentral gyrus, brainstem, corpus callosum anterior, corpus callosum central, corpus callosum posterior), from the left hemisphere (insula, superior parietal lobule, ventral diencephalon), from the right hemisphere (lingual gyrus, paracentral lobule, postcentral gyrus, rostral anterior cingulate cortex).

The second-best SSE in Experiment 1 (about +45% SSE —worse) was obtained by CVR-mSUV-B, which combines bilateral mean SUV. The numerator included

grey-matter regions from both hemispheres (entorhinal cortex, amygdala, inferior temporal gyrus, insula, medial orbitofrontal cortex, paracentral lobule, precuneus, superior frontal gyrus, temporal pole, transverse temporal gyrus). The denominator included grey- and white-matter regions from both hemispheres (brainstem, eroded subcortical white matter, corpus callosum anterior, corpus callosum mid-anterior, caudal anterior cingulate cortex, postcentral gyrus, rostral anterior cingulate cortex, ventral diencephalon). To aid with visualisation, we show a representation of the brain[147], with the regions in the numerator in blue, and the ones in the denominator in red, in Supplementary Figure A.3.

Overall, performance of baselines in Experiment 1 went largely as expected for this preclinical trial scenario. For example, SUVRs with target/numerator regions involving later Braak stages performed poorly. The next worse targets were Braak2 (expected because of off-target binding) and temporoparietal, with mostly SSE above 1 000. The best results of literature-inspired biomarkers in terms of SSE were obtained with the mesial temporal as a target (simple average of its regions' SUV), with an SSE of 206, when using a reference region consisting of the simple average of the eroded subcortical white matter and the cerebellum regions. The next best targets were Braak stages 1 and 3 (SSE of 253 and 256, respectively, using the average SUV), when using a reference region consisting of the average of the eroded subcortical white matter and inferior cerebellum grey matter.

Experiment 2 (hypothetical 18-month clinical AD trial).

Figure 4.3 is a plot of sample size estimate (SSE) against percentage error (longitudinal precision) for all biomarkers in Experiment 2, truncated at $SSE \leq 4\,000$. CVR biomarkers (red crosses) showed superior performance to baseline comparators (blue dots). Error bars show 95% confidence intervals from bootstrapping (see Statistical analysis).

Table 4.3 highlights selected biomarkers for Experiment 2 (all results are included in Supplementary Material File Results.xlsx as a spreadsheet). Namely, specific literature biomarkers, plus the best-performing baseline (biomarker inspired by the literature) and our composite value ratio (CVR) biomarkers found by the algorithm. Overall, Table 4.3 shows that, like with Experiment 1 above, one of the CVR biomarkers always outperforms state-of-the-art biomarkers from the literature

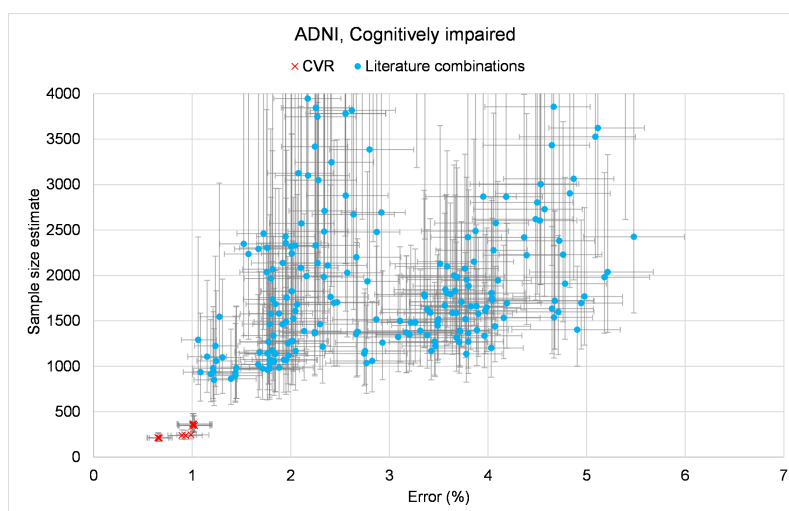


Figure 4.3: Biomarker performance for Experiment 2. The plot shows sample size estimate (SSE) against percentage error (longitudinal precision) for all biomarkers in Experiment 2, truncated at $SSE \leq 4000$. Composite value ratio (CVR) biomarkers (red crosses) showed superior performance to baseline comparators (blue dots), reducing both the SSE and the error. Error bars show 95% confidence intervals from bootstrapping. The SSE shown is the minimum number of individuals that would be needed per branch for a hypothetical clinical trial (cognitively impaired individuals), designed for 80% power and 20% effect size, for a duration of 1.5 years. Figure reproduced from Llorente-Saguer and Oxtoby [131].

—for all metrics and, in particular, our primary goal of smaller SSE. The best results were obtained with CVR-mSUV-L, which —like the results of Experiment 1— is a ratio of simple-mean SUV combinations with laterality. The numerator included grey-matter regions from both hemispheres (inferior temporal gyrus, insula, medial orbitofrontal cortex, superior frontal gyrus, temporal pole), from the left hemisphere (lingual gyrus, pars opercularis, superior parietal lobule), from the right hemisphere (fusiform, pars orbitalis, pars triangularis, pericalcarine cortex, precuneus, rostral middle frontal gyrus, superior temporal gyrus). The denominator included white- and grey-matter regions from both hemispheres (caudal anterior cingulate cortex, rostral anterior cingulate cortex, brainstem, whole cerebellum), from the left hemisphere (lateral orbitofrontal gyrus, pars orbitalis), from the right hemisphere (lateral occipital sulcus, transverse temporal gyrus).

The second-best SSE in Experiment 2 (about 12% bigger SSE —worse) was obtained by CVR-mSUV-B, as with Experiment 1, which involves bilateral mean SUV. The numerator included the following grey-matter regions from both hemispheres (inferior temporal gyrus, insula, medial orbitofrontal cortex, precentral gyrus, superior

Table 4.3: Results of Experiment 2: Biomarker performance for an 18-month clinical trial in an amyloid-positive, cognitively impaired cohort. Numbers in parentheses correspond to the 95% confidence interval. Sample size estimate (SSE) is assessed with 80% power to detect a 20% effect size with respect to placebo, with two visits separated by 18 months. In bold, the best result per metric. Composite = average PET SUV of whole cerebellum, brainstem and eroded subcortical white matter. Inferior_cerebellum_gm = inferior cerebellum grey matter. Our biomarkers have the prefix “CVR”, short for composite value ratio, with mSUV for mean SUV or SUV for volume-weighted SUV, and a suffix L or B indicating a lateral or bilateral regional analysis. Table reproduced from Llorente-Saguer and Oxtoby [131].

Biomarker	SSE	Separation	Repeatability
meta-temp/composite	1539 (1091, 2333)	4.37 (2.46, 6.49)	4.67 (3.97, 5.06)
DDS/inferior_cerebellum_gm	1031 (739, 1539)	3.48 (1.57, 5.7)	3.12 (2.63, 3.69)
SUV Braak5/(ewm+cerebellum)	852 (569, 1415)	4.14 (2.3, 6.35)	1.22 (1.02, 1.46)
CVR-SUV-L	348 (275, 454)	5.09 (3.17, 7.3)	1.02 (0.86, 1.19)
CVR-mSUV-L	212 (174, 265)	6.3 (4.38, 8.42)	0.66 (0.55, 0.76)
CVR-SUV-B	365 (288, 479)	5.15 (3.13, 7.32)	1.01 (0.85, 1.19)
CVR-mSUV-B	238 (195, 297)	5.6 (3.58, 7.6)	0.93 (0.79, 1.1)

frontal gyrus, temporal pole). The denominator included the following white- and grey-matter regions from both hemispheres (brainstem, whole cerebellum, caudal anterior cingulate cortex, lateral orbitofrontal gyrus, rostral anterior cingulate cortex). To aid with visualisation, we show a representation of the brain[147], with the regions in the numerator in blue, and the ones in the denominator in red, in Supplementary Figure A.3.

In this experiment as well, performance of baselines went largely as expected for this clinical trial scenario. The earlier Braak stages (1-3) and the subset of their regions that compose the mesial temporal had the worst performance. The best results of literature-inspired biomarkers in terms of SSE were obtained with the Braak5 composite as a target (volume-weighted SUV), with an SSE of 852, when using a reference region consisting of the average of the eroded subcortical white matter and the cerebellum regions. Following this, the next best SSE was 862, achieved by the same Braak5 target regions (mean SUV) and the composite reference region. The next best target was the temporoparietal, using the composite reference region, with an SSE of 962.

Experiment 3 (hypothetical 18-month clinical AD trial): validation.

Table 4.4 shows the salient results from the discovery set, and their performance in the validation dataset. As mentioned in 4.2.2, we had to use a subset of intersect-

Table 4.4: Results of Experiment 3: Biomarker performance for an 18-month clinical trial in an amyloid-positive, cognitively impaired cohort. Numbers in parentheses correspond to the 95% confidence interval. Sample size estimate (SSE) is assessed with 80% power to detect a 20% effect size with respect to placebo, with two visits separated by 18 months. In bold, the best result per metric. Composite = average PET SUV of whole cerebellum, brainstem and eroded subcortical white matter. Inferior_cerebellum_gm = inferior cerebellum grey matter. Our biomarkers have the prefix “CVR”, short for composite value ratio, with mSUV for mean SUV or SUV for volume-weighted SUV, and a suffix L or B indicating a lateral or bilateral regional analysis. Table reproduced from Llorente-Saguer and Oxtoby [131].

A) Discovery dataset			
Biomarker	SSE	Separation	Repeatability
meta-temp/whole_cerebellum	1722 (1190, 2710)	4.17 (2.28, 6.16)	4.68 (4.01, 5.07)
DDS/inferior_cerebellum_gm	1031 (739, 1539)	3.48 (1.57, 5.7)	3.12 (2.63, 3.69)
mSUV temporoparietal/cerebellum	1136 (719, 2056)	3.52 (1.71, 5.53)	1.84 (1.53, 2.17)
CVR-SUV-L	473 (362, 644)	6.09 (4.17, 8.34)	1.17 (0.98, 1.38)
CVR-mSUV-L	326 (257, 427)	6.38 (4.43, 8.48)	0.6 (0.51, 0.7)
CVR-SUV-B	513 (384, 723)	5.43 (3.57, 7.48)	1.31 (1.1, 1.54)
CVR-mSUV-B	362 (284, 478)	6.36 (4.47, 8.57)	0.85 (0.71, 1.01)
B) Validation dataset			
Biomarker	SSE		Repeatability
meta-temp/whole_cerebellum	690 (394, 1509)		2.16 (1.6, 2.84)
DDS/inferior_cerebellum_gm	417 (271, 723)		2.08 (1.52, 2.77)
mSUV temporoparietal/cerebellum	755 (419, 1748)		2.20 (1.68, 2.87)
CVR-SUV-L	601 (351, 1261)		0.97 (0.72, 1.26)
CVR-mSUV-L	513 (307, 1027)		0.63 (0.47, 0.82)
CVR-SUV-B	630 (361, 1369)		1.17 (0.88, 1.52)
CVR-mSUV-B	334 (217, 579)		0.7 (0.53, 0.91)

ing regions of all datasets, hence the differences with Experiment 2. The best result in the validation set was obtained by the CVR-mSUV-B, which was the second-best BioDisCVR configuration both in Experiment 1 and 2. The CVR-mSUV-B numerator consisted of the following grey-matter regions from both hemispheres (inferior temporal gyrus, insula, medial orbitofrontal cortex, precentral gyrus, superior frontal gyrus, temporal pole), while the denominator consisted of the following white- and grey-matter regions (brainstem, whole cerebellum, caudal anterior cingulate cortex, lateral orbitofrontal gyrus, rostral anterior cingulate cortex). To aid with visualisation, we show a representation of the brain[147], with the regions in the numerator in blue, and the ones in the denominator in red, in Supplementary Figure A.3. The second-best result in the validation set was obtained by the adaptive DDS biomarker,

although it did not perform as well in the discovery dataset.

Additional experiments. Supplementary table A.1 shows BioDisCVR results when fixing either the numerator (meta-temporal) or the denominator (inferior cerebellum grey matter). It can be observed that allowing BioDisCVR to find a data-driven denominator has a considerable impact in improving both the SSE (by a factor of 3 and 4 for Experiment 1 and 2, respectively), while fixing the reference as the inferior cerebellum grey matter achieves comparable results with literature biomarkers. Full information on selected regions is given in Supplementary File Results.xlsx of the online version [131]. Regional ablation analysis (shown in Supplementary Figure A.1) shows remarkably consistent performance, suggesting that no single region dominates the CVR biomarkers, with the exception of the biomarker tested in Experiment 3 – validation set, where removing the inferior temporal gyrus worsens the performance considerably (from 362 to 854 SSE).

Interestingly, the error and SSE of CVR_mSUV-B designed for Experiment 1, but applied to Experiment 2, are 0.90 and 652, respectively. The same configuration (CVR_mSUV-B), but trained for Experiment 2, has an error and SSE of 0.85 and 389, respectively, in Experiment 1. Both are still outperforming the literature biomarkers.

4.4 Discussion

When applied to preclinical (secondary prevention) and clinical (treatment) trial scenarios on Alzheimer's disease, our data-driven framework BioDisCVR discovered uniquely novel ratio-based tau PET biomarkers that are vastly superior to previous work. As discussed below, this can result in remarkable benefits including time- and cost-savings when running clinical trials on Alzheimer's disease.

In a 54-month preclinical trial (Experiment 1), the best BioDisCVR output (mSUV-L) reduced SSE by 82%, decreased repeatability error by 53% and increased group separation by 24% compared to the best result from the literature: the adaptive DDS biomarker[141]. Compared with the best combination of targets and references inspired by literature, the improvement is still vast, at 62% reduced SSE, 25% less error, and about 25 times higher group separation, as the benchmark performed poorly in group separation.

We found similar results in Experiment 2 (18-month clinical trial), where our best biomarker reduced SSE by 79%, decreased repeatability error by 79%, and

improved group separation by 80%, with respect to the DDS[141]. As the SSE is proportional to the inverse squared effect size, the CVR biomarkers will remain superior no matter which effect size is chosen.

In Experiment 3 (18-month clinical trial), results showed that the variant of CVR that had the best generalisation in the validation dataset was the mSUV-B (mean SUV, bilateral analysis), with a 20% reduction of SSE and 65% reduction of error with respect to the best biomarker from the literature (DDS[141]). A key difference between DDS and CVR is that in our configuration, CVR is a ratio of fixed regions, whereas DDS is adaptive, and different individuals will have different target regions. Notably, this CVR vastly outperformed DDS in the discovery set (68% smaller SSE, 80% smaller error).

The mSUV-B CVR for Experiment 1 and 2 had similarities in the numerator (entorhinal cortex, amygdala, precuneus, superior frontal gyrus, transverse temporal gyrus) and the denominator (brainstem, eroded subcortical white matter, caudal anterior cingulate, rostral anterior cingulate), but specialised in different regions to adapt to the stage of the disease: cortical regions that were stable to use in the denominator for earlier stages (e.g., cuneus) disappeared from the denominator and appeared in the numerator in Experiment 2, as they show changes later in the disease process[127]. We advise caution in drawing conclusions comparing the few CVR biomarkers among each other, though, since multiple configurations show similar performance (see ablation analysis in Supplementary Figure A.1).

Our framework demonstrates remarkable flexibility, which we exploited in multiple ways. For example, in supplementary experiments, we incorporated priors and constraints to enhance interpretability and bolster confidence in the main findings. Allowing for hemispheric laterality generally yielded CVR biomarkers with slightly improved SSE (5-30% smaller), which seems to concur with a previous finding of laterality of tau pathology accumulation in Alzheimer's disease[18]. In addition to the volume-weighted composite SUV that is used in previous SUVR studies on statistical ROIs[148, 124, 140, 125, 149], we considered a regional mean SUV, remarkably finding that the regional mean produced CVR biomarkers with better repeatability error (7-48% smaller) and SSE (35-60% smaller). The effect of composition was not significant when performing a Wilcoxon rank-sum test in the literature-inspired

biomarkers ($W = 19944$, $p\text{-value} = 0.7079$, filtering out biomarkers that had above 3 000 SSE). Braak stage SUVR biomarkers (using the composite reference region) produced results comparable to the best biomarkers from the literature, but still worse than CVR (see Supplementary File Results.xlsx, tab Exp1_and_2 of the online version [131]). As could be expected, the earliest Braak Stage 1 was favoured for Experiment 1 (preclinical trial), and the later Braak Stage 5 was favoured for Experiment 2 (clinical trial). Fixing the numerator as the meta-temporal region (and finding a data-driven denominator) improved results compared to using the meta-temp/composite biomarker[78]. Likewise, fixing the denominator as the inferior cerebellum grey matter yielded improved results compared to the literature biomarkers, providing further support for our approach of moving beyond the traditional reference regions used in SUVR analyses. Exploring alternative optimisers as a search algorithm (box-constrained BFGS or Nelder-Mead) did not achieve results on par with the ones obtained by the genetic algorithm (results not shown), but other optimisation approaches could be explored in the future. Different random initialisations generated similarly performing biomarkers, with a low coefficient of variation (0.043) in fitness for Experiment 2 (clinical), but seemed to stop at a local minimum for one of the random seeds in Experiment 1 (preclinical), resulting in a coefficient of variation of 0.17. As observed in the regional ablation analysis (Supplementary Figure A.1), where different biomarkers exhibited similar performances, at the final computed generation of the genetic algorithm, the top 5 biomarkers exhibited comparable performance metrics, with a fitness coefficient of variation below 10% for Experiment 1 (preclinical) and below 5% for Experiment 2 (clinical), (Supplementary File Results.xlsx of the online version [131]). The framework's adaptability extends beyond the two experimental scenarios (54- and 18-month clinical trials) considered in our experiments: it could have easily been employed to identify a biomarker that performs well in both scenarios, thereby providing a unified solution.

Our experiments produced biomarker ratios that included known regions commonly associated with Alzheimer's disease. The numerator often included regions commonly used as targets in SUVR studies of Alzheimer's disease progression, e.g., entorhinal cortex, fusiform gyrus, inferior temporal gyrus, middle temporal gyrus. The denominator often included regions used as SUVR references, e.g., eroded

subcortical white matter, whole cerebellum or cerebellum cortex, brainstem. The fact that our CVR (mean SUV, bilateral) designed for a trial works on the other (better than the specific two biomarkers from the literature) validates, in unseen data, that they are targeting disease-specific regions. Furthermore, both the 95% confidence intervals and the ablation study on influential regions (Supplementary Figure A.1) suggest a robust fit in the available data, free of questionable influential regions. Additional metrics or priors to guide the algorithm could be used to select a smaller set of good-performing biomarkers. Results indicate that using a curated biomarker per clinical trial target group can better characterise relevant disease progression, in line with the recent Revised Criteria for Diagnosis and Staging of Alzheimer's Disease[4], reflecting that different regions are affected at different stages of the disease.

The fast execution of the framework BioDisCVR, together with the usage of readily processed scans (thus avoiding the need to re-run preprocessing methods) makes this approach computationally cheap, positively addressing an increasingly concerning environmental impact of machine learning[150].

We anticipate that data-driven methods such as BioDisCVR could revolutionise the drug development pipeline. The dramatic reduction in SSE produced here has multiple benefits. First, a smaller trial reduces the number of individuals exposed to increased risk of side effects such as ARIA[151, 152, 120, 153]. Second, trials can take considerably less time, e.g., the A4 Study screening time could have been reduced from 44 months[154] to under 12 months, considering a linear relationship between recruitment time and sample size[155]. Third, the cost savings could be immense, e.g., for a cost per individual of \$40k[156], the possible savings per trial using CVR-mSUV-B (as opposed to the best published biomarker) could be $(2 \text{ arms} \times (1031 - 238) \times 40\,000)$ about \$65 million for CI/18-months, or $(2 \text{ arms} \times (438 - 115) \times 40\,000)$ about \$25 million for CU/54-months. In the event that other factors influence the minimum sample size needed and cannot be reduced, using CVR would allow a more precise detection threshold of disease progression. The superior longitudinal precision of our data-driven biomarkers makes them candidates for secondary outcomes or endpoints.

We highlight opportunities for future work. First, it is important to explore

further validation of tau PET ratio-based biomarkers across cohorts, ethnicities, etc. (work in progress, with a collaboration with AIBL and BioFinder groups). Second, our framework can be applied across other PET tracers and beyond PET, e.g., to discover new and improved ratio-based blood biomarkers[157, 4]. Indeed, the small run-time needed for computation facilitates extensive exploration in any data modality and even across modalities, including fluid biomarkers and MRI (less expensive than PET and non-invasive). Not only this, but data acquisition cost could be included in the fitness function to discover a cost-effective ratio-based multimodal biomarker—potentially benefiting low- and middle-income countries, in particular. Third, future work could consider deriving a theoretical guarantee that a given search algorithm will find the global optimum biomarker, but in practice, we found that multiple runs with random seeds produced very similar results (see ablation analysis in Supplementary Figure A.1 and top-n biomarkers in Supplementary File Results.xlsx of the online version [131]).

4.5 Conclusions

Clinical trials of anti-amyloid and anti-tau therapies in Alzheimer’s disease need a precise and accurate biomarker for assessing disease modification (pathology clearance). We have introduced and deployed a data-driven framework for discovering such biomarkers, with experimental results showing state-of-the-art longitudinal precision, and drastically reduced sample size estimates required for modern clinical trials in symptomatic and preclinical Alzheimer’s disease. Our results suggest that considerable time, money, and participant suffering could be saved by incorporating our biomarker framework into the design of future preclinical and clinical trials.

Chapter 5

CVR applied to MRI to track volumetric changes

A preliminary version of the work presented in this chapter was presented in 2023 at the Alzheimer's Association International Conference. Later, a manuscript was prepared with a more thorough analysis, and it is currently under editorial review by the Brain Communications journal (as of February 2025). Neil Oxtoby contributed to the manuscript.

5.1 Introduction

As noted in Background Chapter 2, image-based biomarkers are used as secondary endpoints in clinical trials, often to confirm biological mechanisms to support clinical benefit. Measurement variance can diminish our ability to detect true changes. In Chapter 4, we introduced a biomarker discovery algorithm, applied to tau PET regional values, which produced a new biomarker: the composite value ratio (CVR) [158, 131]. Here, we apply the CVR concept to more widely available and cheaper MRI data, potentially enhancing the practicality and accessibility of longitudinal monitoring. MRI is more relevant to routine healthcare than the specialised tau PET scans, although their sensitivity at different stages of the disease is very different: MRI changes are downstream of pathology accumulation.

Brain atrophy can result naturally from ageing [159, 31, 32, 160, 161] or be accelerated by neurodegenerative diseases (e.g., Alzheimer's disease [32, 162, 30, 163, 4], frontotemporal dementia [30]). Understanding volumetric changes can be useful for diagnosis [163, 4], monitoring disease progression [164], and evaluating the impact of therapeutic interventions [117, 118]. The following regions are reported

to experience a change over time in healthy adults that can be measured with T1- or T2-weighted magnetic resonance imaging (MRI): whole brain [159, 30, 161], hippocampus [31, 165, 161, 163], entorhinal cortex [165], ventricles [159, 32, 160, 161], temporal lobe [31, 161]. In Alzheimer’s disease, these are the regions of interest that have been reported to have an accelerated and detectable volume change over time with respect to healthy controls: whole brain [30], hippocampus [31], ventricles [32], medial temporal lobe [31, 4]. In phase-3 randomised clinical trials aimed at intervening on Alzheimer’s disease progression, MRI-based volumetric measures of the whole brain, the ventricles and the hippocampus have been used as primary [117] and secondary outcome measures [118, 57, 75, 119]. Moreover, visual ratings such as *medial temporal lobe atrophy* (MTA) or *global cortical atrophy* (GCA) are used to subjectively assess atrophy severity. If we discover a CVR that improves measurement sensitivity over traditional biomarkers, then it could provide an objective, quantitative measure to help radiologists and clinicians.

5.2 Materials and Methods

We apply the BioDisCVR framework to find a composite value ratio (CVR) aimed at improving the monitoring of brain volume changes due to Alzheimer’s disease. CVR and traditional biomarkers are measured using three main metrics: (1) the sample size needed to detect a 20% change in the gradient in a hypothetical clinical trial; (2) a measure of longitudinal coherence (translated into a percentage error of the biomarker measure); and (3) group separation of the biomarker trajectory (according to amyloid positivity and cognitive status, as two separate analysis). The BioDisCVR framework uses by default a genetic algorithm to search for combinations of features (a ratio of aggregate brain regional volumes in this case), that improve a fitness function (described below, in Model).

Data. We analysed longitudinal T1-weighted MRI regional volumes from 1381 participants in the ADNI dataset. The segmentation to obtain regional volumes was performed with FreeSurfer 7.1.1 [137]. Table 5.1 shows the number of subjects and visit counts by cognitive impairment (DX), amyloid positivity (AB), sex, and presence of APOE4 (APOE). The median total follow-up time per individual (earliest visit to latest visit) was 157.6 weeks, with an interquartile range (IQR) of 210.3 weeks (range: 4 to 809.6 weeks, average: 214.5 weeks, total follow-up time (adding all

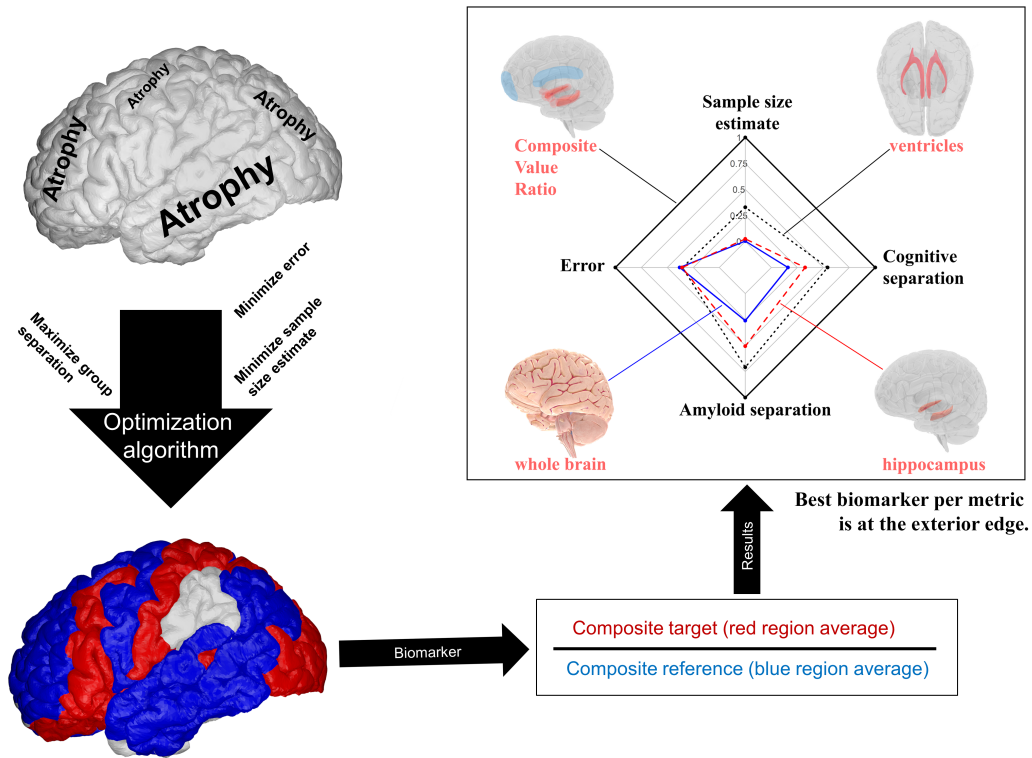


Figure 5.1: Conceptual abstract. Atrophy happens distinctly in different parts of the brain. Regional volumes are fed to an algorithm that searches for a composite value ratio (CVR) that minimises the sample size estimate for a hypothetical clinical trial, maximises group separation (amyloid-positive versus amyloid-negative), and reduces longitudinal error (linear model residuals to fitting the log-transformed biomarker). In the radar plot, the best biomarker per metric is at the external edge.

individuals) over 5676 years). Additional validation is done with the recently available A4 Study data [132].

Study design. We randomly split the data into two subsets, stratified by the covariates in the model formula below (cognitive status, amyloid status, sex, APOE4 presence). The distributions of the two splits are shown in Table 1, and are balanced for all covariate combinations. We report the out-of-sample evaluation metrics of each of the two subsets. In the Statistical analysis section below, we introduce the modelling formula and the evaluation metrics. We consider two cases, or target groups: preclinical (cognitively unimpaired, amyloid-positive) and clinical (cognitively impaired, amyloid-positive). We define amyloid-positive individuals as those being amyloid-positive at any visit (composite amyloid-PET SUVR above or equal to 0.78 for Florbetapir and 0.74 for Florbetaben)[133].

Statistical analysis. We fit a linear mixed-effects model with correlated random

Table 5.1: Data characteristics. Here we can see the available number of scans and individuals per each combination of covariates. DX = diagnosis, where CU = cognitively unimpaired, and CI = cognitively impaired. AB indicates amyloid positivity (shown with the symbol +). APOE4 indicates whether the individuals have no such gene mutations (–) or otherwise (+). The data is randomly split in two subsets for cross-validation of discovered biomarkers.

Covariates				Total		Training set		Test set	
DX	AB	sex	APOEbin	scans	#ID	scans	#ID	scans	#ID
CU	-	female	none	537	106	258	53	279	53
CU	-	female	>0	123	23	75	12	48	11
CU	-	male	none	565	126	276	63	289	63
CU	-	male	>0	147	37	73	19	74	18
CU	+	female	none	161	28	84	14	77	14
CU	+	female	>0	77	17	41	9	36	8
CU	+	male	none	192	39	98	20	94	19
CU	+	male	>0	201	52	88	26	113	26
CI	-	female	none	941	166	506	83	435	83
CI	-	female	>0	249	47	117	24	132	23
CI	-	male	none	655	113	327	57	328	56
CI	-	male	>0	121	23	68	12	53	11
CI	+	female	none	564	108	298	54	266	54
CI	+	female	>0	1183	231	596	116	587	115
CI	+	male	none	494	95	248	48	246	47
CI	+	male	>0	857	170	415	85	442	85

intercepts and slopes, similar to the framework work on tau PET biomarkers for AD (Chapter 4, and [131]). Common covariates (age, sex, APOE4, education) are analysed and included in the model where significant (iterative ANOVA tests), using the ventricles divided by the total intracranial volume as a baseline biomarker. The ventricles are used because they show the most stable longitudinal change (smaller model residuals), compared to other biomarkers from the literature (entorhinal cortex, hippocampus or whole brain), as shown in earlier work [160, 166]. Binary variables for amyloid positivity and cognitive impairment are also included, as their fixed effects will help us evaluate disease signal of biomarkers (positive individuals in either are expected to have more pronounced brain changes than otherwise). The formula used is the following:

$$\log(\text{biomarker}) = \beta_0 + b_0 + \text{Age.bl}(\beta_1 + \beta_2 \cdot \text{sex}) \\ + \text{time}(\beta_3 + b_1 + \beta_4 \cdot \text{DX} + \beta_5 \cdot \text{AB} + \beta_6 \cdot \text{APOE}) + \varepsilon \quad (5.1)$$

Where the age at baseline visit and sex (female = 1, male = -1) contribute to the intercept, and the gradient (base β_3) is modified by cognitive status (DX = 0 for cognitively unimpaired, 1 otherwise), amyloid positivity (AB = 0 when negative, 1 otherwise) and having at least one APOE4 allele (APOE4 = 0 if none, 1 otherwise). We include correlated random effects for the intercept (b0) and gradient (b1), since this provides invariance to time-shifts [144], and we use a relative time point (the average of visit dates, per individual). Education was not included, as it did not have a significant impact on the fitting of the model ($\chi^2 = 0.83$, DF = 1, $p = 0.36$). The low diversity in ethnicity precluded its inclusion in the model: out of the 1381 individuals in the study, only 6 were labelled as Hispanic, 1323 as not Hispanic, and 52 as unknown. When considering a specific target group for a clinical trial (only amyloid-positive, and either cognitively unimpaired or impaired), we use the same equation, but AB and DX are removed because these variables become fixed.

We calculate multiple evaluation metrics to compare biomarkers, namely sample size estimate, measurement percentage error (as per our model fitting), and group separation (by amyloid-positivity and cognition), described below.

Sample size estimate (SSE) is the sample size estimate per branch (placebo and treatment) for a hypothetical clinical trial designed for 80% power, with 20% treatment effect size, with measurements at the initial and final visit of the trial. Length is 18 months for the clinical trial (cognitively impaired, amyloid-positive, similar to the trials ENGAGE/EMERGE [57], CLARITY-AD [79], TRAILBLAZER-ALZ 2 [75]) and 54 months for the preclinical group (cognitively unimpaired, amyloid-positive, similar to the A4 trial [132]). Given the same trial design conditions, a lower SSE indicates that we can achieve the same detection of effect size with a smaller number of participants. Even if a trial is not based solely on this, we can understand this metric as a sensitivity metric, quantifying how good the biomarker is, accounting for the gradient and variance of the model coefficients and residuals.

As an error measure to quantify the residual variability of the linear mixed-effects model, we define $error = 100\%(e^{sd(residuals)} - 1)$, where sd is the standard deviation, and $residuals$ are the model residuals. This transformation back to the original scale allows for a more intuitive interpretation of variability in the context of the biomarkers' actual values. Expressing residual variability as a percentage of the biomarker in

its native space facilitates direct comparisons and meaningful insights into model performance, where a lower error means a better fit.

We further provide two group-separation metrics, corresponding to the t-statistic of the fixed effect of being amyloid-positive ($AB = 1$), and also the t-statistic of the fixed effect of being cognitively impaired ($DX = 1$). This is key to ensuring the biomarker is capturing changes related to the disease. A larger t-statistic corresponds to better group separation. The above metrics are provided with 95% confidence intervals, calculated using the model-based (semi-)parametric bootstrap for mixed models function from the lme4 package [144], version 1.1.35.1.

Additionally, we perform the following tests to assess the fitting of the model: variance inflation factors, as a multicollinearity check, since if strong multicollinearity is present, it will affect the stability and interpretability of the regression coefficients, potentially leading to inflated standard errors and unreliable estimates. Additionally, we do a visual analysis of model residuals and quantile-quantile plots to ensure the model's assumptions about the residuals are not violated.

Further, we examine the usage of volumetric measures in past Phase 3 clinical trials (the A4 Study [132], EMERGE and ENGAGE [57], Clarity AD [79], TRAILBLAZER-ALZ 2 [75]), and calculate the expected detectable effect size for the same clinical trial configuration (number of participants, trial duration and inclusion criteria), for the different biomarkers.

Finally, as an additional CVR validation and application, we obtained the publicly available A4 trial MRI scans (recently available) and processed them using our established pipeline. We compare the annualised percent change for each biomarker for subjects with valid follow-up data. This was computed as $[(\text{endpoint_value} - \text{baseline_value}) / \text{baseline_value}] / \text{duration_in_years} * 100\%$. To estimate the treatment effect, we fitted an Analysis of Covariance (ANCOVA) model for each biomarker. The annualised percent change was the response variable, with trial group (Solanezumab vs. placebo) as the primary predictor. All models were adjusted for key baseline covariates: age, ApoE $\epsilon 4$ carrier status (presence vs. absence), years of education (dichotomised as ≥ 13 vs. < 13 years), and the baseline value of the respective biomarker to increase statistical power and control for regression to the mean. From these models, we extracted the least-squares mean (LS-mean)

difference between treatment arms, its 95

Model. We run the BioDisCVR framework (explained in Chapter 4, section 4.2.3) to obtain a composite value ratio (CVR): the ratio of two data-driven composite regions. Each composite region is the average signal over a number of regions. A genetic algorithm explores combinations of 34 bilateral regions defined by the Desikan-Killiany atlas in FreeSurfer 7.1.1 [137]. Similar to the original framework paper, our algorithm's fitness function to be maximised includes group separation (for disease signal) and a measure of variability (sample size estimate, described below), and is defined as $t_{AB}/(SSE^2)$, where:

t_{AB} = Amyloid group separation: t-statistic of the fixed effect of being amyloid-positive, fitting the model to all data. SSE = Sample size estimates per group for a hypothetical clinical trial designed for 80% power, with 20% treatment effect size. Measurements at the initial and final visit of the trial. Length is 18 months for the clinical trial (cognitively impaired, amyloid-positive) and 54 months for the preclinical group (cognitively unimpaired, amyloid-positive). For the analysis where we fit all groups (regardless of cognitive or amyloid status), we consider a hypothetical clinical trial length of 18 months; this is the experimental configuration used to report both the amyloid and cognitive group separation.

Baselines. Results are compared to volumetric biomarkers that have been used as outcome measures [117, 119, 57, 75] in phase-3 randomised clinical trials aimed at intervening on Alzheimer's disease progression: whole brain, hippocampus, ventricles. We additionally include the entorhinal cortex, for its strong implication with Alzheimer's disease [165]. These volumes of interest are all divided by the concurrent intracranial volume (ICV).

5.3 Results

Results are presented in the following structure: evaluation metrics, biomarker regions, model fitting, and examination of past clinical trials.

Sample size estimate. Table 5.2 shows the sample size estimates for the compared biomarkers, for different groups in each column: all the data (SSE all), only for cognitively impaired amyloid-positive individuals (SSE CI), and only considering cognitively unimpaired amyloid-positive individuals (SSE CU). In parentheses, 95% confidence intervals by bootstrapping. We can observe that the CVR biomarkers

outperform the conventional regions of interest by a considerable margin in all metrics. In particular, CVR requires a 67-73% smaller sample for a clinical trial (cognitively impaired), and a 20-34% smaller sample for a preclinical trial (cognitively unimpaired), compared with the best literature biomarker (ventricles). We also notice the poor performance of the whole brain, entorhinal and hippocampus. Considering all data, CVR biomarkers also show a much lower SSE (in all CVR cases, 45-66% lower) than the best literature biomarker (ventricles).

Table 5.2: Sample size estimate for a hypothetical clinical trial (80% power, 20% effect size). The sample size estimate (SSE) is found by fitting a linear mixed effects model to the target group: all (for reference), cognitively impaired (CI) or cognitively unimpaired (CU). Evaluation is done in each of two splits of the data (Subset 1 and 2), using the other subset for disCVRy (finding CVR biomarkers). CVR_CI was found by minimising the sample size estimate for the cognitively impaired (CI), while CVR_CU was focused on the cognitively unimpaired group (CU). In bold, the best result per subset. Hippocampus, Ventricles, Entorhinal Cortex (EC), and Whole brain are each divided by Intracranial Volume (ICV).

A) Data Subset 1

Biomarker	SSE all	SSE CI	SSE CU
CVR_CI	776 (598, 1048)	342 (264, 462)	97 (70, 142)
CVR_CU	678 (521, 918)	317 (241, 435)	65 (50, 89)
Whole brain	114441 (735869, 44390)	36365 (169796, 15389)	1340 (4011, 663)
Hippocampus	45834 (181106, 20455)	11598 (29443, 6158)	1666 (7696, 707)
EC	26430 (70219, 13749)	6428 (12865, 3844)	859 (1979, 478)
Ventricles	2012 (1497, 2845)	1290 (897, 2013)	99 (73, 143)

B) Data Subset 2

CVR_CI	890 (662, 1260)	302 (227, 424)	80 (60, 112)
CVR_CU	1154 (868, 1608)	395 (288, 577)	81 (60, 113)
Whole brain	86902 (416065, 36502)	66653 (347682, 27313)	1284 (2953, 714)
Hippocampus	50479 (221457, 21775)	12077 (29539, 6524)	1043 (3073, 520)
EC	20813 (58604, 10558)	5222 (11642, 2951)	1545 (7285, 652)
Ventricles	2088 (1586, 2872)	918 (682, 1304)	101 (74, 145)

Percentage error. Next, we examine the percentage error in Table 3. As in Table 2, Table 3 shows results for different groups: all the data (all), only for cognitively impaired amyloid-positive individuals (CI), and only considering cognitively unimpaired amyloid-positive individuals (CU). In parentheses, 95% confidence intervals by bootstrapping. These results show that the CVR biomarkers vastly outperform the conventional regions of interest in all the analyses. In particular, in the evaluation with unseen data, CVR has a 47-66% smaller error in a clinical trial setting (cognitively impaired), and a 6-39% smaller error for a preclinical trial

(cognitively unimpaired), compared with the best literature biomarker at each subset (whole brain). We also notice the large variance of the entorhinal cortex. When evaluating all data, CVR biomarkers also outperform traditional regions of interest (23-67% smaller error).

Table 5.3: Percentage error. Evaluation of the percentage error for the different biomarkers, in both splits of the data (Subset 1 and 2). Our biomarkers (CVR) were found using the other split they are evaluated in. CVR_CI was found by minimising the sample size estimate for the cognitively impaired (CI), while CVR_CU was focused on the cognitively unimpaired group (CU). In bold, the best result per subset. Hippocampus, Ventricles, Entorhinal Cortex (EC), and Whole brain are each divided by Intracranial Volume (ICV).

A) Data Subset 1

Biomarker	Error all	Error CI	Error CU
CVR_CI	2.15 (2.09, 2.22)	2.36 (2.25, 2.47)	2.03 (1.83, 2.25)
CVR_CU	2.52 (2.44, 2.60)	2.69 (2.57, 2.81)	2.37 (2.13, 2.63)
Whole brain	6.50 (6.31, 6.72)	7.98 (7.59, 8.37)	3.91 (3.50, 4.31)
Hippocampus	6.87 (6.65, 7.08)	8.45 (8.06, 8.86)	4.38 (3.90, 4.84)
Entorhinal cortex	9.42 (9.13, 9.73)	11.25 (10.73, 11.78)	7.33 (6.57, 8.16)
Ventricles	6.41 (6.21, 6.61)	7.72 (7.38, 8.06)	4.10 (3.67, 4.54)

B) Data Subset 2

CVR_CI	2.56 (2.48, 2.64)	2.54 (2.43, 2.66)	2.04 (1.82, 2.29)
CVR_CU	4.03 (3.90, 4.15)	3.56 (3.39, 3.73)	3.04 (2.71, 3.40)
Whole brain	5.26 (5.10, 5.41)	6.67 (6.36, 6.94)	3.24 (2.90, 3.52)
Hippocampus	5.70 (5.53, 5.87)	7.10 (6.78, 7.41)	3.96 (3.53, 4.41)
Entorhinal cortex	8.45 (8.17, 8.73)	9.65 (9.19, 10.09)	6.76 (6.05, 7.56)
Ventricles	6.59 (6.39, 6.80)	7.16 (6.82, 7.49)	4.01 (3.60, 4.45)

Group separation. Next, in Table 5.4 we show the group separation: amyloid-positivity and cognitive status. As described in 5.2, we report the t-statistic of the fixed effect of being amyloid-positive, and also the t-statistic of the fixed effect of being cognitively impaired. In parenthesis, 95% confidence intervals by bootstrapping. Here as well, we can observe that the CVR biomarkers outperform the conventional regions of interest. Ventricles have the next best performance, followed by entorhinal cortex and hippocampus, with the whole brain having the worst results. In particular, in the evaluation with unseen data, CVR has a 38-66% higher separation regarding amyloid positivity, and a 23-85% higher separation regarding cognitive impairment, compared with the best literature biomarker at each subset (ventricles).

The detailed region list of the discovered biomarkers aimed at reducing the sample size estimate for a preclinical trial involving cognitively unimpaired, amyloid-

Table 5.4: Group separation, both in amyloid-positivity and cognitively. The separation is given by the t-statistic of the fixed effect of being amyloid-positive or cognitively impaired, fitting a linear mixed-effects model to a split of the data (Subset 1 and 2). Our biomarkers (CVR) were found using the other split they are evaluated in. CVR_CI was found by minimising the sample size estimate for the cognitively impaired (CI), while CVR_CU was focused on the cognitively unimpaired group (CU). The negative separation indicates the direction of the difference between groups. In bold, the best result per subset. Hippocampus, Ventricles, Entorhinal Cortex (EC), and Whole brain are each divided by Intracranial Volume (ICV).

A) Data Subset 1

Biomarker	Separation (amyloid positivity)	Separation (cognition)
CVR_CI	6.90 (4.97, 8.90)	3.16 (1.20, 5.21)
CVR_CU	6.50 (4.54, 8.45)	3.57 (1.72, 5.49)
Whole brain	-2.05 (-4.10, -0.04)	-0.51 (-2.39, 1.51)
Hippocampus	-2.44 (-4.41, -0.36)	-1.31 (-3.30, 0.70)
Entorhinal cortex	-3.03 (-5.08, -1.19)	-1.88 (-3.97, 0.20)
Ventricles	4.15 (2.24, 6.17)	2.57 (0.40, 4.50)

B) Data Subset 2

CVR_CI	7.53 (5.55, 9.39)	4.77 (2.61, 6.79)
CVR_CU	7.35 (5.33, 9.46)	3.92 (2.09, 5.79)
Whole brain	-1.96 (-4.10, 0.02)	-0.76 (-2.63, 1.22)
Hippocampus	-3.81 (-5.81, -1.69)	-1.55 (-3.57, 0.37)
Entorhinal cortex	-3.89 (-5.76, -1.92)	-1.18 (-3.17, 0.83)
Ventricles	5.34 (3.30, 7.38)	2.58 (0.89, 4.67)

positive individuals is shown in Table 5.5. In the numerator, both biomarkers found in the two splits of the data had the ventricles, frontal pole, isthmus of cingulate gyrus, and the mid-posterior part of the corpus callosum. In the denominator, they shared the nucleus accumbens, amygdala, choroid plexus and hippocampus.

The detailed region list of the discovered biomarkers aimed at reducing the sample size estimate for a preclinical trial involving cognitively impaired, amyloid-positive individuals is shown in Table 5.6. Both biomarkers found in the two splits of the data included the ventricles, the central part of the corpus callosum and the choroid plexus in the numerator. In the denominator, they shared the nucleus accumbens, banks of the superior temporal sulcus and the anterior part of the corpus callosum.

In both clinical trial scenarios above, the numerator included the ventricles, and the denominator included the nucleus accumbens. The most notable differentiator between the trial scenarios is that cortical regions featured more in CI clinical trial, whereas subcortical regions were involved for CU preclinical trials.

Table 5.5: Regions of the biomarkers discovered for the cognitively unimpaired group.

Subset 1/2 indicates the subset of data that was used for the discovery of the biomarker. In black, regions that were selected in both subsets of the data.

Numerator		Denominator	
Subset 1	Subset 2	Subset 1	Subset 2
cc_mid_posterior	cc_mid_posterior	accumbens_area	accumbens_area
frontalpole	frontalpole	amygdala	amygdala
isthmuscingulate	isthmuscingulate	choroid_plexus	choroid_plexus
lateral_ventricle	lateral_ventricle	hippocampus	hippocampus
caudalanteriorcingulate	paracentral	cc_anterior	bankssts
cc_central		cc_mid_anterior	entorhinal
lingual		cc_posterior	inferiortemporal
pericalcarine		parahippocampal	thalamus
posteriorcingulate		transversetemporal	ventraldc

Table 5.6: Regions of the biomarkers discovered for the cognitively impaired group.

Subset 1/2 indicates the subset of data that was used for the discovery of the biomarker. In black, regions that were selected in both subsets of the data.

Numerator		Denominator	
Subset 1	Subset 2	Subset 1	Subset 2
cc_central	cc_central	accumbens_area	accumbens_area
choroid_plexus	choroid_plexus	bankssts	bankssts
lateral_ventricle	lateral_ventricle	cc_anterior	cc_anterior
caudate	pericalcarine	amygdala	entorhinal
transversetemporal		brainstem	parstriangularis
		cc_posterior	
		cerebellum_white_matter	
		hippocampus	
		lateralorbitofrontal	
		middletemporal	
		pericalcarine	
		rostralmiddlefrontal	
		ventraldc	

Figure 5.2 summarises all metrics for both subsets of the data for the two cognitive groups (unimpaired and impaired). The outer coordinates correspond to the best biomarker metric, which is used as the reference scale. In all cases and for all metrics, we observe a dominance of the CVR biomarkers. Of the traditional volumetric biomarkers, ventricles outperformed hippocampus and whole brain (entorhinal cortex not shown to avoid cluttering).

Figure 5.3 visualises biomarker variance as model-predicted versus observed $\log(\text{biomarker})$ values, using all data (both subsets). CVR shows a visually narrower spread compared to other biomarkers (in line with the error —over two times lower

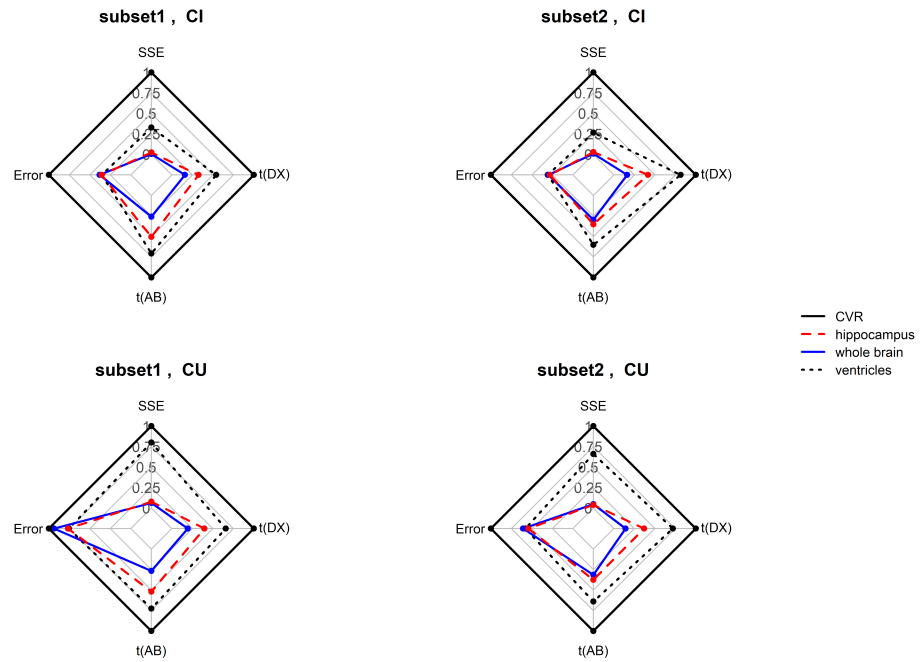


Figure 5.2: Summary metrics report. The outer coordinates correspond to the best biomarker metric, and the rest are scaled relative to that, with metrics that are best when minimum transformed by best/metric and metrics that are best when maximum by metric/best, to normalise all to a maximum of 1. SSE = sample size estimate for a hypothetical clinical trial (80% power, 20% effect size). Error = percentage error of the measurement, as defined by the standard deviation of the residuals of the model, as data is log-transformed. The metric t(AB) corresponds to the t-statistic of the fixed effect for the gradient of amyloid-positive individuals. Finally, the metric t(DX) corresponds to the t-statistic of the fixed effect for the gradient of cognitively impaired individuals.

standard deviation of model residuals— reported in Table 3), suggesting that it explains longitudinal data better. Of the traditional volumetric biomarkers, ventricles again outperform hippocampus and whole brain. We observe the same results in out-of-sample experiments on the two data subsets, shown in Supplementary Figures 1 and 2.

We next report results of our *post hoc* checks. For all the biomarkers, none of the variance inflation factors were above 2, suggesting we don't have a strong collinearity problem in our model [167]. For all biomarkers, the spread of the residuals with respect to the fitted values is a single cluster aligned at zero, showing

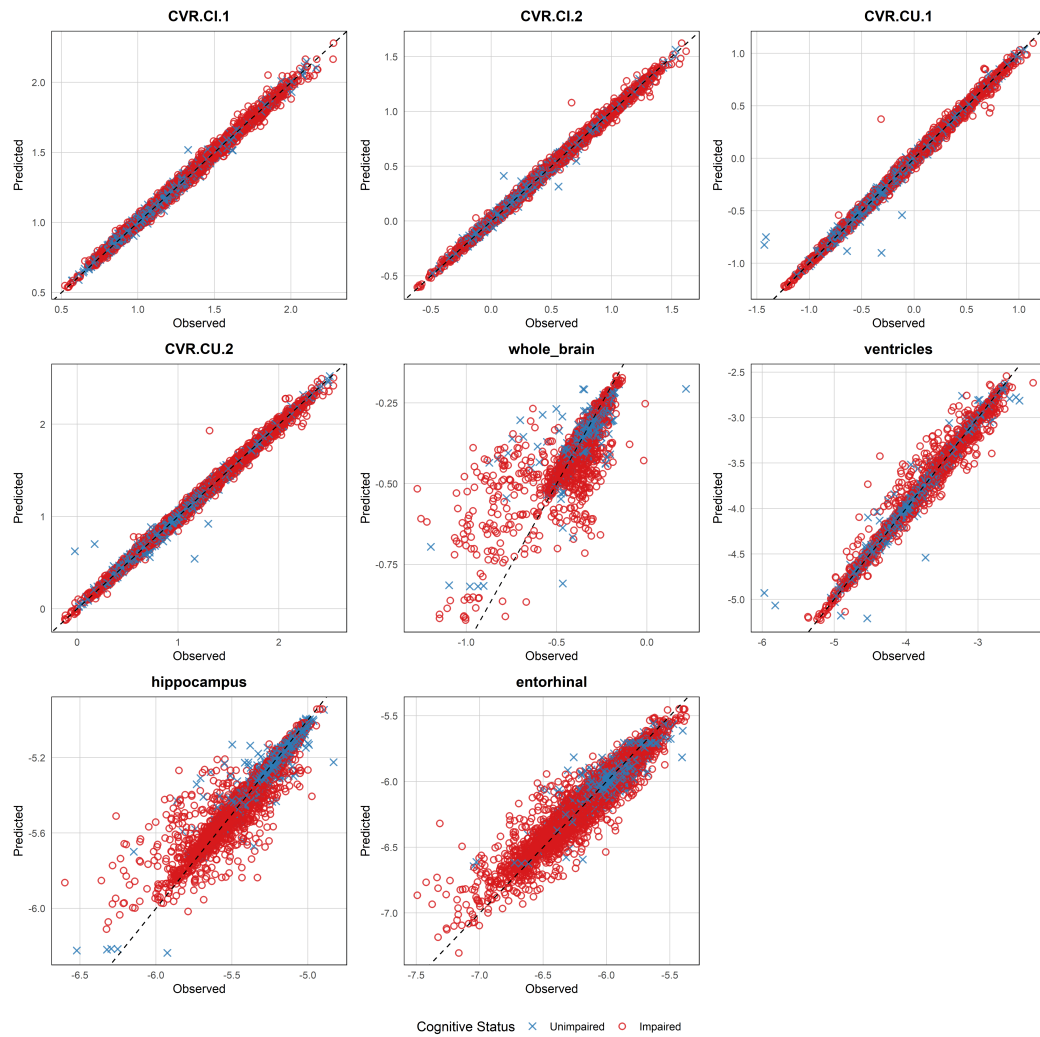


Figure 5.3: Predicted vs observed values. Values are log-transformed. CVR is the composite value ratio of two data-driven regions of interest; the rest are divided by the intracranial volume. The CVR naming coding is as follows: CI = conditioned to minimise the sample size estimate of cognitively impaired individuals; CU = conditioned to minimise the sample size estimate of cognitively unimpaired individuals; suffix 1 or 2 indicate the subset of the data that was used for the CVR discovery.

homogeneity of variance. In quantile-quantile plots, we observe that the central part of the distribution aligns with the expected quantiles, but the tails do not, which suggests the presence of extreme values.

Table 5.7 reports the detectable effect size using past clinical trials' inclusion criteria and configuration. We observe a considerable difference between biomarkers, and their performance from best to worst is: CVR, ventricles, entorhinal cortex, hippocampus, whole brain. To aid readability, we summarise results for the trials

targeting cognitively impaired individuals as CI (EMERGE and ENGAGE [57], Clarity AD [79], TRAILBLAZER-ALZ 2 [75]), and the one trial targeting cognitively unimpaired individuals as CU (A4 Study [132]). Here we summarise the evaluation of the data subsets, to ensure that our reported biomarkers are blindly evaluated, but the table also shows the evaluation using all data. CVR always performed the best, with a detectable effect size below 12% in CI, and around 7% for CU. Ventricles performed second best, with effect sizes between 16-34% in CI and 8% in CU, which rivals CVR. The entorhinal cortex could detect effect sizes of 22-63% for CI, and 24-30% for CU. The hippocampus could detect effect sizes of 30-124% for CI and 26-34% for CU. Last, and worst, the whole brain could detect effect sizes of 60-291% in CI, and 29% in CU.

Regarding the A4 trial MRI data [132], we compared the annualised percent change between the Solanezumab ($N = 401$) and Placebo ($N = 423$) arms for each biomarker using ANCOVA models (Supplementary Table B.1). Our CVRs consistently demonstrated larger, more directionally favourable treatment effects than standard volumetric measures, suggesting superior sensitivity to atrophy in a clinical trial and validating our algorithm's utility. Specifically, the CVR.CU.1 biomarker (discovered in ADNI's Subset 1) showed the largest numerical slowing of atrophy, with the solanezumab group exhibiting a 0.28% smaller annualised decline than the placebo group (LS-mean difference = -0.28% , 95% CI $[-0.67, 0.12]$, $p = 0.17$). The CVR.CU.2 measure (discovered in ADNI's Subset 2) showed a similar trend (difference = -0.20% , 95% CI $[-0.59, 0.19]$, $p = 0.32$). In contrast, the effect on traditional volumetrics was negligible; for example, the adjusted difference for hippocampus/ICV was nearly zero (0.01% , 95% CI $[-0.17, 0.18]$, $p = 0.95$).

5.4 Discussion

This chapter introduces and validates a novel approach to monitoring brain atrophy in the context of Alzheimer's disease by employing composite value ratios (CVRs) of regional brain volumes derived from MRI, thanks to the `BioDisCVR` framework introduced in Chapter 4. The findings robustly demonstrate that these data-driven CVR biomarkers significantly outperform traditional volumetric measures (including whole brain, hippocampus, entorhinal cortex and ventricular volume) across multiple key metrics relevant to clinical trial design and disease monitoring. Specifically,

CVRs consistently exhibited lower sample size estimates, reduced measurement error (as per the linear mixed-effects model fitting), and enhanced group separation between amyloid-positive and amyloid-negative individuals, as well as between cognitively unimpaired and impaired individuals. This performance advantage was particularly pronounced in more advanced stages (cognitive and neurodegenerative) of disease.

Considering **Sample Size Estimates** (SSE), the advantage of CVR biomarkers is particularly notable. Our analysis revealed that CVRs require a substantially smaller sample size to achieve the same statistical power in hypothetical clinical trials compared to traditional volumetric measures. Specifically, in trials on cognitively impaired individuals, such as in early Alzheimer's disease (e.g., TRAILBLAZER-ALZ-2), CVRs reduced the required sample size by 67-73% compared to ventricles, the best-performing traditional biomarker in this context. The reduction was still notable for preclinical trials in cognitively unimpaired individuals (such as the A4 Study), ranging from 20-34%. This translates to potentially considerable cost- and time-savings in clinical trials, underscoring the enhanced sensitivity of CVRs to detect changes to disease-related biomarker trajectories with fewer participants. Worth mentioning here that Alzheimer's disease clinical trials are not typically powered by volumetric biomarkers, as cognition or early pathologies are preferred. Nevertheless, as new biomarkers emerge, they could complement cognitive and pathological measures, or even challenge the old choices of trial endpoints, or provide insights into disease progression. On another note, if trials are eventually moving to earlier stages of the disease, then cognitive outcomes might not even be a useful measure, within the scope of the trial.

In terms of longitudinal measurement of **Error**, CVR biomarkers again demonstrated clear superiority. The percentage error, representing the residual variability of the linear mixed-effects models, was consistently and considerably lower for CVRs compared to traditional measures across all analyses. In cognitively impaired individuals, CVRs exhibited a 47-66% reduction in percentage error compared to whole brain volume, the best traditional biomarker in this metric for this group. Even in preclinical populations, CVRs showed a 6-39% error reduction. This reduced measurement error signifies a more precise and reliable assessment of longitudinal

volume changes, which is crucial for accurately monitoring disease progression and evaluating treatment effects. Furthermore, such improved accuracy in individual trajectories could enhance prognostic value and inform clinical decision support in routine healthcare settings for individuals at risk of or living with Alzheimer's disease.

Regarding **Group Separation**, CVRs consistently provided enhanced differentiation between groups relevant to AD pathology. CVRs showed a 38-66% higher t-statistic for amyloid positivity and a 23-85% higher t-statistic for cognitive impairment compared to ventricles (the best-performing traditional biomarker). This improved group separation indicates that CVRs are more effective at capturing the disease signal, distinguishing individuals based on key AD-related characteristics. This enhanced ability to discriminate between groups is vital for biomarker validation, for stratifying patients in clinical trials (as demonstrated in STX), and potentially for use in diagnostic applications in memory clinics.

The detectable effect size differences are considerable, using configurations of past clinical trials. It should be noted that our evaluated biomarkers were designed for a wider population, whereas the clinical trials' inclusion criteria had various restrictions (MMSE, CDR, age). This makes the performance even more remarkable, as it demonstrates that MRI CVRs can achieve robust and reliable results without being tailored to highly selected cohorts, underscoring their potential applicability across diverse groups, at different stages of the disease.

We hypothesise that the superior performance of CVRs might be due to their inherent capacity to normalise for inter- and intra-subject variability and global factors, thanks to their ratio-based construction. This normalisation may reduce noise and amplify the signal specific to regional volume changes associated with AD pathology. This enhanced signal-to-noise ratio and reduced measurement error not only benefit group-level analyses in clinical trials but also hold significant value at the individual level. Clinicians could leverage these biomarkers to detect subtle, clinically meaningful changes earlier and more confidently, facilitating timely and personalised interventions for better patient care. The identified CVRs consistently featured ventricles in the numerator, suggesting that global changes reflected by ventricular expansion are a crucial component of atrophy monitoring. An interesting advantage of the ventricles is that they are likely summarising atrophy happening

elsewhere, and their size and contrast to brain tissue likely make them an easy target for accurate segmentation. The presence of the nucleus accumbens in the denominator across different CVRs is more intriguing, but not unexpected: a study by Nie et al. [168] found a correlation between the nucleus accumbens volume and clinical rating scales (MMSE and MoCA).

Although the A4 trial did not meet its primary clinical endpoint, our re-analysis of its imaging data (MRI) provides valuable insights into biomarker sensitivity. Although our analysis did not identify a statistically significant treatment effect for any biomarker (indeed, this difference may not exist), we noted a consistent pattern in which our data-driven CVRs produced considerably larger effect sizes and lower p-values than traditional volumetric measures like hippocampus/ICV. This finding is important for two main reasons. First, it suggests that our CVRs may have greater sensitivity to detect subtle, treatment-related signals of slowed atrophy, even in a trial where the overall therapeutic effect was modest or non-existent. This enhanced sensitivity could be critical for future trial design, potentially enabling the detection of a biological effect with smaller sample sizes or shorter durations. Second, it highlights the potential limitations of relying solely on traditional, single-region biomarkers in preclinical or prevention trial settings where expected changes are minimal. The results from this single trial are, of course, exploratory and require confirmation. However, they strongly motivate the usage of these optimised CVRs in pooled analyses across multiple prevention studies or in future trials to definitively establish their added value as sensitive endpoints for measuring disease modification.

There are limitations to the application of volume-based biomarkers in Alzheimer's disease trials. Firstly, anti-amyloid immunotherapies have been reported to show accelerated "pseudo" atrophy in the whole brain, hippocampus, and ventricles [169]. This could confound any volumetric biomarker in such a trial, including CVR, but would be easy to test on patient-level data from recent clinical trials. An additional limitation is that the cohorts contributing data to this study currently lack ethnic diversity, which may affect generalisability. Work is underway to rectify this in ADNI [170] (and elsewhere), but it may be some time before the impact of ethnicity on brain atrophy quantification is fully understood. Another limitation of the study is that the BioDisCVR framework was applied to data obtained by a specific

pipeline (from the image acquisition to the segmentation, even if from 61 different sites), so potentially useful region candidates could have been discarded by the final CVR biomarker due to high variance (e.g., segmentation errors inducing a lower signal-to-noise ratio). So, while the reported biomarkers have proven useful in our experiments, further work is possible to advance understanding of region-specific biological contributions.

Beyond addressing limitations, there are other avenues for future work. Validation on patient-level data from completed clinical trials in Alzheimer's disease is a clear priority, where smaller effect sizes could be detected on secondary imaging biomarker outcomes. The challenge here is the general reluctance for trial sponsors to share patient-level data, e.g., Clarity-AD (see Supplementary Material from van Dyck et al. [79]). Indeed, given the notable impact of the clinical trial configuration (Supplementary Table ST1), BioDisCVR holds promise for clinical trial design. The scope of this study was specific to Alzheimer's disease, but perhaps CVR could prove useful in application to other diseases, or even normative modelling.

5.5 Conclusions

In conclusion, this study demonstrates the superior performance of composite value ratio (CVR) biomarkers for monitoring brain atrophy in Alzheimer's disease. By applying the BioDisCVR framework, originally developed for tau PET, to MRI-based volumetric measures, we have successfully translated the benefits of ratio-based biomarkers to a more widely accessible and cost-effective imaging modality. The enhanced disease signal extraction achieved through CVRs (evidenced by significantly reduced noise and increased group separation) translates directly into tangible advantages for both research and clinical practice, over and above traditional volumetric biomarkers, including ventricles, whole brain, and hippocampus.

Table 5.7: Detectable effect size per example clinical trial. Given a past clinical trial design (number of participants, trial duration and inclusion criteria), we show the hypothetical detectable effect size per biomarker. The calculation is given by first fitting a linear mixed-effects model to the log-transformed biomarker, with covariates age, sex, and APOE4. All classical biomarkers (whole brain, ventricles, hippocampus, entorhinal cortex) are divided by the intracranial volume to improve robustness. Our biomarkers' names (CVR) indicate the target cognitive group and data subset they have been trained for: CU = cognitively unimpaired; CI = cognitively impaired; suffix .1 and .2 indicate the subset that was used to obtain the biomarker. In bold, for the subset analysis, the best measures per trial, not including the biomarkers that used the same subset data. Asterisk (*) indicates that the region was used for the trial as secondary outcome. Non-CVRs are divided by Intracranial Volume (ICV). TB-ALZ 2 is the TRAILBLAZER-ALZ 2 clinical trial.

Detectable effect size, considering all data (and inclusion criteria per trial)

Biomarker	A4 Study	EMERGE/ENGAGE	Clarity AD	TB-ALZ 2
CVR.CI.1	0.081	0.112	0.083	0.112
CVR.CI.2	0.068	0.106	0.105	0.112
CVR.CU.1	0.065	0.113	0.152	0.141
CVR.CU.2	0.065	0.12	0.124	0.125
whole brain	0.299*	1.554*	0.659*	2.347*
ventricles	0.080*	0.176*	0.263*	0.241*
hippocampus	0.301*	0.734*	0.312*	1.142*
entorhinal	0.282	0.532	0.268	0.491

Detectable effect size, considering Subset 1 (and inclusion criteria per trial)

Biomarker	A4 Study	EMERGE/ENGAGE	Clarity AD	TB-ALZ 2
CVR.CI.1	0.084	0.102	0.066	0.097
CVR.CI.2	0.073	0.106	0.114	0.119
CVR.CU.1	0.063	0.11	0.167	0.165
CVR.CU.2	0.074	0.121	0.132	0.139
whole brain	0.291*	1.636*	0.829*	2.911*
ventricles	0.083*	0.168*	0.338*	0.212*
hippocampus	0.262*	0.784*	0.302*	1.245*
entorhinal	0.307	0.633	0.224	0.554

Detectable effect size, considering Subset 2 (and inclusion criteria per trial)

Biomarker	A4 Study	EMERGE/ENGAGE	Clarity AD	TB-ALZ 2
CVR.CI.1	0.079	0.123	0.106	0.131
CVR.CI.2	0.063	0.106	0.089	0.107
CVR.CU.1	0.067	0.116	0.117	0.121
CVR.CU.2	0.057	0.118	0.108	0.116
whole brain	0.296*	1.396*	0.609*	1.569*
ventricles	0.077*	0.184*	0.342*	0.277*
hippocampus	0.345*	0.673*	0.333*	1.093*
entorhinal	0.243	0.455	0.267	0.42

Chapter 6

Weighted Composite Value Ratio, a case with amyloid peptides

This chapter introduces a novel extension of CVR, introducing feature importance through a weighted combination of inputs. This approach was specifically developed in response to a clinically-driven need identified by our laboratory-based colleagues, listed below, who sought to understand the relative importance of different amyloid-beta peptides in Alzheimer's disease pathogenesis. The work described in this chapter is based on a paper sent to Alzheimer's Research & Therapy for publication [December 2024], where I devised the data analysis (methods, results, discussion). Other coauthors were involved in the creation of new data (one out of the four datasets used, the iPSC one), facilitated funding, gave feedback on the manuscript, and put the work in the context of the literature (Rebecca Gabriele, Teisha Bradshaw, Claire Leckey, Rohan de Silva, Nick C. Fox, Selina Wray, Neil Oxtoby and Charles Arber).

6.1 Introduction

In Chapter 4 we introduced the concept of a composite value ratio (CVR), along with a framework to discover specific such biomarkers. In that case-study, the explored composition only included the mean of regional PET signal and a volume-weighted mean, but the concept is flexible, and can accommodate other ways of combining the different regions or inputs, like in Chapter 5, where CVR was applied to MRI-derived volumes. Here, we use a linear combination of inputs (weighted composite value ratio, or wCVR), as a means to explore feature importance, in particular, with another characteristic feature of Alzheimer's disease: amyloid-beta ($A\beta$) peptides.

Changes to the relative abundance of A β peptides are characteristic of Alzheimer's disease. Induced pluripotent stem cells (iPSC) neuronal models offer a physiological model of A β production. In the literature, different biomarkers, using different A β peptides are proposed, with ratios being the norm. This chapter will use the BioDisCVR framework, with a weighted-mean composition to investigate combinations of A β peptides as AD biomarkers and the relative contribution of peptides to AD pathogenesis.

6.2 Methods

6.2.1 Data

Data analysed in this work is the A β measures from iPSC and cell-based neuronal models, collected from four independent sites. The iPSC is from University College London colleagues (described in our manuscript currently submitted for publication, with pre-print available at <https://doi.org/10.1101/2024.11.23.624811>). The cell data consists of published A β profiles generated by overexpression of familial AD-linked PSEN1 mutations in PSEN1/PSEN2 double knockout cell lines. This includes mouse embryonic fibroblast cell lines by Petit and colleagues (2022) [171], and by overexpression in HEK 293T cell lines by Liu and colleagues (2023) [172] and Schultz and colleagues (2024) [173]. A key distinction is that the iPSC model provides a more accurate representation of human neuronal physiology and pathology compared to traditional cell-based models. The available data numbers from controls and *PSEN1*-mutations are shown in Table 6.1. The publication from Schultz et al. [173] included data from the study from Liu et al. [172]; we only considered the non-overlapping, unique, data from Shultz et al.

Table 6.1: Data used in this study comes from four sites. The smaller variability in controls peptide proportional counts allows to work with smaller sample sizes, compared to the mutation-carriers. Only the unique data from Shultz et al. is included, as the study used data from Liu et al.

Site	Controls	PSEN1 mutations
iPSC	3	10
Petit	2	21
Liu	3	120
Schultz	1	56

6.2.2 Modelling and analysis

Data harmonisation. Samples with any peptide measure equal to zero were removed ($N = 14$). Zeros could potentially be due to assay detection limits or errors, and impede the usage of the peptide as a denominator, for these samples. Samples were averaged across batches to reduce the influence of measurement error. Data harmonisation across all datasets was performed in two steps. First, each sample's data was divided by the sum of the available peptides within that sample. This operation adjusts for differences in total peptide amounts across samples, as we are interested in their relative amounts. Second, each peptide measure was further divided by the average of the control samples for the same dataset to correct for any assay-specific systematic biases.

To evaluate biomarkers that include the addition of multiple peptides, each peptide measurement was scaled back to the relative control proportions based on data from Petit et al. [171]. Ratio biomarkers like the short/long [171] involve a weighted sum where each peptide is weighted by its naturally occurring abundance in controls. This scaling step is essential because the relative abundances of A β peptides differ by orders of magnitude. Not scaling the data back would distort the setup for which the biomarker was originally designed.

Statistical analysis We evaluate biomarkers in case-control classification and age-at-onset regression scenarios. In classification experiments we analyse the area under the receiver operator characteristic curve (ROC AUC), which plots sensitivity, or true positive rate $TP/(TP+FN)$ against false positive rate $FP/(FP+TN)$, where T/F indicates True/False and P/N indicates Positive/Negative. Since the data is strongly unbalanced (very few control samples), we also look at the precision-recall curve (PR AUC) of the smaller group, which plots control precision $TN/(TN+FN)$ against control recall $TN/(TN+FP)$, since PR AUC is a preferred metric in this scenario.

To evaluate the correlation between the biomarker and the age at onset (AAO), we employ the coefficient of determination (R^2), using Pearson correlation, since peptide biomarkers have shown good linear correlation [171].

Median and 95% confidence intervals are reported for all metrics using 2000 bootstrapping samples, stratified in the classification case (to accommodate the heavy class imbalance).

We bootstrap the weighted composite value ratio (wCVR, explained below), running the search algorithm five times per bootstrap with random initialisations, to investigate the effect of the weighting of the peptides. We also do a grid-search analysis, where we fixed the A β 42/40 ratio as baseline for R^2 (state-of-the-art in terms of correlation with age-at-onset [171]), and then calculated the metrics of all combinations of the other peptide weights from 0 to 1 by 0.02 (either in the numerator or denominator). We did the same with precision-recall of the controls, but the baseline ratio was A β 37/42 (state-of-the-art in terms of classification in AD [172]).

Performance Evaluation and Additional Analyses. To assess model generalisation and to test whether the superior performance of wCVR reflected overfitting, we compared it with alternative biomarkers using full leave-one-out cross-validation (LOOCV) across all datapoints (mutation carriers and controls). For regression, each biomarker was fitted with a linear model to predict age at onset (AAO). For classification, in each training fold, the decision threshold was set to the most extreme control biomarker value (in the abnormal direction), prioritising specificity for the control group (the smaller class).

Alongside wCVR, we evaluated: (i) literature biomarkers described in this chapter, (ii) a linear model and random forest models trained separately for regression and classification, and (iii) a novel weighted version of the `theta` model (global peptide profile angle) [174], introduced in Chapter 7. For wCVR and wTheta, the three binary digits in the model name indicate whether the BioDisCVR optimisation considered, respectively, Pearson correlation with AAO, AUC, and precision–recall (combined multiplicatively).

Performance was quantified from LOOCV predictions. For regression, we computed the Mean Absolute Error (MAE); for classification, we calculated F1-Score. Uncertainty in point estimates was assessed via non-parametric bootstrap (2000 replicates) to obtain 95% confidence intervals. Because all models were evaluated on identical holdout sets, pairwise comparisons used a paired framework: MAE differences were tested with the paired Wilcoxon signed-rank test, and classification metric differences were assessed via empirical two-sided p -values from the bootstrap distribution (2000 replicates). Multiple testing was controlled using

the Benjamini–Hochberg procedure (FDR 5%), with adjusted $p < 0.05$ considered statistically significant.

Biomarkers Different ratio-based biomarkers have been proposed in the field, namely $A\beta_{42/40}$, $A\beta_{37/42}$ [172] and the short/long ratio [171], which is $A\beta_{(37+38+40)/(42+43)}$. We ran BioDisCVR [158] to find which peptides are selected to go to the numerator, and which are selected to go to the denominator, along with their linear combination weights, obtaining a weighted composite value ratio (wCVR). The general formula is:

$$wCVR = \frac{\sum_{i \in X_1} k_i \cdot X_{1i}}{\sum_{j \in X_2} k_j \cdot X_{2j}} \quad (6.1)$$

, where X_1 and X_2 are non-overlapping non-empty sets of the available peptides, and k is a weighting factor for each peptide. To drive the algorithm, we consider different configurations, which we indicate with the suffixes -R for R^2 , -P for PR AUC, and RP for both (i.e., when the objective function is the product of both metrics), since the BioDisCVR framework allows for a user-chosen search algorithm. We chose to run a genetic algorithm, as described in the original work [158].

Additionally, we evaluated all possible combinations of ratios (180 possible combinations) where either the numerator or denominator is one peptide, or the addition of multiple peptides, without repetition, and conserving the different proportions of the peptides from Petit et al. [171]. In this case, the search algorithm was not needed, as we could evaluate all possible combinations.

6.3 Results

Figure 6.1 shows, for the UCL-generated data, the literature biomarkers from the peptides measured in conditioned media from 10 PSEN1 mutation-carrying patient-derived iPSC-neuronal lines in triplicate, as well as 3 controls. As expected for EOFAD data, there is visual separation between cases and controls, as well as variation between PSEN1 patient-derived lines.

Figure 6.2 shows the peptide distribution of all EOFAD data, relative to the controls. Most deviations of $A\beta_{37}$, $A\beta_{38}$ and $A\beta_{40}$ occur below the mean of controls (91.3%, 82.6% and 72.9%, respectively), while most deviations of $A\beta_{42}$ and $A\beta_{43}$ occurred above the mean of controls (97.1% and 84.5%, respectively).



Figure 6.1: Literature biomarkers from the peptides measured in conditioned media from 10 PSEN1 mutation-carrying patient-derived iPSC-neuronal lines in triplicate, as well as 3 controls. This is for the newly-generated data from UCL.

To allow comparison of Aβ data from published datasets and my colleagues' newly generated data, harmonisation was required because of widely acknowledged inter-assay variability (Figure 6.3, left panels). Harmonisation was achieved by weighting peptides based on the average relative abundance of each peptide found in control cell data within each dataset. Converting to a dimensionless, normative scale (Figure 6.3, right panels) facilitates both data interpretation and combination or comparison of multi-assay datasets (Aβ peptide ratios have been shown to be highly replicable between control lines [90]). Figure C.1 shows the relative peptide amount of control cell lines for each dataset before harmonisation. Although all datasets share similar general relative abundance (e.g., Aβ40 being the most abundant, and Aβ43 being the least), there are some important differences, especially for our cohort (iPSC), where the relative abundance of Aβ37 is lower and the Aβ38 is higher than the rest of the cohorts.

Table 6.2 shows the uncorrected correlation (Pearson's r and Kendall's τ) with the age-at-onset for all peptides, per dataset (pre-transformation), and for all data

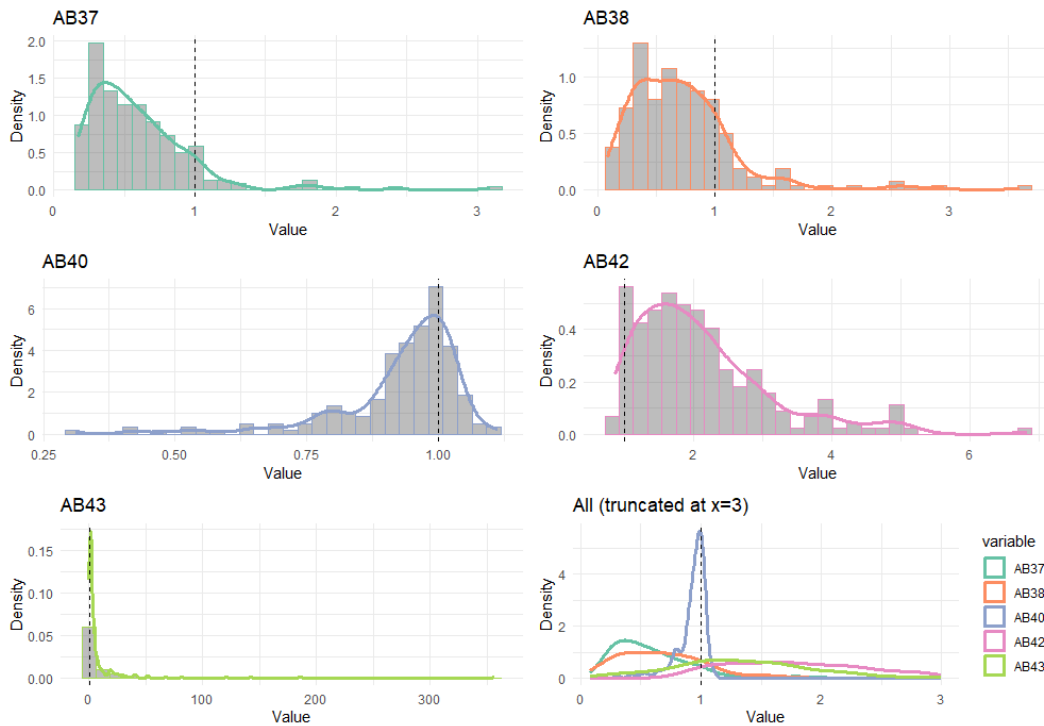


Figure 6.2: Peptide distribution relative to the controls, of all merged and harmonised datasets. We can observe that the deviations from the controls (value = 1) are not symmetrical.

merged (post-transformation). The measures are all relative to the sum of the five peptides. All the statistically significant comparisons ($p \leq 0.05$, in bold, and thus the subset of corrected p-values) share the same message: positive correlation for A β 37, A β 38 and A β 40, and negative correlation for A β 42 and A β 43.

Using all harmonised data, we assessed the association between the position of the mutation in PSEN1 and age-at-onset. Kendall's τ coefficient was calculated to be 0.0104 ($z = 0.228$, $p = 0.820$). The Pearson correlation coefficient was 0.0118 ($t = 0.175$, $df = 219$, $p\text{-value} = 0.861$). The practically zero coefficients and high p-values suggest that the observed correlation is not statistically significant. Figure S4 shows a scatterplot of AAO versus location, and A β 42/40 versus location.

A dataset-specific evaluation of biomarkers is shown as heatmaps in Figure 6.4. We compare established ratios (A β 42/40, A β 40/42, A β 37/42 and A β short/long) with weighted combined value ratio (wCVR) analyses with different configurations (see methods), where wCVR was optimised per cohort. For both correlation with age-at-onset (coefficient of determination, Figure 6.4, left panel) and case-control classification (precision-recall, Figure 6.4, right panel), the heatmaps show per-

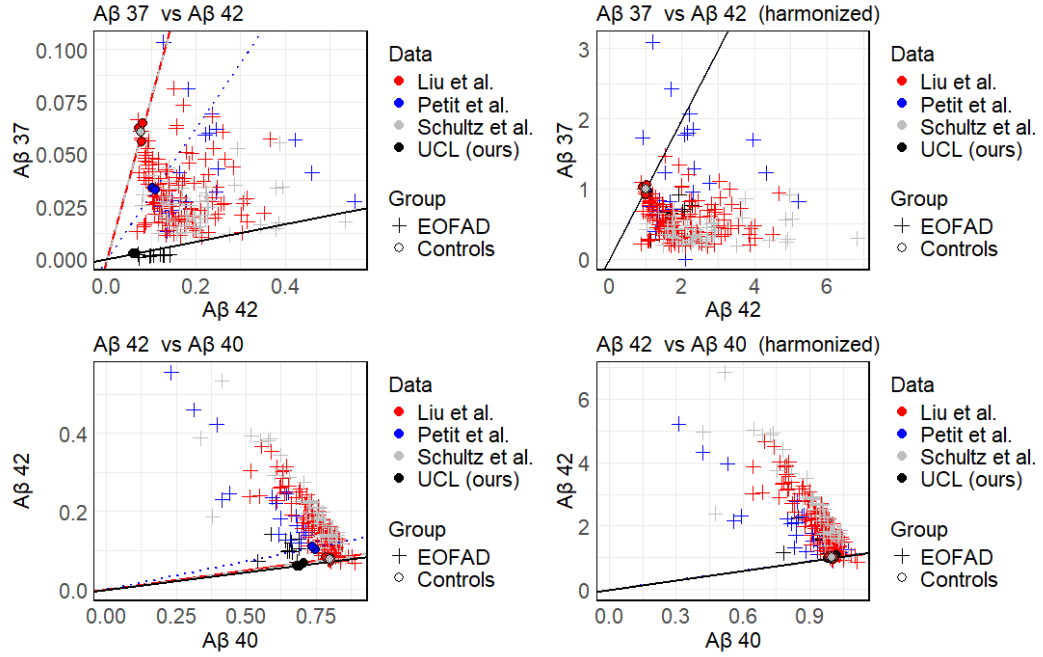


Figure 6.3: Example of the need for harmonisation, and the result of it. Top-left ($A\beta_{37/42}$) is the most striking example, where we can observe three distinctive ratios (directions) for controls, rendering the data useless if merged. This is solved in the top-right panel, where all controls are aligned, and we can define a global threshold for classification. The bottom row shows a similar case, but with other peptides ($A\beta_{42/40}$), where the need for harmonisation is not as evident.

formance heterogeneity between the cohorts. Data from iPSC, Petit or Schultz classify well (perfect classification for all biomarkers except for $A\beta_{42/40}$, $A\beta_{37/42}$ and $A\beta_{40/42}$), and data from Petit exhibits highest coefficient of determination, followed by iPSC. As for the biomarkers, it shows that wCVR achieves the best classification whenever it is directed at precision-recall of the controls (wCVR-P), and wCVR-R shows the best correlation with age-at-onset when driven by the coefficient of determination, together with the $A\beta_{short/long}$ ratio and the $A\beta_{40/42}$, whenever it incorporates this metric.

Figure 6.5 shows the results of our ratio biomarker experiments. The map of R^2 (AAO) vs precision-recall AUC shows wCVR biomarkers towards the upper right quadrant, demonstrating superior performance than most simple combinations of peptides. The literature biomarkers also do well in one metric each, but not both: $A\beta_{short/long}$ shows high R^2 , and $A\beta_{37/42}$ outperforms most simple ratios in precision-recall AUC.

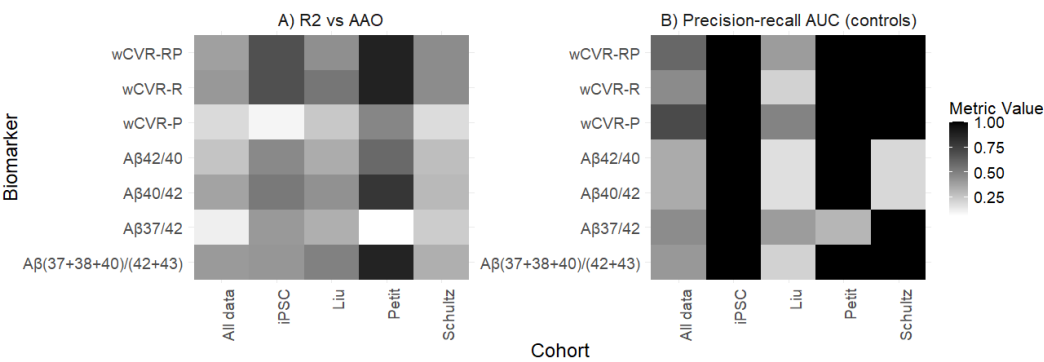


Figure 6.4: Heatmap of results per biomarker, per site. In both panels, 1 (black) represents the best possible metric. Panel A shows the coefficient of correlation of the biomarker against age-at-onset. Panel B represents the precision-recall area under the curve for the controls.

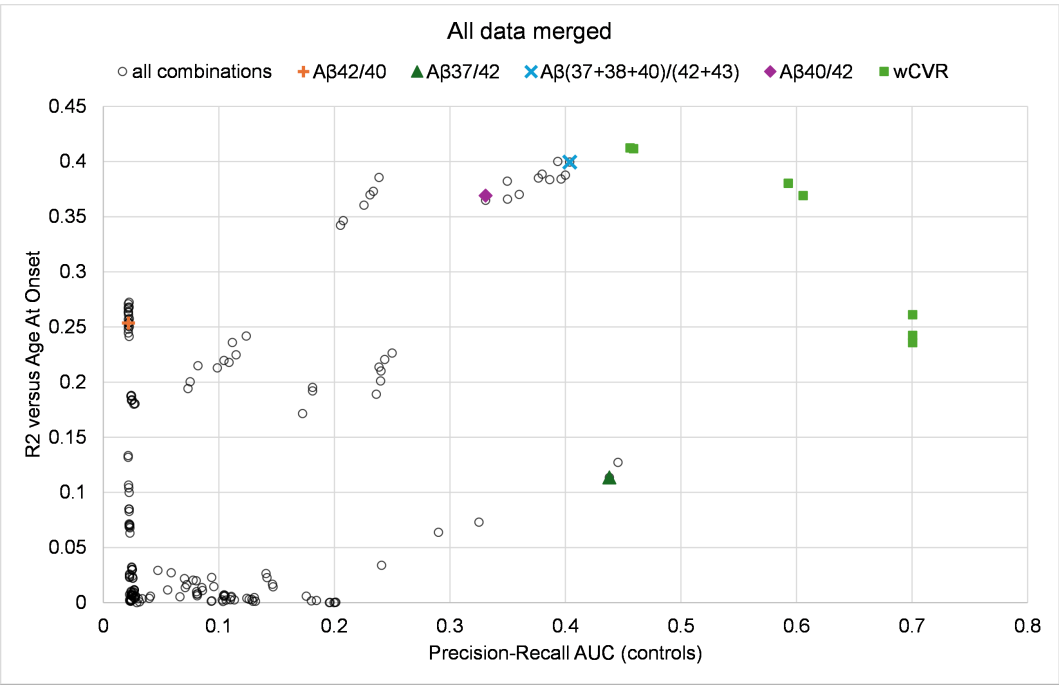


Figure 6.5: Visual plot of evaluation metrics for all biomarkers. Higher number = better, in both dimensions.

Table 6.2: Correlation of peptides and age-at-onset. In bold, uncorrected p-values below or equal to 0.05.

	Peptide	Pearson		Kendall	
		r	p-value	τ	p-value
UCL	AB37	0.114	0.754	-0.068	0.787
	AB38	-0.156	0.666	-0.068	0.787
	AB40	0.462	0.179	0.296	0.241
	AB42	-0.559	0.093	-0.341	0.176
	AB43	0.358	0.310	0.159	0.528
Petit	AB37	-0.141	0.512	-0.167	0.268
	AB38	-0.093	0.666	-0.022	0.902
	AB40	0.781	0.000	0.534	0.000
	AB42	-0.834	0.000	-0.710	0.000
	AB43	-0.292	0.167	-0.241	0.101
Liu	AB37	0.253	0.004	0.172	0.004
	AB38	0.194	0.026	0.190	0.001
	AB40	0.486	0.000	0.342	0.000
	AB42	-0.641	0.000	-0.521	0.000
	AB43	-0.238	0.006	-0.144	0.016
Schultz	AB37	0.043	0.751	0.092	0.319
	AB38	-0.007	0.958	0.187	0.043
	AB40	0.467	0.000	0.334	0.000
	AB42	-0.581	0.000	-0.430	0.000
	AB43	0.001	0.995	0.063	0.497
All	AB37	0.031	0.644	0.074	0.104
	AB38	0.066	0.330	0.103	0.024
	AB40	0.493	0.000	0.343	0.000
	AB42	-0.580	0.000	-0.447	0.000
	AB43	-0.108	0.109	-0.095	0.038

Table 6.3 contains the quantification of the performance of the salient biomarkers, with 95% confidence intervals from bootstrapping (see Methods). It is noteworthy that wCVR is able to improve on established ratios from the literature, however, different wCVR biomarkers excel at precision-recall versus coefficient of determination.

Table 6.4 lists the peptide weights (denominator in red, with an asterisk) for the different wCVR configurations, along with their classification and regression performance. There is a trade-off between the two evaluation metrics (classification and regression), which is determined by the weight shifting and position in the ratio of the peptides. In all cases, A β 40 is in the numerator and A β 42 and A β 43 are in the denominator. A β 38 is actively (10-19%) in the numerator when the objective function includes the coefficient of determination, and A β 37 appears in the numerator when classification (precision-recall) is considered. When considering both metrics of

Table 6.3: Quantification of biomarker performance, with 95% confidence intervals.

Our biomarkers are the weighted composite value ratios (wCVR), with the suffix -R indicating it was optimised for R^2 (versus age-at-onset), the suffix -P indicating it was optimised to maximise precision-recall of controls, or both metrics (-RP). AUC is the classical area under the receiver-operating characteristic curve.

Biomarker	AUC	PR AUC controls	R^2
wCVR-R	0.977 (0.953, 0.994)	0.456 (0.287, 0.805)	0.412 (0.284, 0.526)
wCVR-RP	0.986 (0.966, 0.997)	0.593 (0.363, 0.934)	0.380 (0.255, 0.497)
wCVR-P	0.992 (0.977, 1.00)	0.700 (0.435, 1.00)	0.242 (0.145, 0.350)
A β 42/40	0.961 (0.929, 0.983)	0.345 (0.217, 0.615)	0.259 (0.187, 0.343)
A β 37/42	0.977 (0.953, 0.993)	0.454 (0.29, 0.785)	0.114 (0.048, 0.230)
short/long	0.974 (0.947, 0.991)	0.411 (0.265, 0.729)	0.403 (0.275, 0.517)

classification and regression, the composite value ratio becomes similar to the short/long ratio, with notably more weighting of the A β 42 in the denominator, and A β 40 in the numerator. Table 4B shows the wCVR for the individual datasets, demonstrating the different contribution of each peptide to combined performance (precision-recall plus coefficient of determination) for each dataset. Note that A β 42 and A β 43 are consistently in the denominator, while A β 37, A β 38 and A β 40 are in the numerator.

Table 6.4: Peptide relative weights, to the denominator (in red, with an asterisk ×) and the numerator (black, no asterisk). Although the ranking and position (numerator/denominator) of the relative peptide weights are the same, in pre-harmonised data, there are differences in the proportions. We observe a trade-off in performance from optimising for classification (precision-recall area under the curve, for controls) or regression (versus age-at-onset). The prefix of the weighted composite value ratio (wCVR) indicates the metric it was optimised for: R for regression, P for precision-recall, and RP for both (multiplied).

A) Data-driven weighting and combination of peptides (wCVR)

A β 37	A β 38	A β 40	A β 42	A β 43	biomarker	PRAUC	R^2
*0.2%	19.0%	81.0%	*97.8%	*2.0%	wCVR-R	0.456	0.412
21.0%	9.8%	69.2%	*93.6%	*6.4%	wCVR-RP	0.593	0.382
88.9%	*27.8%	11.1%	*59.0%	*13.2%	wCVR-P	0.705	0.185

B) Short/long relative weights per cohort (pre-harmonisation)

A β 37	A β 38	A β 40	A β 42	A β 43	Cohort
0.3%	25.7%	74.1%	*99.7%	*0.3%	iPSC (UCL)
3.7%	13.0%	83.3%	*99.4%	*0.6%	Petit
6.6%	6.7%	86.7%	*98.6%	*1.4%	Liu
6.6%	6.6%	86.8%	*98.6%	*1.4%	Schultz

Results from a permutation feature importance analysis are shown in Supplementary Table C.1. For all metrics, and for both displayed biomarkers (A β short/long

and wCVR-RP), A β 42 is the peptide providing the biggest impact. The data-driven wCVR-RP better employs information from A β 43, as the importance values are higher in all metrics.

When bootstrapping the aggregated data and running BioDisCVR three times to get the best wCVR, different patterns were observed depending on the metric used for optimisation. Figures C.2, C.3 and C.4 show the boxplots of the weights for 200 bootstrapped samples, optimised for R^2 , precision-recall, and both metrics multiplied. However, these weights were obtained by fitting different resampled data, so to have a better understanding of their effect, Figure C.5 shows the evaluation metrics against peptide weights. We can observe a clear convex space of data points in all cases, with a slight truncation for the precision-recall plots. A similar pattern of an optimal weight combination can be observed in Figure 6.6, where the grid-search analysis of weighted combinations shows a single peak of performance (only the top 10% of points are shown, amounting to 0.9% of the total explored combinations), which corresponds to the reported wCVR weights. This figure was created by calculating the two evaluation metrics for each combination of peptide weights, ranging from -5 to 5 in 0.1 increments, where negative weight meant the peptide was in the denominator, and the baseline ratio was A β 42/40 for R^2 and A β 37/42 for PR AUC, upon which we introduced the rest of the peptides. For example, if the weights assessed for A β 37, 38 and 43 are -0.25, 0.10 and 2.0, respectively, then the biomarker assessed for R^2 would be $\frac{A\beta42 \cdot 1 + A\beta38 \cdot 0.1 + A\beta43 \cdot 2.0}{A\beta40 \cdot 1 + A\beta37 \cdot 0.25}$. These plots all show that employing multiple peptides improves performance in each case.

Leave-One-Out Cross-Validation Analysis We evaluated biomarker models for classification (mutation carriers vs controls) and regression (predicting age at onset, AAO) using leave-one-out cross-validation (LOOCV). Full results are in Table 6.5; pairwise statistical comparisons are in Supplementary File wCVR.results.xlsx.

For classification, several models achieved near-perfect performance. The unweighted *theta* ratio (global peptide profile angle) achieved the highest F1-Score at 0.993 (95% CI: 0.984–1.000), driven by high recall (0.996) and precision (0.991). Custom-optimised weighted *theta* models also performed strongly ($F1 = 0.986$), and, consistent with our specificity-prioritising threshold, most models had precision

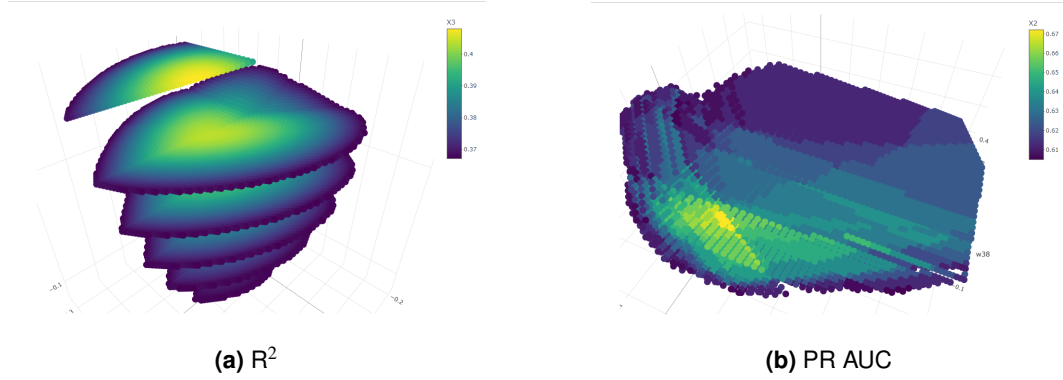


Figure 6.6: Heatmap of evaluation metrics. The goal of this figure was to visualise how the metrics shifted when the proportions of peptides shifted, and to see if there were multiple or a single “hot spot” (local/global maxima). We only found a single hotspot. Left panel (a) is the R^2 versus age-at-onset, and left panel (b) is the precision-recall area-under-the-curve of controls. Given a base ratio of A β 42/40 for R^2 , we explore how the metric changes when incorporating other peptides, in a grid-search (0 to 5 in 0.1 increments, both for numerator and denominator). Panel b) shows the same, but for the precision-recall of controls, and the base ratio is, in this case, A β 37/42. Only the top-performing 10% of points are shown, which amounts to 0.9% of the total explored combinations.

> 0.99 except the `RandomForest_Classifier` (0.974). Recall was the main differentiator: `shortlong` (short/long peptide ratio) and `RandomForest_Regressor` had markedly lower recall (0.753 and 0.852, respectively), resulting in significantly lower F1-Scores (adjusted $p < 0.05$).

For regression, the ranking shifted slightly. The optimised weighted `theta` models achieved the lowest MAEs, with `wtheta.111` at 6.045 years (95% CI: 5.450–6.670), significantly outperforming literature biomarkers. `wCVR` variants and the `LinearModel_Regressor` also performed well (MAE 6.1–6.3 years). In contrast, `shortlong` and `RandomForest_Regressor` had higher errors (> 7.0 years), indicating reduced accuracy for AAO prediction.

6.4 Discussion

In this chapter we used A β measurements of five A β peptides to investigate A β ratios in an unbiased, data-driven manner. We explored the relative contribution of each peptide to disease classification and onset prediction by extending the CVR concept to include weights, interpretable as “importance” for the given predictive/diagnostic task of interest. The results show a differential contribution of longer peptides (A β 42 and A β 43) and shorter peptides (A β 40, A β 38 and A β 37) to disease. A β 40 and

Table 6.5: Cross-validated performance of biomarker models for classification and regression tasks. Classification accuracy is reported as the F1-Score (harmonic mean of precision and recall; higher is better) for distinguishing mutation carriers from controls. Regression accuracy is reported as the Mean Absolute Error (MAE, in years; lower is better) for predicting age at onset (AAO). All values are point estimates with 95% confidence intervals in parentheses, obtained via leave-one-out cross-validation and 2000 bootstrap samples. The three binary digits after the models wCVR and wTheta correspond to whether the model considered the following metrics during optimisation through the BioDisCVR framework: Pearson correlation (against age-at-onset), AUC and precision-recall, respectively (multiplying them). In bold, the best-performing model for each metric.

biomarker_name	F1_Score	MAE
Abeta42/40	0.963 (0.944-0.980)	6.850 (6.167-7.594)
Abeta37/42	0.982 (0.968-0.993)	6.892 (6.234-7.546)
RandomForest_Classifier	0.982 (0.968-0.993)	7.870 (7.042-8.740)
RandomForest_Regressor	0.918 (0.889-0.943)	6.370 (5.722-6.999)
LinearModel_Classifier	0.968 (0.949-0.984)	7.009 (6.351-7.642)
LinearModel_Regressor	0.984 (0.970-0.996)	6.351 (5.704-7.008)
wCVR.100	0.984 (0.972-0.995)	6.206 (5.597-6.824)
wCVR.101	0.984 (0.972-0.995)	6.205 (5.601-6.823)
wCVR.001	0.982 (0.968-0.993)	7.730 (6.898-8.606)
shortlong	0.857 (0.817-0.893)	7.706 (6.972-8.446)
theta	0.993 (0.984-1.000)	7.049 (6.400-7.703)
wtheta.100	0.986 (0.975-0.996)	6.051 (5.461-6.672)
wtheta.101	0.986 (0.975-0.996)	6.053 (5.461-6.678)
wtheta.111	0.986 (0.975-0.996)	6.045 (5.450-6.670)

A β 42 represent the major contributing factors; however, the inclusion of additional peptides further improves the model. Notably, the weightings were distinct when optimising for classification versus onset prediction, suggesting differential relevance of A β peptides. Finally, we observed that data from iPSC models show greater weighting for A β 38 and less for A β 40 compared with data from PSEN1/2 double knockout cell lines [173, 172, 171], though the sample size for the iPSC data was small (3 controls. 10 *PSEN1*-mutations).

Arber [175] comments that employing a patient-derived iPSC-neuronal model, this study complements previous studies in PSEN1/PSEN2 double-knockout cell lines due to the presence of one healthy PSEN1 allele and two healthy PSEN2 alleles, analogous to the patient setting. This model therefore represents a physiological model of A β production in human neurons, which Arber et al. have previously shown to faithfully correlate with clinical plasma data from the same donor [90, 176]. As such, this model of A β production is relevant for current efforts in developing blood

plasma biomarkers.

Previously reported ratios performed well: A β 37:42 [172] distinguished patient and control lines better than other unweighted ratios and A β short/long [171] performed best in terms of explaining age-at-onset variance (higher R^2). However, weighted composite value ratios (wCVR) were able to outperform linear ratios in both metrics, suggesting that increasing the number of A β peptides and assigning weight to each is able to further improve biomarker performance. The optimal ratio for all data, centred around both disease classification and explaining age-at-onset variance was:

$wCVR = (21 \cdot A\beta37 + 10 \cdot A\beta38 + 69 \cdot A\beta40) / (94 \cdot A\beta42 + 6 \cdot A\beta43)$. The numerator and denominator are precisely the same peptides of Petit et al. [171], with varying weights on the peptides.

To test whether the superior performance of wCVR reflected overfitting rather than genuine predictive value, we compared it under cross-validation with a range of alternative biomarkers. This revealed a task-dependent hierarchy: simple magnitude-based metrics such as unweighted theta excelled at distinguishing mutation carriers from controls, whereas predicting age at onset required more complex, weighted models. In the latter case, wCVR and weighted theta outperformed both simple ratios and standard machine-learning regressors, indicating that their gains stem from embedding biologically informed weighting rather than fitting noise. These results support the robustness of wCVR and reinforce that biomarker choice should be matched to the clinical question—simple interpretable metrics for diagnostic screening, and domain-optimised weighted approaches for prognostic modelling.

A diversity of A β peptides exists, and *PSEN1*-mutations alter their relative abundance differently. We analysed wCVR to infer the relative contribution of each peptide to classifying disease and for predicting age at onset. The fact that weights differ between the two biomarker analyses indicates that pathogenicity and severity are somewhat distinct; for example, changes to A β 38 can aid in predicting onset but are negatively associated with classification. Additionally, we observed differences in the wCVR in iPSC-models compared to cell models, such that A β 38 has a higher weighting at the expense of A β 40. Arber [175] comments that this may inform on the presence of a healthy *PSEN1* allele in iPSC models, speaking to the

overrepresentation of A β 40 compared with other peptides.

There is variance in the onset of familial AD, even within families with the same mutation. Our analyses found no association between the location of the mutation across the PSEN1 protein and age-at-onset, similar to previous recent observations (2024) [173]. Arber [175] comments that these data suggest that the severity of each mutation depends upon the biochemical effect of each amino acid substitution on the tertiary/quaternary protein structure.

This study comes with some limitations to consider. One limitation is the low sample sizes. More donors and data replicates will further improve the confidence in these findings, for which replication studies are required. Arber [175] comments that another limitation is that the panel of iPSC samples tested comprise an overrepresentation of mutations that produce relatively high levels of A β 43 (R278I [90], E280G [177] and P436S [178]), potentially skewing the relative contribution of A β 43. However, our analysis of other datasets which do not have the same issue mitigates this limitation. Finally, age at onset is known to be variable, and so the regression analyses should be considered with this caveat in mind. Future analyses could consider using errors-in-variables models.

In conclusion, this study presents new data from collaborators, a new fluid biomarker harmonisation method, and a new ratio-based biomarker. Our experimental results demonstrate superior performance compared to the previous state-of-the-art ratio biomarkers, and also offer insights into disease biology/mechanisms. Now that plasma biomarkers have entered clinical use [89], our findings are more relevant than ever. The proposed framework for data harmonisation and multi-peptide biomarker construction enhances diagnostic precision, particularly in complex cases such as early-onset familial Alzheimer's disease. This work provides a critical refinement to support the translation of first-generation tools into robust, next-generation clinical diagnostics.

Chapter 7

Theta, a multidimensional-input, one-dimensional output ratio-based biomarker

The work described in this chapter was presented at the 2024 AAIC conference, with a poster. An updated preprint is available at medRxiv [174], and R and Python code are available on GitHub (github.com/isaac-6/theta). In this Chapter, we consolidate the idea of a multidimensional ratio as a biomarker.

7.1 Introduction

As we saw in Chapter 6, different ratio biomarkers have been proposed in the literature, using different amyloid peptides, for AD classification. In the particular case of familial Alzheimer's disease classification versus controls, the current ratios proposed are the canonical $A\beta_{42/40}$ [179] and the recent $A\beta_{37/42}$ [172]. If they are *different*, and *useful*, then perhaps we can use a combination of them. Inspired by the combined reference formulation derived in Chapter 3 (Section 3.4), we plan to unify all available peptides into a single ratio-based biomarker, and evaluate its performance against the other biomarkers from the literature.

7.2 Material and Methods

Data. Data used for the analysis in this chapter is the same as in Chapter 6, and consists of measurements of $A\beta$ peptides from iPSC and cell cultures. Data harmonisation was justified and solved in Chapter 6, so we base the subsequent analysis on that. Here, we correct the pathogenic classification for two mutations: E318G (benign

[180, 181]), E69D (benign [180, 182]). Note that this was done retrospectively, after finding that their theta value was very close to the controls, and inquiring about them (see Supplementary Figure D.1). The following analysis considers them controls. This finding suggests that theta may have potential applicability in determining the pathogenicity of novel mutations.

In Chapter 3, Section 3.4, we showed how a two-dimensional ratio can be expressed as a function of the angle α between two vectors corresponding to two measures, and how this can work in multiple dimensions. Here, we move forward towards the application of the biomarker. The ratio-based biomarkers here are used for classification, comparing their values of *PSEN1*-mutation carriers against the controls.

Theta Each data sample includes five peptide measures, which can be represented as a point in a five-dimensional space. Simple ratio biomarkers, such as A β 42/40 and A β 37/42, work in the projection of the multidimensional space onto a plane defined by the two selected peptides (see Figure 7.1). The peptide ratio is the tangent of the angle α formed between the vector from the origin to the data point (b) and the direction of the denominator peptide dimension (see Figure 7.1). Since peptide measures cannot be negative, all possible data points lie within the positive orthant, and therefore, this angle α is a monotonic function of the ratio, and likewise for its complementary angle. After all peptide measures are divided by the average of the controls in each dataset, all controls are centred around the direction defined by the vector-of-ones $\vec{j}$. We can then take this direction as a reference, and calculate the angle θ with respect to any other origin-datapoint vector $\vec{b}$ (Figure 7.1). The angle θ is given by the following equation: $\theta = \arccos\left(\frac{\vec{j} \cdot \vec{b}}{\|\vec{j}\| \|\vec{b}\|}\right)$, where the numerator is the dot product of vectors $\vec{j}$ and $\vec{b}$, and the denominator is the multiplication of the magnitudes (norms) of vectors j and b , respectively. Note that due to the data harmonisation, this method does not require any fitting of parameters, and can be used in any independent dataset.

Benchmarks. We compare the performance of theta to A β 42/40 [179, 183], A β 37/42 [172], and “short/long” = (A β 37+38+40)/(A β 42+43) [171].

Evaluation metrics. As argued in Chapter 6, due to the heavy class imbalance of the data (relatively few controls), we employ the precision-recall curve (PR AUC)

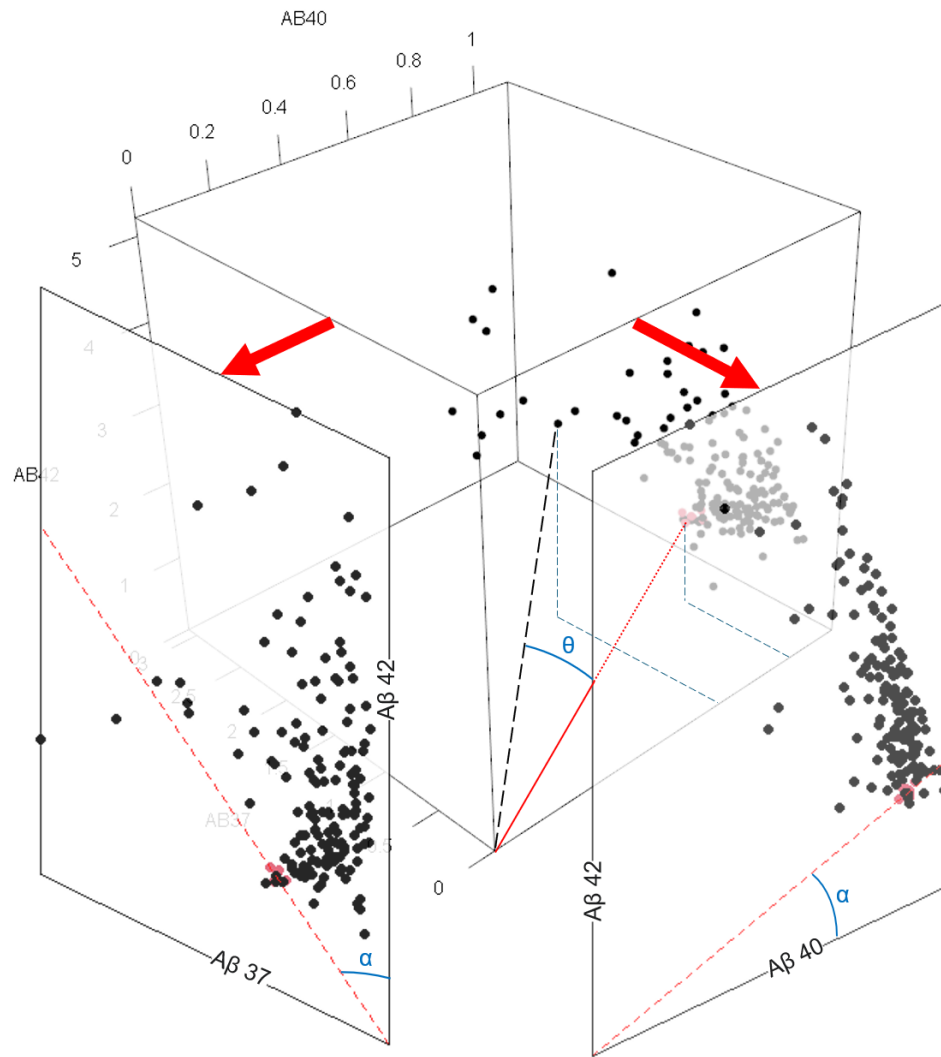


Figure 7.1: Distribution of the three merged cell datasets in three dimensions, corresponding to A β 37, 40 and 42. Red dots represent controls. The simple ratio peptides like A β 42/40 and A β 37/42 are equal to the tangent of the angle α on the plane they project the data to. Theta (θ) is the angle between the direction defined by the vector-of-ones (red) and any origin-datapoint vector.

of the smaller group, which plots control *precision* $TN/(TN+FN)$ against control *recall* $TN/(TN+FP)$, where T/F indicates True/False and P/N indicates Positive/Negative. 95% confidence intervals are provided using 2000 bootstrapped samples stratified by class (controls or mutation-carriers).

7.3 Results

We evaluated the classification performance of our new approach (theta) and benchmarks on multiple datasets independently, as well as merged datasets (all cell-based and all data). Table 7.1 shows the results in terms of 1) the area under the receiver operating characteristic curve (AUC) and 2) area under the precision–recall curve (PRAUC) of the controls, with 95% confidence intervals (CI). Across all datasets, theta consistently yielded excellent discrimination, with AUC and PR AUC values of 1.0 (perfect classification), even the lowest bound of the confidence interval. As expected by the class imbalance, all benchmark biomarkers scored high in AUC (above 90%), but the performance plummeted when evaluating PR AUC in some of the datasets, and when merging all data (less than 0.48 when looking at all cell data).

Figure 7.2 shows the controls precision-recall curves, using all datasets merged, for the classical A β 42/40, the best biomarker for classification from the literature (A β 37/42), and theta. It's easy to see that theta has a perfect classification, as it fills up the entire plot.

7.4 Discussion

In summary, we devised an elegant and efficient mathematical model as a multidimensional ratio to investigate combinations of A β 37, A β 38, A β 40, A β 42 and A β 43 peptides and their use as biomarkers in early-onset familial Alzheimer's Disease. The efficiency stems from its parameter-free nature and ability to synthesise information from multiple inputs into a single, interpretable metric.

In all datasets independently, and also after merging all data, theta outperforms other ratios, both for the AUC (true positive rate versus false positive rate) and the PR AUC (precision versus recall), with perfect classification. Although biomarkers from the literature all achieve high classification scores in the receiver operating characteristic curve (AUC), they use different biological information (as in, they are

Table 7.1: Classification metrics per dataset, independently and merged. In parentheses, 95% confidence intervals by bootstrapping. AUC = area under the receiver operating characteristic curve. PR AUC = area under the precision–recall curve of the controls.

Data	Biomarker	AUC	PR AUC
UCL	42/40	1 (1, 1)	1 (1, 1)
UCL	37/42	1 (1, 1)	1 (1, 1)
UCL	(37+38+40)/(42+43)	1 (1, 1)	1 (1, 1)
UCL	theta	1 (1, 1)	1 (1, 1)
Liu	42/40	0.956 (0.912, 0.986)	0.311 (0.177, 0.676)
Liu	37/42	0.986 (0.958, 1)	0.588 (0.317, 1)
Liu	(37+38+40)/(42+43)	0.969 (0.932, 0.993)	0.377 (0.221, 0.865)
Liu	theta	1 (1, 1)	1 (1, 1)
Petit	42/40	1 (1, 1)	1 (1, 1)
Petit	37/42	0.905 (0.762, 1)	0.307 (0.159, 1)
Petit	(37+38+40)/(42+43)	1 (1, 1)	1 (1, 1)
Petit	theta	1 (1, 1)	1 (1, 1)
Schultz	42/40	0.964 (0.911, 1)	0.189 (0.088, 1)
Schultz	37/42	1 (1, 1)	1 (1, 1)
Schultz	(37+38+40)/(42+43)	1 (1, 1)	1 (1, 1)
Schultz	theta	1 (1, 1)	1 (1, 1)
Cell data	42/40	0.968 (0.939, 0.988)	0.372 (0.233, 0.741)
Cell data	37/42	0.981 (0.958, 0.996)	0.478 (0.3, 0.911)
Cell data	(37+38+40)/(42+43)	0.981 (0.959, 0.997)	0.475 (0.303, 0.925)
Cell data	theta	1 (1, 1)	1 (1, 1)
All data	42/40	0.969 (0.943, 0.988)	0.437 (0.295, 0.713)
All data	37/42	0.982 (0.961, 0.996)	0.544 (0.368, 0.924)
All data	(37+38+40)/(42+43)	0.982 (0.961, 0.996)	0.541 (0.371, 0.912)
All data	theta	1 (1, 1)	1 (1, 1)

different peptides). By expanding the ratio concept to a multidimensional space, our biomarker theta can make use of deviations of multiple peptides. Evaluating the precision-recall curve for the minority class (controls), the performance of the biomarkers from the literature plummeted, whereas theta achieved a perfect PR AUC score of 1, after merging all data. This is especially important, since the prevalence of EOFAD in the world population is so low that a low performance detecting non-mutation carriers would render the biomarker impractical for clinical usage, as it would translate into mostly false positives.

This study comes with some limitations to consider, and shares the limitations reported in Chapter 6 regarding the data. A limitation with the new ratio-based biomarker theta is that it is equally sensitive to deviations in any peptide. While our

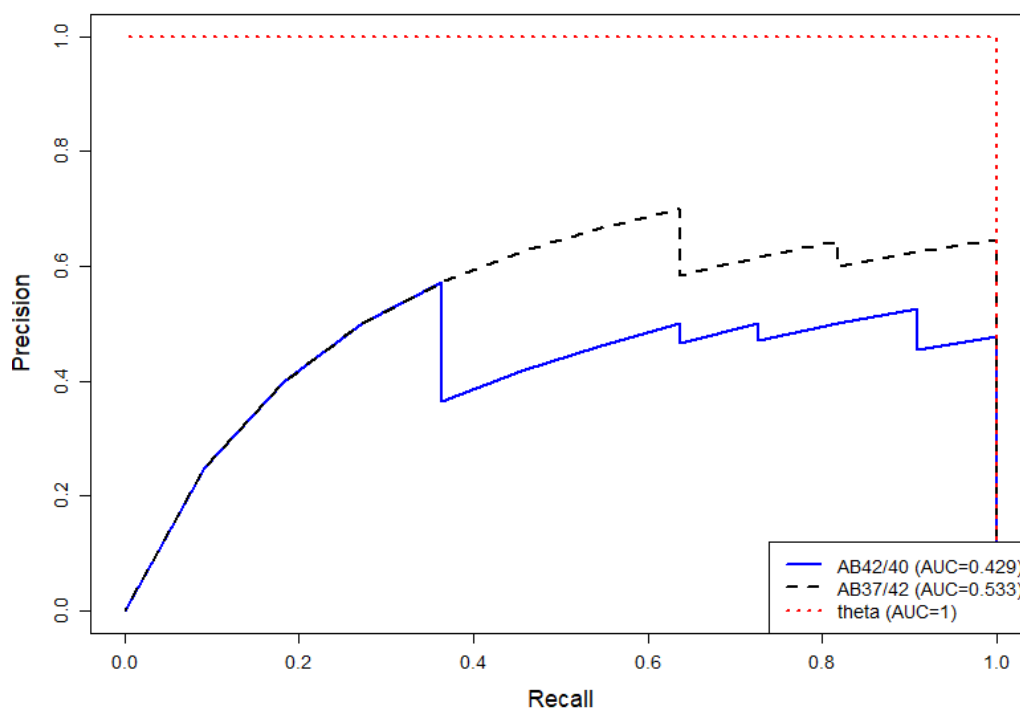


Figure 7.2: Precision-recall (of controls) curves for biomarkers A β 42/40 (solid blue line), A β 37/42 (dashed black line), and theta (dotted red line). The performance of each biomarker is evaluated through the precision-recall area under the curve (AUC) metric, with values of 0.429, 0.533, and 1.00, respectively.

experimental results suggest that theta is a far superior biomarker for classification, future work could take a data-driven approach to exploring the contribution of the different peptides, and thus propose a weighted-theta, or theta using a subset of peptides. On that end, theta is adaptable, as it can work with any number of peptides or inputs, provided they are first normalised. Update from August 2025: weighted theta has been implemented (by adapting the BioDisCVR framework) and tested in a leave-one-out cross-validation setup (see previous Chapter's Table 6.5 of out-of-fold performance), and theta leads the classification performance at 0.993 (95%: 0.984-1.000) F1-score, while weighted theta follows closely in classification, and leads the regression performance with a mean average error of 6.045 (95%: 5.450-6.670) years.

7.5 Conclusions

In conclusion, this study presents a new ratio-based biomarker, called theta. Our experimental results demonstrate superior performance to the previous state-of-

the-art ratio biomarkers. Specifically, theta achieves perfect classification of familial Alzheimer's disease mutations versus controls across multiple independent datasets and in merged data, a level of performance unmatched by existing biomarkers, particularly when considering precision-recall metrics crucial for rare disease applications. This exceptional performance highlights the power of a multidimensional, data-driven approach to biomarker design, effectively leveraging the combined information from multiple amyloid-beta peptides. The parameter-free nature of theta also offers advantages in terms of robustness and generalisability, simplifying its potential clinical translation. Theta's ability to virtually eliminate false positives in our evaluation suggests it could be a transformative tool for accurate identification of familial AD risk and for research focused on early detection and targeted interventions in this devastating disease.

Chapter 8

Discussion

This chapter critically evaluates the findings presented in this thesis. It starts with an in-depth interpretation of their significance, examines their broader implications for Alzheimer's disease biomarker research and clinical practice, and explicitly addresses the methodological limitations identified. The chapter concludes with a concise summary of the thesis's main contributions and their implications.

8.1 Chapter 3: Reference regions in SUVR

The investigation into the foundational SUVR methodology in tau PET imaging, as detailed in Chapter 3, unveiled a key and concerning discovery: a lack of correlation between conventionally accepted reference regions, specifically the inferior cerebellar grey matter and eroded subcortical white matter. This finding fundamentally challenges the assumption of their interchangeability in tau PET quantification and demands a deeper inquiry into the underlying biological mechanisms that might influence tracer kinetics or signal-to-noise ratios in these regions. The inconsistency highlights a critical vulnerability in standard SUVR calculations, potentially leading to unreliable biomarker measures.

In response to these identified inconsistencies, the chapter proceeded by rigorously examining whether a composition of regions could improve the expected longitudinal stability of tau PET measurements (Section 3.2). This analysis, detailed in Section 3.3, evaluated a published composite reference region against commonly used single reference regions. The results corroborated other studies [78] in highlighting the superior performance of a composite reference region (encompassing the whole cerebellum, eroded subcortical white matter, and brainstem), a concept initially proposed for amyloid PET [124]. A crucial distinction of our work from

Schwarz et al. [78] lies in our specific assessment: we focused on longitudinal consistency within individuals and regions expected to be pathology-free, rather than evaluating longitudinal increase of tau signal. This approach provided a more direct measure of reference region stability. Furthermore, this work reinforced the counsel of Young et al. [123] (published later, in November 2021), who also advised against the inferior cerebellar grey matter for longitudinal ^{18}F -AV1451 PET studies in favour of eroded subcortical white matter. The ongoing relevance of these findings is evident in numerous recent studies that continue to employ the inferior cerebellar grey matter, despite the accumulating evidence for the quantified superiority of composite reference regions. To further expand the knowledge on this area, actual test-retest data is invaluable to help advance the understanding of radiotracer measurement in non-pathological regions, and likely, across all stages of the disease.

The finding of the clear disparity in regional longitudinal behaviour (measured SUVR vs SUV of another reference region) is another key point that would require further investigation. Other previous papers note the regional differences in the measured signal[184]. However, we noted a difference beyond a shift of SUVR values, since the gradients were distinctively different as well. While the correlation of SUVR factors of paired-scans with an unused reference region SUV suggests that combining them can correct for that longitudinal disparity, it still does not address the diverse regional needs. Again, test-retest data would provide a better ground-truth to base an analysis on.

Having concluded that the inclusion of multiple reference regions can benefit the stability of the measurements, we conceive a mathematical formulation to expand the simple ratio (division of two factors) to include multiple references as directions in a high-dimensional space. This is later applied in Chapter 7, with excellent classification results. This formulation maintains the simplicity and concept of the simple ratio, allowing multiple references to influence how they transform the numerator. At this point, this is purely theoretical, and its utility depends on the regions selected. A question about optimal reference region selection arises, which leads to Chapter 4.

We examine the effects of the PVC method proposed by ADNI (using the Geometric Transfer Matrix approach), and find out that it severely worsens all

test-retest metrics (of 18F-AV-1451 tau PET measurements), for all considered biomarkers. A limitation of the study is the experimental design, where we look at the composite regions that make up the Braak stages 5 and 6, for individuals with no or mild cognition impairment. A finer study that sieves through specific regions for specific individuals that makes sure there is no tau accumulation could further validate the findings. In December 2021, Groot et al. [121] commented that “no consensus has yet been reached on whether to use partial-volume correction of tau PET data.”, while a more recent paper[185] (2024) using the same data as we did (from ADNI) comments on the utility of PVC for tau PET, stating “PVC showed very limited benefits in increasing statistical power in group-level-style analysis of longitudinal tau PET metrics”. A limitation of their analysis, however, is that they evaluated group differences, which I think can dilute biomarker performance comparison: all individuals that are away from the cut-point will likely be easily separable by most biomarkers/pipelines, so the notable difference is when studying individuals close to the classification threshold. It should be noted that a limitation of not using PVC for 18F-AV-1451 tau PET data is that the hippocampus becomes unusable due to off-target signal bleeding from the choroid plexus. In this case, a specific subregion of the hippocampus could be used.

8.2 Chapter 4: Composite Value Ratio

The previous chapter focused on the reference regions of SUVR. Here, we approached the biomarker evaluation in terms of statistical power; that is, we want to find useful (tau-PET) signal, and evaluate the biomarkers with multiple metrics. A further notable leap with respect to the previous chapter, is that the data-driven search for an optimal composition of regions of interest is that it is applied to both the numerator and denominator of the ratio. A key difference with other related literature work is that in SUVR, or PERSI (an individualised subset of white matter voxels) is that the proposed composite value ratio (CVR) can adapt to different stages of the disease to maximise its performance, making use of regions that, while prohibitive in the traditional way, can be used as part of a reference because they contribute to a better biomarker (smaller model residuals, better group separation, smaller sample size needed to detect gradient changes). The results show a clear superiority of the new data-driven biomarkers, which can help anyone who uses tau PET data,

be it for research, clinical monitoring, or clinical trials. On that end, CVR can have impactful implications for clinical trials, potentially leading to substantial reductions in sample size (over 79% reduction in our experiments), faster trial completion, and reduced costs, ultimately accelerating the development of effective AD therapies. The ablation study shows an interesting picture: the biomarker composition is flexible, with little consequences to its performance metrics. My thoughts on this is that as tau appears, multiple regions are eventually affected, and so there is collinearity between different regions. This poses no problem, although future work could consider this, perhaps with the mathematical model proposed in Section 3.4. I am currently involved in a similar study, but involving multiple tracers, the intersection of which is more likely to give us biological insights on tau deposition. Future work on tau PET could incorporate amyloid as well, as it has been shown to exist a synergy between both pathologies [186].

8.3 Chapter 5: CVR applied to MRI

This chapter extended the application of CVR to MRI volumetric data, achieving similarly outstanding improvements in biomarker performance for tracking volumetric changes associated with AD. The success of CVR across both PET and MRI modalities highlights the versatility and robustness of the BioDisCVR framework in leveraging multi-regional information to create more powerful biomarkers. While the anti-amyloid clinical trials have shown signs of pseudo-atrophy (reduction in brain volume not attributed to a decline in neuronal count), 1) they still need biomarkers to detect changes, and 2) not all clinical trials for AD are anti-amyloid. Future work could include a collaboration with a neurologist, where BioDisCVR could be limited to certain regions as targets, providing a better proxy of the regional volume. A key advantage of CVR is that it takes into account the noise or variability. That is, while other approaches might be focused on just the target, all measurements are susceptible to errors or variability (e.g., some regions might be more challenging to segment), which prevents them from achieving optimal sensitivity or statistical power. BioDisCVR, by design, mitigates this by incorporating information from multiple regions in both the numerator and denominator and evaluating the biomarker as a whole.

Apart from discovering new biomarkers that are well-suited for monitoring

brain changes due to AD, this chapter highlights the bad performance of traditional biomarkers like the whole brain, or the hippocampus. While whole brain and hippocampal volumes remain clinically relevant measures, our findings underscore their limitations in terms of statistical power and precision for detecting subtle longitudinal changes, especially compared to the data-driven CVR-MRI biomarkers.

A limitation of the study is that results are inevitably tied to the data used, as is with all data-driven discoveries. In this case, even though the data came from over 60 different sites, they all used a similar processing pipeline for acquisition, and the regional values were extracted with a specific software (FreeSurfer 7.1.1). This means that if a specific region produced higher variance measurements, it would be discarded by BioDisCVR. While this is useful in practice (it discards noisy regions), it can limit the biological interpretation: perhaps a discarded region is very relevant to AD, but the high variance of its measurements prevents it from being useful (or as useful). Future work across datasets and processing pipelines could explore this, similar to Schwarz et al. [78].

8.4 Chapter 6: Weighted CVR

In this chapter, we shift the focus to amyloid-beta peptides ($A\beta$), another hallmark of AD pathology. BioDisCVR is used here with an added bonus: we allow each composite (numerator and denominator) to be a linear combination of features, obtaining a weighted-CVR (wCVR). This allowed to reveal differential contributions of various $A\beta$ peptides to AD pathogenesis and demonstrated the best classification (best wCVR achieved 0.700 precision-recall AUC, compared to 0.454 from $A\beta_{37/42}$) and explained variance with age-at-onset ($0.412 R^2$, with a close 0.403 from $A\beta_{(37+38+40)/(42+43)}$) compared to both traditional and published ratios, and compared to all possible non-weighted peptide combinations.

There are multiple takeaways from this chapter. First, data harmonisation is imperative in order to mix data from different sites/assays, and we propose a simple, but effective method that does not require external data or training of a model; as controls are expected to be homogeneous [90], we just use them as reference. Second, different peptides contribute different information, so there's a performance benefit from incorporating multiple peptides (as with Petit et al. short/long peptide ratio biomarker [171]), and from allowing their contributions to be weighted. From

the analysed data, we see that most of the mutations diminish the proportion of A β 37,38,40 with respect to the controls, while augmenting the proportion of A β 42 and A β 43, which coincides with the proposed biomarker from Petit et al. [171]. Supplementary Figure C.5 and Figure 6.6 support the robustness of the findings (within the used data), showing the peptide weights optimised for each metric are a global, not local, optimal point. This convergence with the findings of Petit et al. [171] strengthens the biological plausibility of our wCVR biomarker and highlights the importance of considering the full spectrum of A β peptides in AD biomarker development, possibly gaining mechanistic insights.

A limitation, in practice, might be the added cost of extra analysis. Indeed, future research could explore simplified versions of wCVR or strategies to reduce assay costs, by incorporating the cost in the fitness function of BioDisCVR. Another limitation that complicates the understanding of mutations is their heterogeneity: we are not comparing controls to a single group, but to multiple specific groups (at least one per mutation). Future work, with additional data, could look into mutation-specific patterns, instead of considering them all a single group.

8.5 Chapter 7: Theta, a multidimensional ratio

While the previous chapter (wCVR for A β peptides, Chapter 6) argued that multiple peptides offer complementary information, here we formalised and tested the theoretical multidimensional ratio theorised in Chapter 3 for classification. The results are outstanding: theta has a precision-recall area-under-the-curve of 1.00 (95% CI 1.00, 1.00), which is a perfect classification. The next best biomarker is A β 37/42, with 0.544 (0.368, 0.924). At the beginning of the study, we identified four mutations that were very close to the controls, in their theta values. Further investigation revealed two to be benign, while the remaining two had unknown pathogenicity. One advantage of theta, as presented in this work, is that it is not a model with trainable parameters; theta is just a representation of multidimensional data, summarised in one angle. This parameter-free nature of theta makes it inherently robust and less susceptible to overfitting, enhancing its potential for generalisability across different datasets and cohorts. Nevertheless, future work could explore a parametric theta (as briefly introduced in an additional analysis in Chapter 6), where the implication of each feature (peptides in our case) is weighted for mechanistic insight. This

could also bridge the gap between the non-parametric theta and the weighted CVR approach, combining the strengths of both.

8.6 Synthesis of thesis Contributions and their Implications

Having interpreted the specific findings of each chapter in the context of the existing literature and addressed their respective limitations, this section provides a lightweight, summarised overview of the principal contributions of this thesis. Each contribution is presented alongside its direct implications for the field of Alzheimer's disease research, clinical trials, and biomarker development, thereby summarising the core arguments of this discussion.

- **On the Foundations of Tau PET Quantification (Chapter 3):**
 - **Contribution:** Exposed the critical inconsistency of conventional SUVR reference regions and the detrimental impact of the GTM partial-volume correction method on measurement stability. A novel mathematical framework for a multi-reference ratio was theorised.
 - **Implication:** This work provides quantitative evidence that calls for a fundamental reconsideration of standard tau PET quantification pipelines. It urges the field to move towards more robust, composite reference regions and exercise caution with PVC (GTM in AV-1451), directly impacting the validity of data from ongoing and future clinical trials.
- **On Data-Driven Biomarker Discovery for Neuroimaging (Chapters 4 & 5):**
 - **Contribution:** Developed the BioDisCVR framework and validated the novel Composite Value Ratio (CVR) biomarker across both tau PET and structural MRI data.
 - **Implication:** CVR is established as a modality-agnostic biomarker that offers transformative gains in statistical power, precision, and sensitivity over traditional methods. Its adoption can lead to smaller, faster, and more cost-effective clinical trials, accelerating the development of effective AD therapies, and they can also improve monitoring of pathological changes.

- **On the Analysis of Amyloid-Beta Peptides (Chapter 6):**

- **Contribution:** Introduced the weighted CVR (wCVR) to reveal the differential diagnostic and prognostic contributions of the full spectrum of A β peptides. A simple, robust method for harmonising multi-assay data was also proposed.
- **Implication:** This demonstrates that moving beyond the traditional A β 42/40 ratio to a weighted, multi-peptide signature can yield superior biomarkers. This data-driven approach can uncover novel biological insights into peptide processing in familial AD. Furthermore, by incorporating a weighted composition in CVR, the algorithm can adapt to different feature importances.

- **On a New Paradigm for Ratio-Based Biomarkers (Chapter 7):**

- **Contribution:** Formalised and validated theta, a parameter-free, multi-dimensional ratio capable of integrating multiple features into a single, highly sensitive biomarker.
- **Implication:** Theta achieved state-of-the-art classification of pathogenic mutations, proving its potential as a powerful tool for summarising complex, multi-feature data. As a generalisation of the widely accepted ratio concept, theta offers a readily interpretable yet powerful new approach for fluid biomarker development in AD and beyond.

Taken together, the findings presented in this thesis support a shift toward a data-driven and quantitatively rigorous approach to biomarker development. Beyond introducing novel biomarkers, this work lays the groundwork for a scalable framework that can guide future discovery efforts. These insights set the stage for the concluding chapter, which reflects on the broader implications of this research and outlines directions for clinical translation and future investigation.

Chapter 9

Conclusions

In conclusion, this PhD thesis has demonstrated that the current understanding and utilisation of (PET and MRI) brain scans in Alzheimer's disease can be vastly enhanced through the application of data-driven methodologies. By exposing the limitations of traditional SUVR and introducing the data-driven BioDisCVR framework and its associated CVR biomarker, as well as theta, this work offers a powerful new toolkit for the field. The composite value ratio, validated across multiple modalities, provides a more robust and longitudinally consistent measure of pathology, with enhanced statistical power. Theta, on the other hand, is able to incorporate multiple features (iPSC and cell-based amyloid-beta peptides, in our case study) and is sensitive to deviations in any of them, achieving a perfect classification (controls versus *PSEN1* mutations) in all datasets evaluated. I believe that theta should be readily acceptable by the field, since it is essentially a generalisation of the ratio biomarker concept: a concept that is widely accepted in the field, but has not been leveraged in such a way to include more than two features.

These new conceptual biomarkers of CVR and theta pave the way for the development of biomarkers that are not only statistically superior but also offer the potential to advance our fundamental understanding of Alzheimer's disease, ultimately contributing to more effective diagnosis, faster/cheaper clinical trials, treatment, and hopefully prevention of this devastating pandemic-like disease in the era of precision medicine and technological advances. As mentioned in the introduction, to change the trajectory of the disease, we must first understand its path, and these models and biomarkers are bridges between the chaos of data and the clarity of understanding.

Appendix A

BioDisCVR Supplementary Material

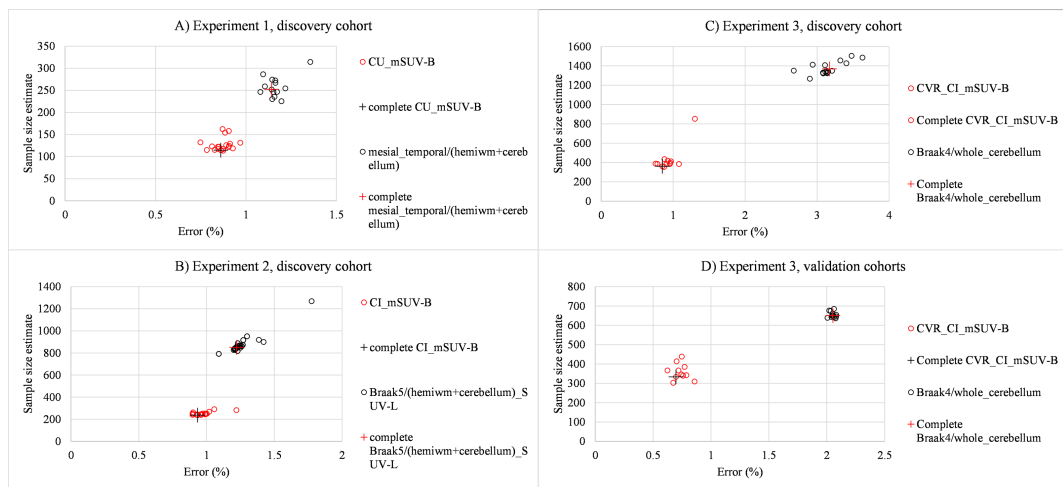


Figure A.1: Repeatability (error) and SSE performance of selected biomarkers, removing one region at a time, for cognitively unimpaired (A) and cognitively impaired (B). Panels C and D show the performance of the discovery and validation sets, respectively, using the brain regions that were available in both datasets. The crosses represent the performance of the full, original biomarker. The vast worsening of SSE meta-temp/composite is due to removing the eroded subcortical white matter region from the denominator, which is also present in both CVR biomarkers. A similar observation is seen when the inferior temporal gyrus is removed from the numerator in (C) CVR_mSUV-B, highlighting its importance in the biomarker. Abbreviations: CVR = composite value ratio, as in “the ratio of two composite regions”. SUV = standardised uptake value, the signal measured in a positron emission tomography scan over a volume. The prefix “m” in SUV indicates that it is the mean SUV over a number of regions; otherwise, it is the volume-weighted SUV. The suffix -B after SUV indicates that the analysis was bilateral, as in “considering regions in both hemispheres”, as opposed to -L, which would indicate laterality. Figure reproduced from [131].

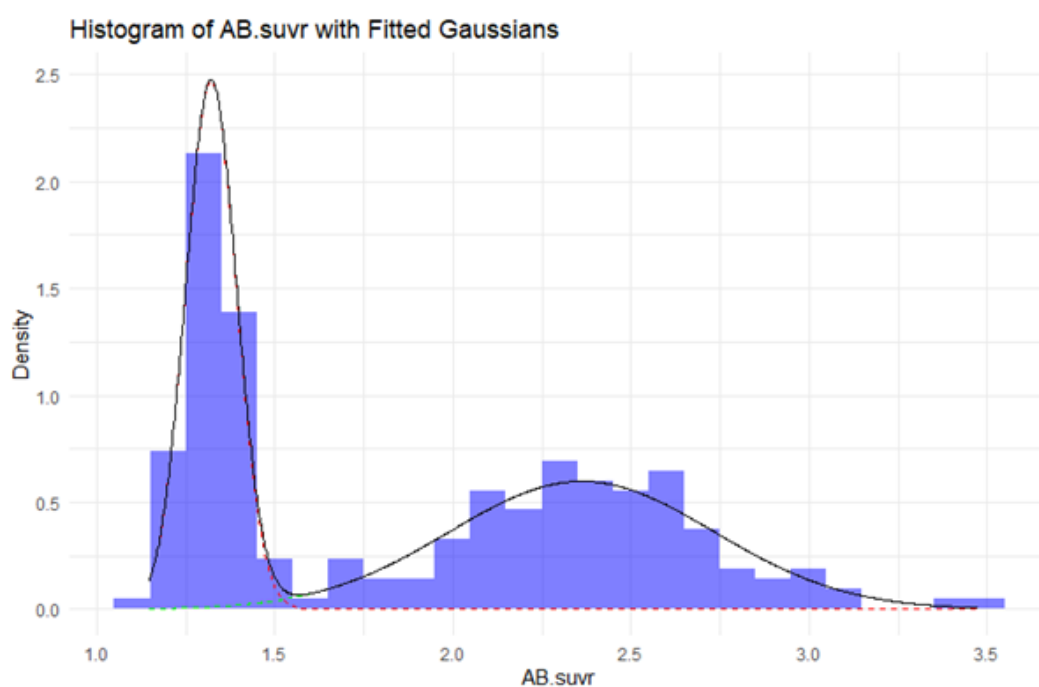


Figure A.2: Amyloid distribution of the validation set. The cutoff for amyloid positivity was defined as the intersection of the two Gaussians that lies between their means. SUVR is the standardised uptake value ratio. The reference was the inferior cerebellum grey matter, and the target was the cortical summary region from Lee J, Murphy A, Ward T, Harrison T, Landau S, Jagust W.: Amyloid PET Processing Methods, published online 2023 in the ADNI database. Figure reproduced from [131].

Table A.1: BioDisCVR fixing either the numerator or denominator as known literature ROIs. We have different designs here: we fix the numerator as the meta-temporal, and find a fitting denominator (CVR_r(ft.meta)). We do the same, but starting by fixing the denominator as the inferior cerebellum grey matter (igm). This configuration of BioDisCVR, considers bilateral regions (using both hemispheres jointly) and takes the mean standardised uptake values (SUV) of the regions. The two first rows show biomarkers from the literature, while the third row for each sub-table shows the best-performing biomarker out of 220 literature-inspired combinations of composite targets and references. We can observe the big impact of allowing our algorithm to find an appropriate data-driven composite to be used as a reference (denominator). Abbreviations: composite = average of whole cerebellum, brainstem and eroded subcortical white matter; DDS = adaptive individualised biomarker by Leuzy et al., 2023; ewm = eroded subcortical white matter; igm = inferior cerebellum grey matter; SSE = sample size estimate (80% power, 20% effect size, 4.5 years for Experiment 1, and 1.5 years for Experiment 2). Table reproduced from [131].

A) Experiment 1 (cognitively unimpaired), with fixed numerator or denominator				
Numerator	Denominator	SSE	Separation	Repeatability
meta_temporal	composite	626 (353, 1401)	4.37 (2.46, 6.49)	1.06 (0.82, 1.31)
DDS	igm	438 (259, 895)	3.48 (1.57, 5.7)	1.66 (1.32, 2.06)
mesial_temporal (mean)	(ewm+cerebellum)	206 (143, 325)	0.17 (-1.84, 2.09)	1.03 (0.82, 1.3)
meta_temporal	CVR_r(ft.meta)	207 (143, 329)	2.23 (0.25, 4.18)	1.16 (0.91, 1.43)
CVR_t(igm)	igm	321 (201, 591)	2.82 (0.81, 4.78)	1.46 (1.13, 1.79)

B) Experiment 2 (cognitively impaired), with fixed numerator or denominator				
Numerator	Denominator	SSE	Separation	Repeatability
meta_temporal	composite	1539 (1091, 2333)	4.37 (2.46, 6.49)	4.67 (3.97, 5.06)
DDS	igm	1031 (739, 1539)	3.48 (1.57, 5.7)	3.12 (2.63, 3.69)
Braak5	(ewm+cerebellum)	852 (569, 1415)	4.14 (2.3, 6.35)	1.22 (1.02, 1.46)
meta_temporal	CVR_r(ft.meta)	372 (290, 495)	3.8 (1.85, 5.87)	1.55 (1.32, 1.82)
CVR_t(igm)	igm	1052 (722, 1672)	4.49 (2.54, 6.64)	2.58 (2.16, 3.03)

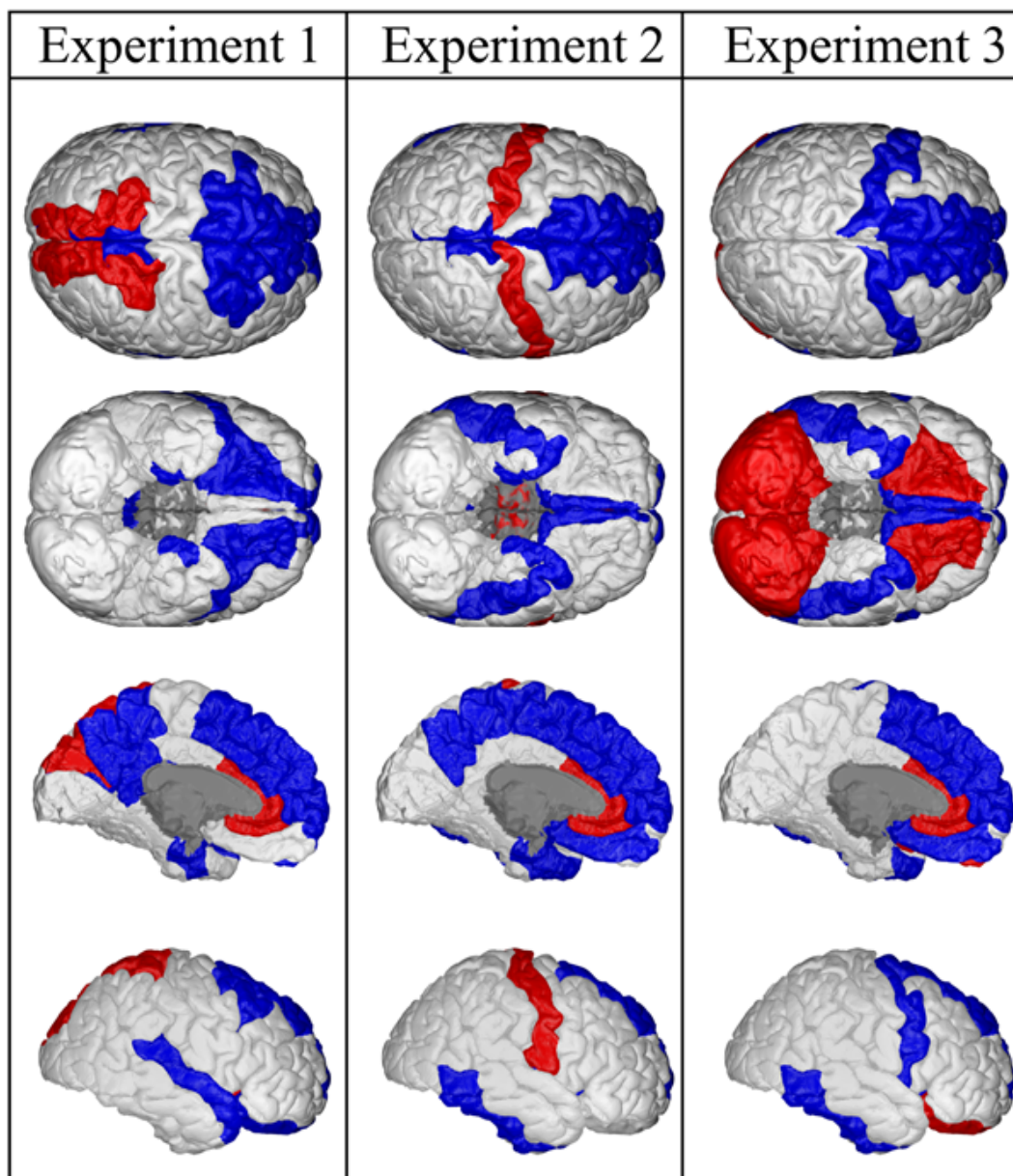


Figure A.3: Visualization of numerator (blue) and denominator (red) regions for our biomarker CVR-mSUV-B, for Experiments 1, 2 and 3. CVR stands for "composite value ratio". In this case, the composition is the mean of the regions. The suffix -B indicates that the design considered bilateral regions (joint left- and right-hemisphere). Figure reproduced from [131].

Code for BioDisCVR. Given the regional values and volumes, diagnosis and amyloid positivity, the code finds suitable composite value ratio (CVR) biomarkers, for preclinical and clinical trials.

```

1 # Script *example* for BioDisCVR, discovery of biomarkers.
2 # Author: Isaac Llorente Saguer, University College London
3 # For the paper in Brain Communications, by Llorente-Saguer and
  Oxtoby, 2024
4
5 ### load libraries
6 # Package manager
7 install.packages("pacman")           # Install pacman package
8 library("pacman")                   # Load pacman package
9
10 # Install/Load other packages
11 p_load(lme4, longpower, GA, dplyr, parallel)
12
13 ### Data preparation / load
14
15 ## load or create dataframe "data"
16 # It should have these columns:
17 # - SUVR: biomarker value
18 # - AB: Amyloid positivity (T/F)
19 # - time: time difference with respect to the average visit
  date, per individual
20 # - RID: unique individual ID
21 # - DX: Diagnostic. 0 = cognitively unimpaired. 1 = cognitively
  impaired (MCI or AD)
22 # data <- ...
23
24 ## load or create dataframe "data.suv" with the SUV or SUVR
25 # It should have a column per region, and the same matching
  visits (rows) as "data"
26
27 ## load or create dataframe "data.vol"
28 # Similar to data.suv, but with regional volumes
29

```

```

30 ## load or create dataframe "data.uv"
31 data.uv <- data.suv * data.uv
32
33 ## get list of possible regions
34 keep_regs <- names(data.suv)
35
36
37 ### functions
38
39 get_suvr <- function (chromosome) {
40   t = keep_regs[chromosome < 1]
41   r = keep_regs[chromosome > 2]
42
43   t <- t[!is.na(t)]
44   r <- r[!is.na(r)]
45
46   len_t <- length(t)
47   len_r <- length(r)
48
49   if ((len_t & len_r) == 0) {return (-99999)}
50
51   if (len_t == 1) {
52     tt = data.suv[,t]
53   } else {
54     if (var_composition == 0) {
55       # average weighted by another variable (e.g., volume)
56       tt = rowSums(data.uv[,t])/rowSums(data.vol[,t])
57     } else if (var_composition == 1) {
58       # simple average
59       tt = rowMeans(data.suv[,t])
60     }
61
62   }
63
64   if (len_r == 1) {
65     rr = data.suv[,r]

```

```

66   } else {
67
68     if (var_composition == 0) {
69       # average weighted by another variable (e.g., volume)
70       rr = rowSums(data.uv[,r])/rowSums(data.vol[,r])
71     } else if (var_composition == 1) {
72       # simple average
73       rr = rowMeans(data.suv[,r])
74     }
75   }
76
77   return(log(tt/rr))
78 }
79
80
81 # fit = t/(sse^2)
82 # correlated intercept and slopes
83 fitness_tss <- function(aux) {
84
85   # Get metrics from cohort 0 (ADNI in the example)
86   data <- data0
87   data$SUVR <- get_suvr(aux, 0)
88
89   # if SUVR failed for whatever reason, return a very bad
      fitness
90   if (all(data$SUVR == -99999)) {return(-99999)}
91
92   if (var_SSEgroup == "CI") {
93     # evaluating CI A+
94     aux.data <- data[data$DX == 1,];
95     aux.data <- aux.data[aux.data$AB == T,];
96     aux.data <- aux.data[!is.na(aux.data$AB),];
97     aux.time <- 1.5
98
99   } else if (var_SSEgroup == "CU") {
100     # evaluating only CU A+

```



```

122
123   eSep <- fSep(aux.model)
124
125   return(eSep/(eSiz^2))
126
127 }
128
129 ### discovery
130
131 ## Define variables
132 var_SSEgroup <- "CU" # "CU" for cognitively unimpaired, or "CI"
    for cognitively impaired
133 var_composition <- 1 # 0 is volume-weighted SUV, 1 is mean SUV
134
135 ## Run an exploration/optimization algorithm, such as this one
136 GA <- ga(type = "real-valued",
137         fitness = fitness_tss,
138         lower = c(rep(0,length(keep_regs))), upper = c(rep
            (2.5,length(keep_regs))),
139         lower = low_val, upper = high_val,
140         selection = "gareal_tourSelection",
141         crossover="gareal_blxCrossover",
142         mutation = "gareal_raMutation",
143         elitism = 2, # keep top 2 biomarkers of the previous
            generation
144         pmutation = 0.5, # mutation rate prob
145         popSize = 32, # the number of individuals (biomarkers
            per generation)
146         maxiter = 300, # total runs or generations
147         monitor = F, # F = nothing printed. T = print each
            generation's metrics.
148         parallel = T, # allow parallel processing
149         seed=42 # for reproducibility purposes
150 )
151
152 ## results

```

```
153  
154 chrom <- GA@solution[1,]  
155 t = keep_regs[chrom < 1]  
156 r = keep_regs[chrom > 2]  
157 cat(c("Numerator regions: \n",t))  
158 cat(c("Denominator regions: \n",r))  
159 cat(c("Biomarker fitness: ", fitness_tss(chrom)))
```

Listing A.1: Code for BioDisCVR

Appendix B

CVR in MRI for atrophy monitoring

Supplementary Material

Table B.1: Solanezumab treatment effect on the change of atrophy in the A4 trial cohort. Least-squares mean difference in the annualised percent change of volumetric biomarkers between the Solanezumab ($N = 401$) and Placebo ($N = 423$) treatment arms. A negative difference indicates less atrophy (treatment benefit) in the solanezumab group compared to placebo. All estimates are derived from Analysis of Covariance (ANCOVA) models adjusted for baseline age, sex (female vs. male), education (≥ 13 vs. < 13 years), ApoE $\epsilon 4$ carrier status, and the baseline value of the biomarker. Values in parentheses represent the 95% confidence interval for the difference. CVR.CU.x = Composite Value Ratio trained on the Cognitively Impaired Subset data (50%) from ADNI; ICV = Intracranial Volume.

Biomarker	Difference in % change	p-value
CVR.CU.1	-0.277 (-0.672, 0.118)	0.169
CVR.CU.2	-0.198 (-0.587, 0.191)	0.318
Ventricles/ICV	-0.132 (-0.504, 0.240)	0.486
WholeBrain/ICV	0.0311 (-0.058, 0.120)	0.494
Hippocampus/Ventricles	-0.0643 (-0.461, 0.332)	0.751
Hippocampus/ICV	0.0059 (-0.170, 0.182)	0.948

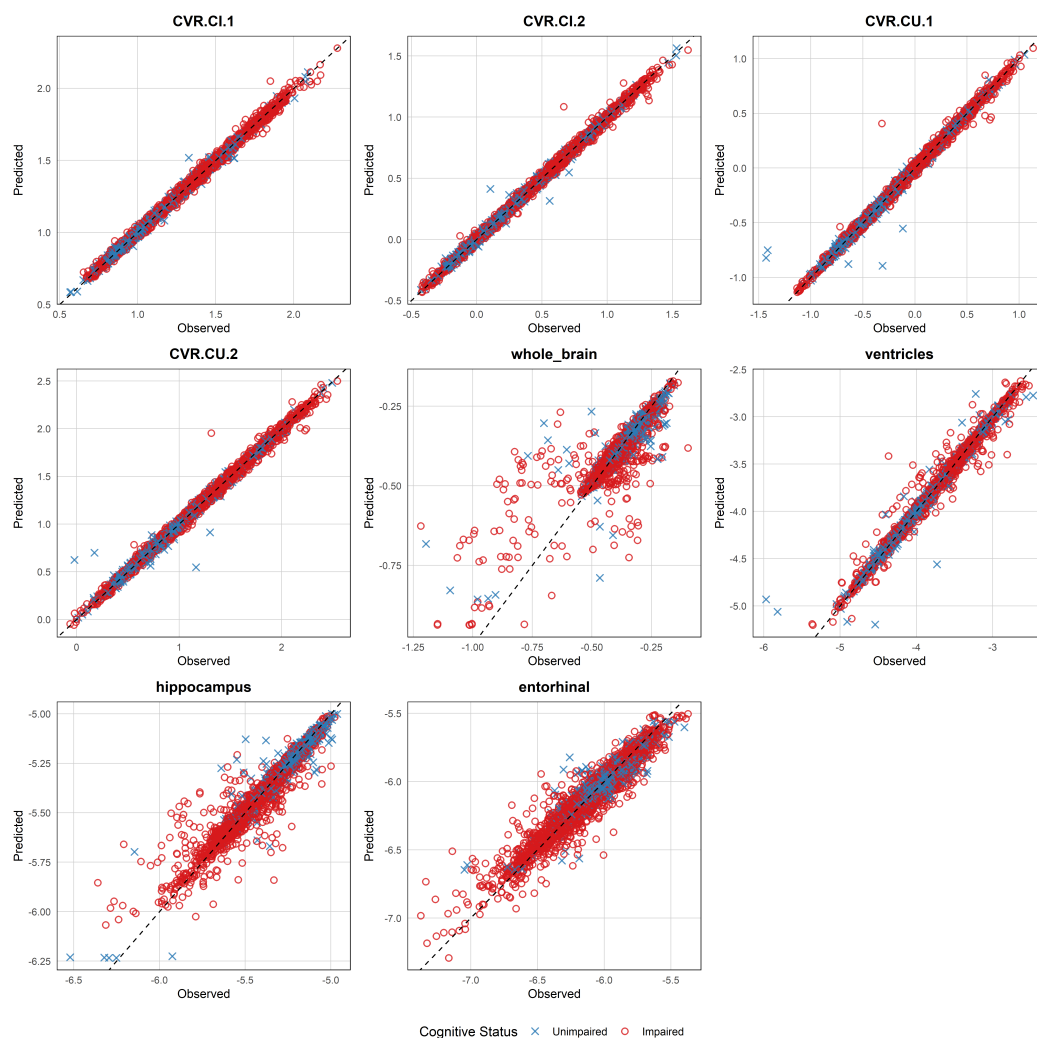


Figure B.1: Predicted vs observed values, for subset 1. Values are log-transformed. CVR is the composite value ratio of two data-driven regions of interest; the rest are divided by the intracranial volume. The CVR naming coding is as follows: CI = conditioned to minimise the sample size estimate of cognitively impaired individuals; CU = conditioned to minimise the sample size estimate of cognitively unimpaired individuals; suffix 1 or 2 indicate the subset of the data that was used for the CVR discovery.

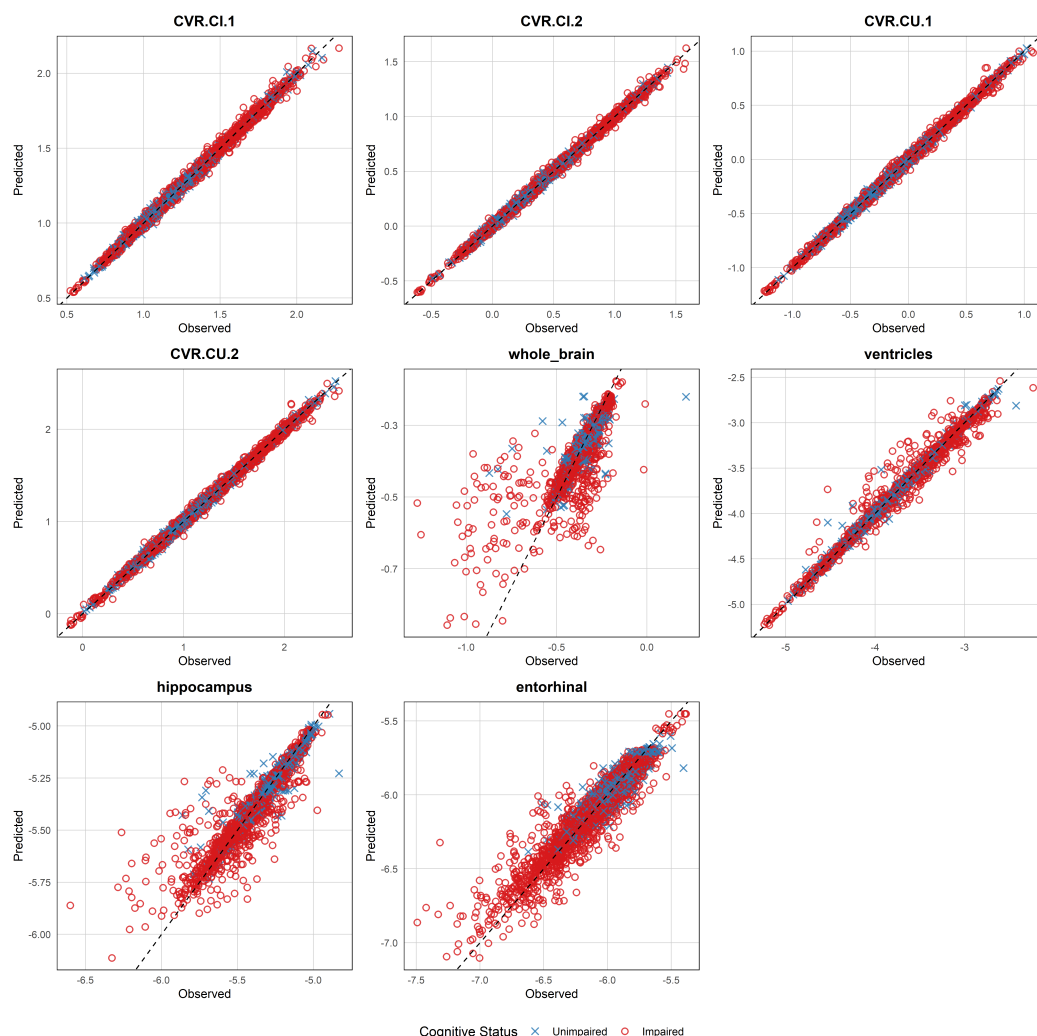


Figure B.2: Predicted vs observed values, for subset 2. Values are log-transformed. CVR is the composite value ratio of two data-driven regions of interest; the rest are divided by the intracranial volume. The CVR naming coding is as follows: CI = conditioned to minimise the sample size estimate of cognitively impaired individuals; CU = conditioned to minimise the sample size estimate of cognitively unimpaired individuals; suffix 1 or 2 indicate the subset of the data that was used for the CVR discovery.

Appendix C

Weighted CVR Supplementary Material

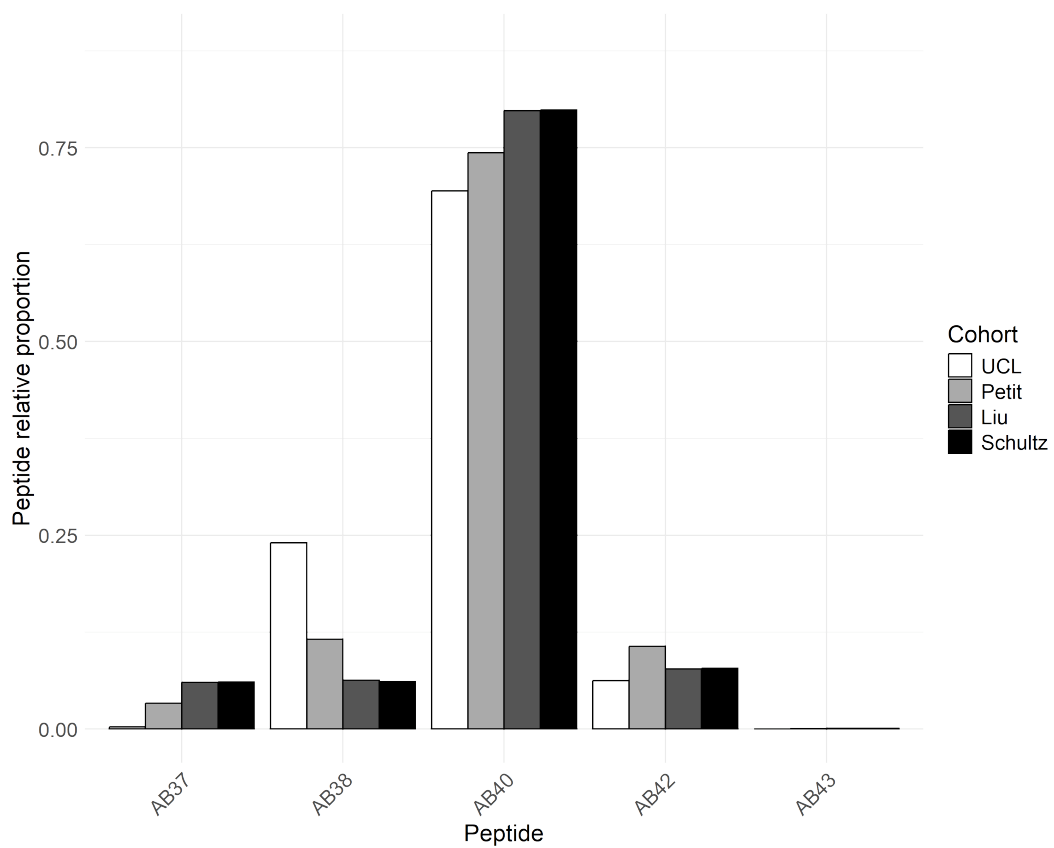


Figure C.1: Relative abundance of controls, by dataset. We see similar trends, but distinct enough to warrant a harmonisation pre-process if data is to be merged or a general, universal threshold be proposed.

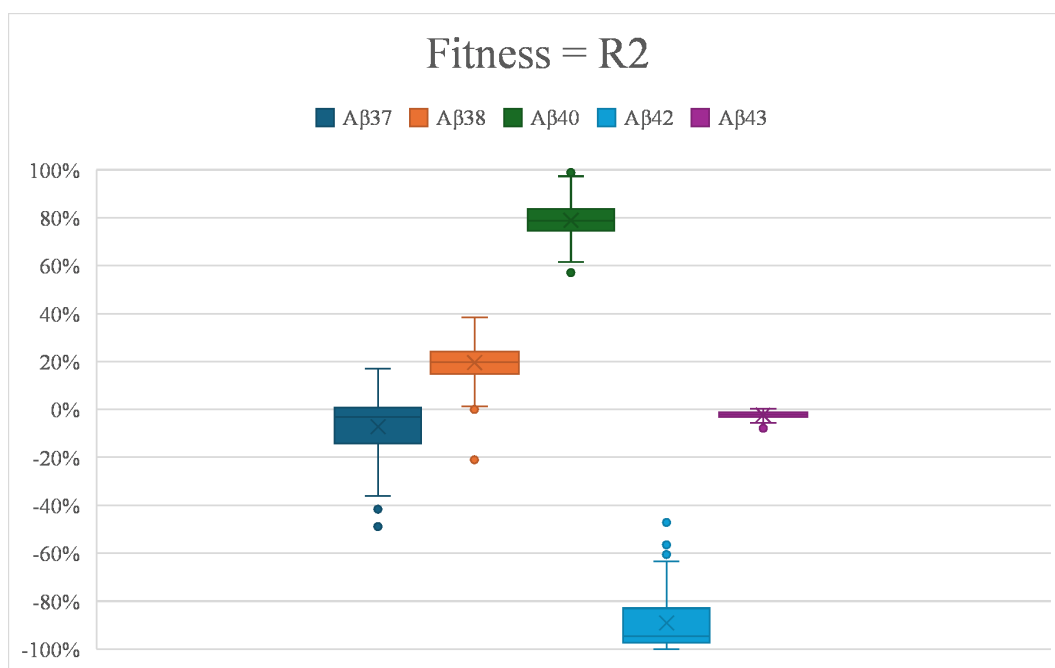


Figure C.2: Distribution of peptide relative weights of 200 bootstrapped samples, **optimized for R^2 vs age-at-onset**. The weights are relative to all peptides in the numerator (positive) or denominator (negative).

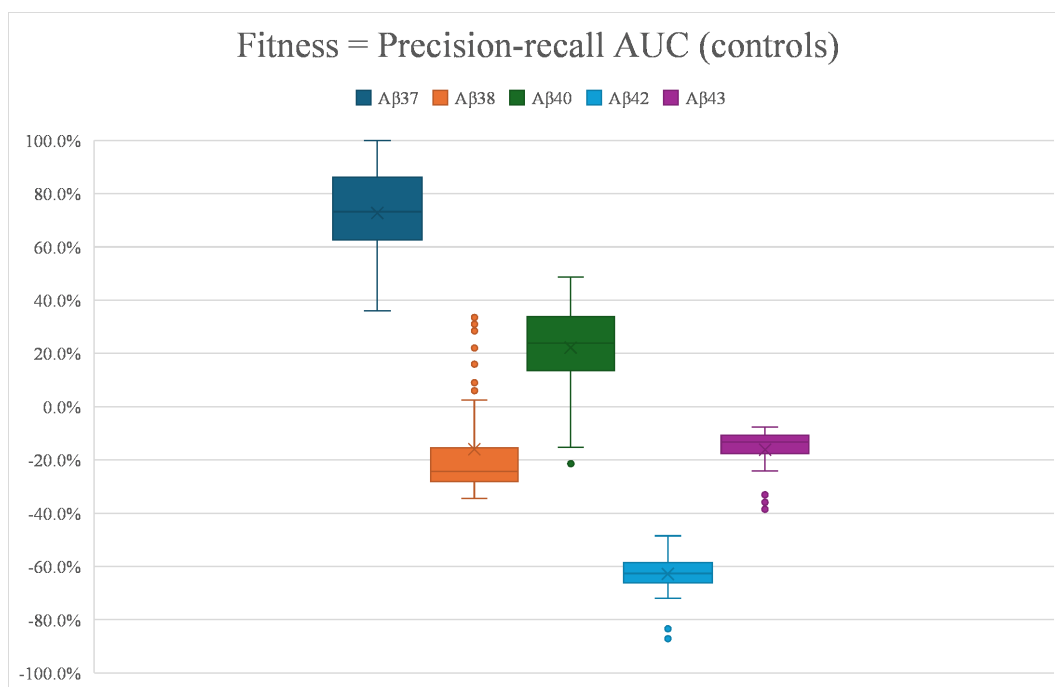


Figure C.3: Distribution of peptide relative weights of 200 bootstrapped samples, **optimized for precision-recall of controls**. The weights are relative to all peptides in the numerator (positive) or denominator (negative).

Table C.1: Permutation feature importance. For all metrics, and for both displayed biomarkers ($A\beta$ short/long and wCVR-RP), $A\beta$ 42 is the peptide providing the biggest impact. The data-driven wCVR-RP better employs information from $A\beta$ 43, as the importance values are higher in all metrics.

	1 - Precision-recall AUC		1 - (R^2 vs AAO)	
	short/long	wCVR-RP	short/long	wCVR-RP
$A\beta$ 37	0.99 (0.98, 1.04)	1.13 (1.02, 1.19)	1.00 (0.99, 1.00)	0.99 (0.97, 1.02)
$A\beta$ 38	1.10 (1.01, 1.17)	1.09 (1.06, 1.19)	1.05 (1.02, 1.09)	1.03 (1.00, 1.05)
$A\beta$ 40	1.06 (0.92, 1.14)	1.18 (1.00, 1.24)	1.06 (1.05, 1.09)	1.06 (1.05, 1.08)
$A\beta$ 42	1.58 (1.56, 1.61)	1.99 (1.82, 2.09)	1.61 (1.54, 1.66)	1.52 (1.45, 1.57)
$A\beta$ 43	1.02 (1.01, 1.19)	1.55 (1.18, 1.57)	1.03 (1.00, 1.05)	1.19 (1.10, 1.22)

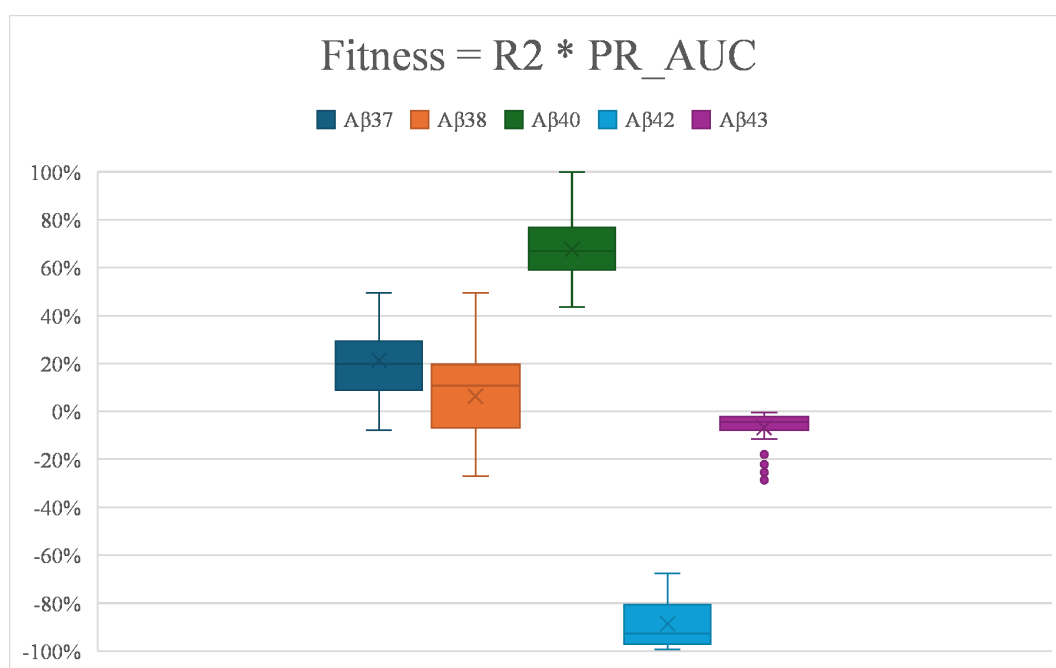


Figure C.4: Distribution of peptide relative weights of 200 bootstrapped samples, **optimized for R^2 vs age-at-onset, multiplying the precision-recall area-under-the-curve of controls**. The weights are relative to all peptides in the numerator (positive) or denominator (negative).

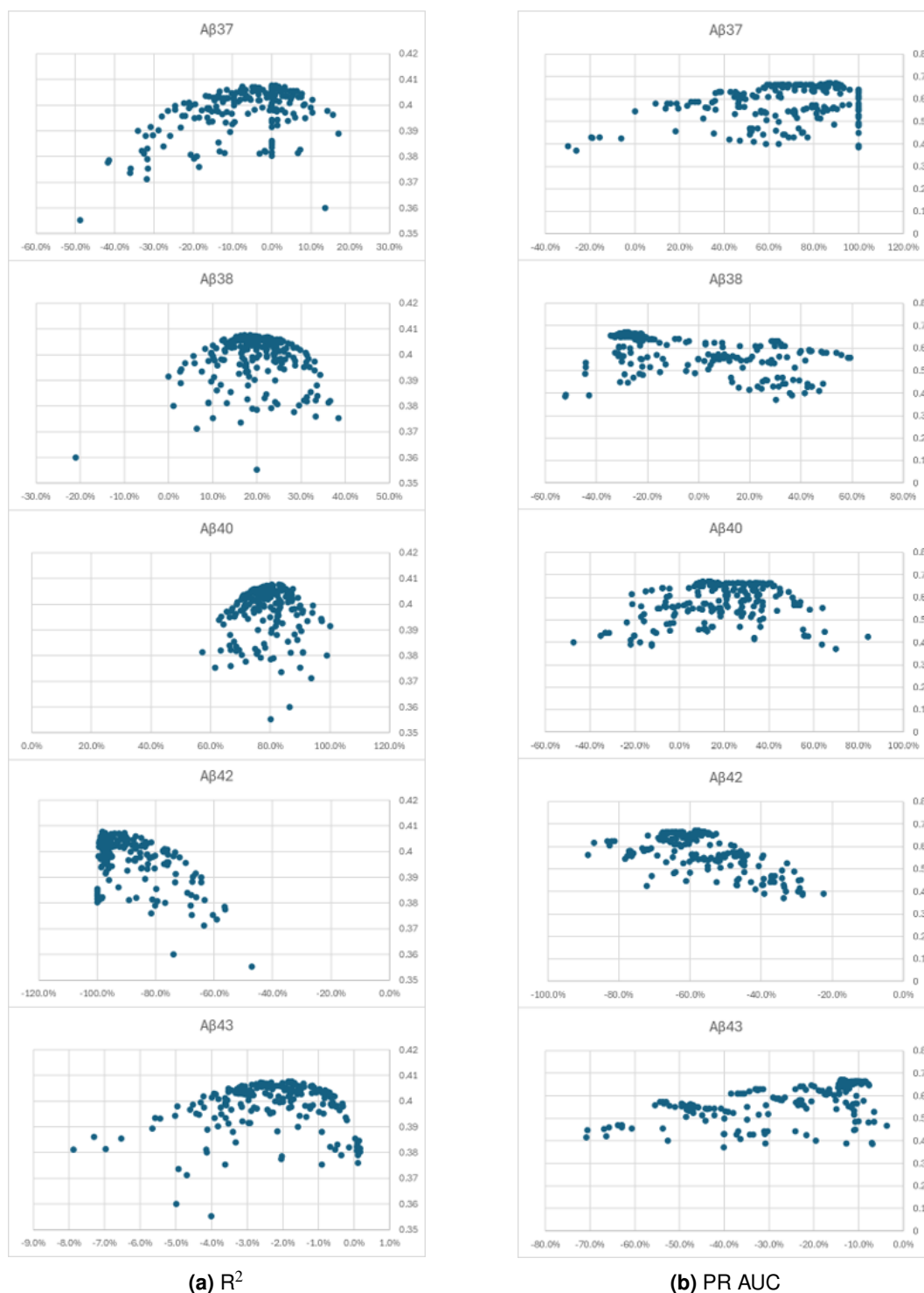


Figure C.5: Evaluation metrics versus peptide weights, from 200 bootstrapped samples. Left panel (a) is the R^2 versus age-at-onset, and left panel (b) is the precision-recall area-under-the-curve of controls. Weights above 0 mean the peptide is included in the numerator, and negative weights imply the peptide is included in the denominator.

Appendix D

Theta Supplementary Material

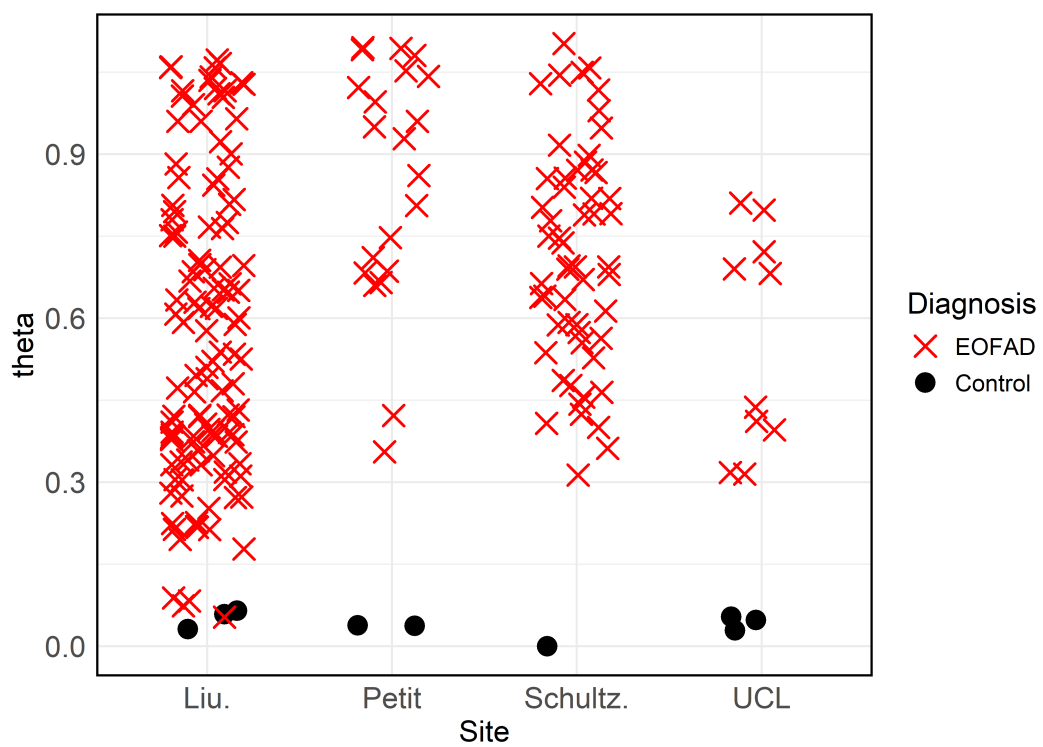


Figure D.1: Strip plot of theta values per site. Two of the four mutations in Liu are actually benign, and the other two are of unknown pathogenicity. The benign ones were re-labeled for the analysis of Chapter 7.

Appendix E

Colophon

This document was set in the Arial typeface using $\text{\LaTeX}$ and $\text{\BibTeX}$, composed with pdfLaTeX. Template from <https://github.com/UCL/ucl-latex-thesis-templates>.

Bibliography

- [1] M Monica Moore, Mirella Díaz-Santos, and Keith Vossel. “Alzheimer’s association 2021 facts and figures report”. In: *Alzheimers Dementia* 17.3 (Mar. 2021), pp. 327–406.
- [2] MAPPING A Better. “Alzheimer’s disease facts and figures”. In: *Alzheimers Dement* 19.4 (2023), pp. 1598–1695.
- [3] David M Cash et al. “Imaging endpoints for clinical trials in Alzheimer’s disease”. In: *Alzheimer’s research & therapy* 6 (2014), pp. 1–10.
- [4] Clifford R. Jack Jr. et al. “Revised criteria for diagnosis and staging of Alzheimer’s disease: Alzheimer’s Association Workgroup”. en. In: *Alzheimer’s & Dementia* n/a.n/a (2024). ISSN: 1552-5279. DOI: 10.1002/alz.13859.
- [5] Julie Ottoy et al. “Recent advances in neuroimaging of Alzheimer’s disease and related dementias”. In: *Alzheimer’s & dementia: the journal of the Alzheimer’s Association* 21.9 (2025), e70648.
- [6] World Health Organization. *Dementia*. 2023.
- [7] Alzheimer’s Society. *Annual Report 2024*. 2024.
- [8] Maoshing Ni. *The yellow emperor’s classic of medicine: a new translation of the neijing suwen with commentary*. Shambhala Publications, 1995.
- [9] Caleb E Finch and Stanley M Burstein. “Dementia in the Ancient Greco-Roman World was minimally mentioned”. In: *Journal of Alzheimer’s Disease* Preprint (2024), pp. 1–8.
- [10] Bernard Ineichen. “Introducing the Subject”. In: *Senile Dementia: Policy and services*. Boston, MA: Springer US, 1989, pp. 1–5. ISBN: 978-1-4899-3476-5. DOI: 10.1007/978-1-4899-3476-5_1.

- [11] Alzheimer's Society. *Who Gets Alzheimer's Disease?* 2023.
- [12] Alois Alzheimer. "Über eigenartige Erkrankung der Hirnrinde". In: *All Z Psychiatr* 64 (1907), pp. 146–148.
- [13] National Institute on Aging. *Alzheimer's Disease Genetics Fact Sheet*. 2024.
- [14] Matthew A Lalli et al. "Origin of the PSEN1 E280A mutation causing early-onset Alzheimer's disease". In: *Alzheimer's & Dementia* 10 (2014), S277–S283.
- [15] Leon J Thal et al. "The role of biomarkers in clinical trials for Alzheimer disease". In: *Alzheimer Disease & Associated Disorders* 20.1 (2006), pp. 6–15.
- [16] Alexander J Ehrenberg et al. "Relevance of biomarkers across different neurodegenerative diseases". In: *Alzheimer's research & therapy* 12 (2020), pp. 1–11.
- [17] Alexandra L Young et al. "Uncovering the heterogeneity and temporal complexity of neurodegenerative diseases with Subtype and Stage Inference". In: *Nature communications* 9.1 (2018), p. 4273.
- [18] Jacob W. Vogel et al. "Four distinct trajectories of tau deposition identified in Alzheimer's disease". en. In: *Nature Medicine* 27.5 (May 2021), pp. 871–881. ISSN: 1546-170X. DOI: 10.1038/s41591-021-01309-6.
- [19] Ranjan Duara and Warren Barker. "Heterogeneity in Alzheimer's disease diagnosis and progression rates: implications for therapeutic trials". In: *Neurotherapeutics* 19.1 (2022), pp. 8–25.
- [20] Timothy Rittman. "Functional Networks: A New Window Into Alzheimer's Disease Subtypes". In: *Biological Psychiatry* 93.9 (2023), pp. 752–753.
- [21] Nicholas Yew Wei Heng and Timothy Rittman. "Understanding ethnic diversity in open dementia neuroimaging data sets". In: *Brain Communications* 5.6 (2023), fcad308.
- [22] Seyed Hani Hojjati et al. "Heterogeneous tau deposition patterns in the preclinical stage link to domain-specific cognitive deficits". In: *Alzheimer's & Dementia* 21.5 (2025), e70153.

- [23] Hubert M Fonteijn et al. "An event-based disease progression model and its application to familial Alzheimer's disease". In: *Biennial International Conference on Information Processing in Medical Imaging*. Springer. 2011, pp. 748–759.
- [24] Neil P Oxtoby and Daniel C Alexander. "Imaging plus X: multimodal models of neurodegenerative disease". In: *Current opinion in neurology* 30.4 (2017), p. 371.
- [25] Timothy Rittman. "Neurological update: neuroimaging in dementia". In: *Journal of Neurology* 267.11 (2020), pp. 3429–3435.
- [26] Anja Laader et al. "1.5 versus 3 versus 7 Tesla in abdominal MRI: A comparative study". In: *PloS one* 12.11 (2017), e0187528.
- [27] Jason P Stockmann and Lawrence L Wald. "In vivo B0 field shimming methods for MRI at 7 T". In: *Neuroimage* 168 (2018), pp. 71–87.
- [28] MRI Master. *TR and TE in MRI*. <https://mrimester.com/tr-and-te-in-mri/>. 2025.
- [29] Prashanthi Vemuri and Clifford R Jack Jr. "Role of structural MRI in Alzheimer's disease". In: *Alzheimer's research & therapy* 2.4 (2010), p. 23.
- [30] D. Chan et al. "Rates of global and regional cerebral atrophy in AD and frontotemporal dementia". en. In: *Neurology* 57.10 (Nov. 2001), pp. 1756–1763. ISSN: 0028-3878, 1526-632X. DOI: 10.1212/WNL.57.10.1756.
- [31] Clifford R. Jack et al. "MR-based hippocampal volumetry in the diagnosis of Alzheimer's disease". en. In: *Neurology* 42.1 (Jan. 1992), pp. 183–183. ISSN: 0028-3878, 1526-632X. DOI: 10.1212/WNL.42.1.183.
- [32] Hans Förstl et al. "Brain Atrophy in Normal Ageing and Alzheimer's Disease: Volumetric Discrimination and Clinical Correlations". en. In: *British Journal of Psychiatry* 167.6 (Dec. 1995), pp. 739–746. ISSN: 0007-1250, 1472-1465. DOI: 10.1192/bjp.167.6.739.
- [33] Bruce Fischl and Anders M Dale. "Measuring the thickness of the human cerebral cortex from magnetic resonance images". In: *Proceedings of the National Academy of Sciences* 97.20 (2000), pp. 11050–11055.

- [34] Nick C Fox and Peter A Freeborough. "Brain atrophy progression measured from registered serial MRI: validation and application to Alzheimer's disease". In: *Journal of Magnetic Resonance Imaging* 7.6 (1997), pp. 1069–1075.
- [35] Ferran Prados et al. "Measuring brain atrophy with a generalized formulation of the boundary shift integral". In: *Neurobiology of aging* 36 (2015), S81–S90.
- [36] Philip Scheltens et al. "Atrophy of medial temporal lobes on MRI in" probable" Alzheimer's disease and normal ageing: diagnostic value and neuropsychological correlates." In: *Journal of Neurology, Neurosurgery & Psychiatry* 55.10 (1992), pp. 967–972.
- [37] Ji-seon Lee et al. "Impact of the ventricle size on alzheimer's disease progression: A retrospective longitudinal study". In: *Dementia and Neurocognitive Disorders* 23.2 (2024), p. 95.
- [38] Kim A Celone et al. "Alterations in memory networks in mild cognitive impairment and Alzheimer's disease: an independent component analysis". In: *Journal of Neuroscience* 26.40 (2006), pp. 10222–10231.
- [39] Federica Agosta et al. "Resting state fMRI in Alzheimer's disease: beyond the default mode network". In: *Neurobiology of aging* 33.8 (2012), pp. 1564–1578.
- [40] Olivier Naggara et al. "Diffusion tensor imaging in early Alzheimer's disease". In: *Psychiatry Research: Neuroimaging* 146.3 (2006), pp. 243–249.
- [41] Lin Zhuang et al. "White matter integrity in mild cognitive impairment: a tract-based spatial statistics study". In: *Neuroimage* 53.1 (2010), pp. 16–25.
- [42] Hui Zhang et al. "NODDI: practical in vivo neurite orientation dispersion and density imaging of the human brain". In: *Neuroimage* 61.4 (2012), pp. 1000–1016.
- [43] Madelaine Daianu et al. "Disrupted brain connectivity in Alzheimer's disease: Effects of network thresholding". In: *Computational Diffusion MRI and Brain Connectivity: MICCAI Workshops, Nagoya, Japan, September 22nd, 2013*. Springer. 2013, pp. 199–208.

- [44] E Mark Haacke et al. "Imaging cerebral amyloid angiopathy with susceptibility-weighted imaging". In: *American journal of neuroradiology* 28.2 (2007), pp. 316–317.
- [45] Takashi Yoshiura et al. "Arterial spin labelling at 3-T MR imaging for detection of individuals with Alzheimer's disease". In: *European radiology* 19.12 (2009), pp. 2819–2825.
- [46] Elizabeth Joe et al. "1H MRS spectroscopy in preclinical autosomal dominant Alzheimer disease". In: *Brain imaging and behavior* 13.4 (2019), pp. 925–932.
- [47] Clifford R Jack Jr et al. "NIA-AA research framework: toward a biological definition of Alzheimer's disease". In: *Alzheimer's & dementia* 14.4 (2018), pp. 535–562.
- [48] Satoshi Minoshima et al. "18F-FDG PET imaging in neurodegenerative dementing disorders: insights into subtype classification, emerging disease categories, and mixed dementia with copathologies". In: *Journal of nuclear medicine* 63.Supplement 1 (2022), 2S–12S.
- [49] Giuseppe Pasqualetti, David J Brooks, and Paul Edison. "The role of neuroinflammation in dementias". In: *Current neurology and neuroscience reports* 15 (2015), pp. 1–11.
- [50] Michel M Ter-Pogossian et al. "A positron-emission transaxial tomograph for nuclear imaging (PETT)". In: *Radiology* 114.1 (1975), pp. 89–98.
- [51] Bernd J Pichler, Martin S Judenhofer, and Christina Pfannenberger. "Multimodal imaging approaches: Pet/ct and pet/mri". In: *Molecular Imaging I* (2008), pp. 109–132.
- [52] Karl Herholz. "FDG PET and differential diagnosis of dementia". In: *Alzheimer Disease & Associated Disorders* 9.1 (1995), pp. 6–16.
- [53] Lisa Mosconi et al. "Multicenter standardized 18F-FDG PET diagnosis of mild cognitive impairment, Alzheimer's disease, and other dementias". In: *Journal of nuclear medicine* 49.3 (2008), pp. 390–398.

- [54] William E Klunk et al. "Imaging brain amyloid in Alzheimer's disease with Pittsburgh Compound-B". In: *Annals of Neurology: Official Journal of the American Neurological Association and the Child Neurology Society* 55.3 (2004), pp. 306–319.
- [55] Dean F Wong et al. "In vivo imaging of amyloid deposition in Alzheimer disease using the radioligand 18F-AV-45 (flobetapir F 18)". In: *Journal of nuclear medicine* 51.6 (2010), pp. 913–920.
- [56] Christopher M Clark et al. "Use of florbetapir-PET for imaging β -amyloid pathology". In: *Jama* 305.3 (2011), pp. 275–283.
- [57] S. Budd Haeberlein et al. "Two Randomized Phase 3 Studies of Aducanumab in Early Alzheimer's Disease". en. In: *The Journal of Prevention of Alzheimer's Disease* (2022). ISSN: 22745807. DOI: 10.14283/jpad.2022.30.
- [58] Willemijn J Jansen et al. "Prevalence of cerebral amyloid pathology in persons without dementia: a meta-analysis". In: *Jama* 313.19 (2015), pp. 1924–1938.
- [59] Michael J Pontecorvo et al. "Relationships between flortaucipir PET tau binding and amyloid burden, clinical diagnosis, age and cognition". In: *Brain* 140.3 (2017), pp. 748–763.
- [60] Marta Marquié et al. "Lessons learned about [F-18]-AV-1451 off-target binding from an autopsy-confirmed Parkinson's case". In: *Acta neuropathologica communications* 5.1 (2017), p. 75.
- [61] Adam J Schwarz et al. "Regional profiles of the candidate tau PET ligand 18 F-AV-1451 recapitulate key features of Braak histopathological stages". In: *Brain* 139.5 (2016), pp. 1539–1550.
- [62] Suzanne L Baker, Anne Maass, and William J Jagust. "Considerations and code for partial volume correcting [18F]-AV-1451 tau PET data". In: *Data in brief* 15 (2017), pp. 648–657.
- [63] Tobey J Betthausen et al. "In vivo characterization and quantification of neurofibrillary tau PET radioligand 18F-MK-6240 in humans from Alzheimer disease dementia to young controls". In: *Journal of Nuclear Medicine* 60.1 (2019), pp. 93–99.

- [64] Antoine Leuzy et al. "Tau PET imaging in neurodegenerative tauopathies—still a challenge". In: *Molecular psychiatry* 24.8 (2019), pp. 1112–1134.
- [65] Dominique Gouilly et al. "Neuroinflammation PET imaging of the translocator protein (TSPO) in Alzheimer's disease: an update". In: *European Journal of Neuroscience* 55.5 (2022), pp. 1322–1343.
- [66] Negin Holland et al. "Synaptic loss in primary tauopathies revealed by [11C] UCB-J positron emission tomography". In: *Movement disorders* 35.10 (2020), pp. 1834–1842.
- [67] Mika Naganawa et al. "First-in-human evaluation of 18F-SynVesT-1, a radioligand for PET imaging of synaptic vesicle glycoprotein 2A". In: *Journal of Nuclear Medicine* 62.4 (2021), pp. 561–567.
- [68] Priya Singh et al. "The 18-kDa translocator protein PET tracers as a diagnostic marker for neuroinflammation: development and current standing". In: *ACS omega* 7.17 (2022), pp. 14412–14429.
- [69] Solveig Tiepolt et al. "PET imaging of cholinergic neurotransmission in neurodegenerative disorders". In: *Journal of Nuclear Medicine* 63.Supplement 1 (2022), 33S–44S.
- [70] Chris WJ van der Weijden et al. "Quantitative myelin imaging with MRI and PET: an overview of techniques and their validation status". In: *Brain* 146.4 (2023), pp. 1243–1266.
- [71] Robert B Innis et al. "Consensus nomenclature for in vivo imaging of reversibly binding radioligands". In: *Journal of Cerebral Blood Flow & Metabolism* 27.9 (2007), pp. 1533–1539.
- [72] Brian J Lopresti et al. "Simplified quantification of Pittsburgh Compound B amyloid imaging PET studies: a comparative analysis". In: *Journal of Nuclear Medicine* 46.12 (2005), pp. 1959–1972.
- [73] S. M. Landau et al. "Amyloid PET imaging in Alzheimer's disease: a comparison of three radiotracers". eng. In: *European Journal of Nuclear Medicine and Molecular Imaging* 41.7 (July 2014), pp. 1398–1407. ISSN: 1619-7089. DOI: 10.1007/s00259-014-2753-3.

- [74] Adam S. Fleisher et al. "Use of white matter reference regions for detection of change in florbetapir positron emission tomography from completed phase 3 solanezumab trials". en. In: *Alzheimer's & Dementia* 13.10 (2017), pp. 1117–1124. ISSN: 1552-5279. DOI: 10.1016/j.jalz.2017.02.009.
- [75] John R. Sims et al. "Donanemab in Early Symptomatic Alzheimer Disease: The TRAILBLAZER-ALZ 2 Randomized Clinical Trial". en. In: *JAMA* 330.6 (Aug. 2023), p. 512. ISSN: 0098-7484. DOI: 10.1001/jama.2023.13239.
- [76] Isaac Llorente Saguer and Neil P Oxtoby. "Choosing wisely: Evaluating reference regions for improved longitudinal 18F-AV-1451 tau PET consistency". In: *Alzheimer's Association International Conference. ALZ.* 2024.
- [77] Kjell Erlandsson et al. "A review of partial volume correction techniques for emission tomography and their applications in neurology, cardiology and oncology". In: *Physics in Medicine & Biology* 57.21 (2012), R119.
- [78] Christopher G. Schwarz et al. "Selecting software pipelines for change in flortaucipir SUVR: Balancing repeatability and group separation". In: *NeuroImage* 238 (Sept. 2021), p. 118259. ISSN: 1053-8119. DOI: 10.1016/j.neuroimage.2021.118259.
- [79] Christopher H. Van Dyck et al. "Lecanemab in Early Alzheimer's Disease". en. In: *New England Journal of Medicine* 388.1 (Jan. 2023), pp. 9–21. ISSN: 0028-4793, 1533-4406. DOI: 10.1056/NEJMoa2212948.
- [80] William E. Klunk et al. "The Centiloid Project: Standardizing quantitative amyloid plaque estimation by PET". en. In: *Alzheimer's & Dementia* 11.1 (Jan. 2015), p. 1. ISSN: 1552-5260, 1552-5279. DOI: 10.1016/j.jalz.2014.07.003.
- [81] Vincent Dore et al. "CenTauRz: A standardized quantification of tau PET scans". In: *Alzheimer's & Dementia* 18 (2022), e061177.
- [82] Oskar Hansson et al. "Prediction of Alzheimer's disease using the CSF A β 42/A β 40 ratio in patients with mild cognitive impairment". In: *Dementia and geriatric cognitive disorders* 23.5 (2007), pp. 316–320.

- [83] Ross W Paterson et al. "Cerebrospinal fluid in the differential diagnosis of Alzheimer's disease: clinical utility of an extended panel of biomarkers in a specialist cognitive clinic". In: *Alzheimer's research & therapy* 10 (2018), pp. 1–11.
- [84] Ashvini Keshavan et al. "CSF biomarkers for dementia". In: *Practical Neurology* 22.4 (2022), pp. 285–294. ISSN: 1474-7758. DOI: 10.1136/practneurol-2021-003310. eprint: <https://pn.bmj.com/content/22/4/285.full.pdf>.
- [85] Nicolas R Barthélemy et al. "CSF tau phosphorylation occupancies at T217 and T205 represent improved biomarkers of amyloid and tau pathology in Alzheimer's disease". In: *Nature aging* 3.4 (2023), pp. 391–401.
- [86] Kanta Horie et al. "CSF MTBR-tau243 is a specific biomarker of tau tangle pathology in Alzheimer's disease". In: *Nature Medicine* 29.8 (2023), pp. 1954–1963.
- [87] Niels Andreasen, Magnus Sjögren, and Kaj Blennow. "CSF markers for Alzheimer's disease: total tau, phospho-tau and A β 42". In: *The world journal of biological psychiatry* 4.4 (2003), pp. 147–155.
- [88] Burak Arslan, Henrik Zetterberg, and Nicholas J Ashton. "Blood-based biomarkers in Alzheimer's disease—moving towards a new era of diagnostics". In: *Clinical Chemistry and Laboratory Medicine (CCLM)* 62.6 (2024), pp. 1063–1069.
- [89] Janice Hopkins Tanne. *FDA approves blood test to diagnose Alzheimer's*. 2025.
- [90] Charles Arber et al. "Familial Alzheimer's disease patient-derived neurons reveal distinct mutation-specific effects on amyloid beta". en. In: *Molecular Psychiatry* 25.11 (Nov. 2020), pp. 2919–2931. ISSN: 1476-5578. DOI: 10.1038/s41380-019-0410-8.
- [91] Marshal F Folstein, Susan E Folstein, and Paul R McHugh. "'Mini-mental state': a practical method for grading the cognitive state of patients for the clinician". In: *Journal of psychiatric research* 12.3 (1975), pp. 189–198.
- [92] John C Morris. "The Clinical Dementia Rating (CDR) current version and scoring rules". In: *Neurology* 43.11 (1993), pp. 2412–2412.

- [93] John C Morris. "Clinical dementia rating: a reliable and valid diagnostic and staging measure for dementia of the Alzheimer type". In: *International psychogeriatrics* 9 (1997), pp. 173–176.
- [94] Wilma G Rosen, Richard C Mohs, and Kenneth L Davis. "A new rating scale for Alzheimer's disease." In: *The American journal of psychiatry* 141.11 (1984), pp. 1356–1364.
- [95] David P Graham et al. "The Alzheimer's Disease Assessment Scale-Cognitive subscale: normative data for older adult controls". In: *Alzheimer Disease & Associated Disorders* 18.4 (2004), pp. 236–240.
- [96] Jacqueline K Kueper, Mark Speechley, and Manuel Montero-Odasso. "The Alzheimer's disease assessment scale–cognitive subscale (ADAS-Cog): modifications and responsiveness in pre-dementia populations. a narrative review". In: *Journal of Alzheimer's Disease* 63.2 (2018), pp. 423–444.
- [97] Michael C Donohue et al. "The preclinical Alzheimer cognitive composite: measuring amyloid-related decline". In: *JAMA neurology* 71.8 (2014), pp. 961–970.
- [98] Kathryn V Papp et al. "Optimizing the preclinical Alzheimer's cognitive composite with semantic processing: the PACC5". In: *Alzheimer's & Dementia: Translational Research & Clinical Interventions* 3.4 (2017), pp. 668–677.
- [99] Paul Maruff et al. "Validity of the CogState brief battery: relationship to standardized tests and sensitivity to cognitive impairment in mild traumatic brain injury, schizophrenia, and AIDS dementia complex". In: *Archives of clinical neuropsychology* 24.2 (2009), pp. 165–178.
- [100] Claire Lancaster et al. "Evaluating the feasibility of frequent cognitive assessment using the Mezurio smartphone app: observational and interview study in adults with elevated dementia risk". In: *JMIR mHealth and uHealth* 8.4 (2020), e16142.
- [101] Irene B Meier et al. "Using a Digital Neuro Signature to measure longitudinal individual-level change in Alzheimer's disease: the Altoida large cohort study". In: *NPJ digital medicine* 4.1 (2021), p. 101.

- [102] Jeffrey Kaye et al. "Unobtrusive measurement of daily computer use to detect mild cognitive impairment". In: *Alzheimer's & Dementia* 10.1 (2014), pp. 10–17.
- [103] Hessa Alfalahi et al. "Diagnostic accuracy of keystroke dynamics as digital biomarkers for fine motor decline in neuropsychiatric disorders: a systematic review and meta-analysis". In: *Scientific reports* 12.1 (2022), p. 7690.
- [104] U.S. Food and Drug Administration. *Digital Health Technologies for Remote Data Acquisition in Clinical Investigations: Guidance for Industry, Investigators, and Other Stakeholders*. <https://www.fda.gov/media/155022/download>. Dec. 2023.
- [105] Antoine Piau et al. "Current state of digital biomarker technologies for real-life, home-based monitoring of cognitive function for mild cognitive impairment to mild Alzheimer disease and implications for clinical care: systematic review". In: *Journal of medical Internet research* 21.8 (2019), e12785.
- [106] Elizabeth Ford, Richard Milne, and Keegan Curlewis. "Ethical issues when using digital biomarkers and artificial intelligence for the early detection of dementia". In: *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery* 13.3 (2023), e1492.
- [107] Jeffrey L. Cummings et al. "Alzheimer's disease drug development pipeline: 2025". In: *Alzheimer's & Dementia: Translational Research & Clinical Interventions* 11.2 (2025), e70098. DOI: <https://doi.org/10.1002/trc2.70098>. eprint: <https://alz-journals.onlinelibrary.wiley.com/doi/pdf/10.1002/trc2.70098>.
- [108] Susan Abushakra et al. "APOLLOE4 Phase 3 study of oral ALZ-801/valiltramiprosate in APOE $\epsilon 4/\epsilon 4$ homozygotes with early Alzheimer's disease: Trial design and baseline characteristics". In: *Alzheimer's & Dementia: Translational Research & Clinical Interventions* 10.3 (2024), e12498.
- [109] Susan Abushakra et al. *Topline Results from Pivotal APOLLOE4 Phase 3 Trial of Oral Valiltramiprosate/ALZ-801 in Patients with Early Alzheimer's Disease Carrying Two Copies of APOE4 Gene*. en. 2025.

- [110] Amanda L Edwards et al. "Exploratory tau biomarker results from a multiple ascending-dose study of BIIB080 in Alzheimer disease: a randomized clinical trial". In: *JAMA neurology* 80.12 (2023), pp. 1344–1352.
- [111] Jeffrey L Cummings et al. "evoked and evoked+: design of two large-scale, double-blind, placebo-controlled, phase 3 studies evaluating efficacy, safety, and tolerability of semaglutide in early-stage symptomatic Alzheimer's disease". In: *Alzheimer's research & therapy* 17.1 (2025), p. 14.
- [112] Julia Elise Cabral et al. "Targeting the NLRP3 inflammasome for inflammatory disease therapy". In: *Trends in Pharmacological Sciences* (2025).
- [113] Kan Yin Wong et al. "Relationships between mitochondrial dysfunction and neurotransmission failure in Alzheimer's disease". In: *Aging and disease* 11.5 (2020), p. 1291.
- [114] Lauren Priest et al. "Early positive signals of cognitive outcomes in mild to moderate Alzheimer patients treated with the synaptic regenerative small molecule, SPG302". In: *Alzheimer's Association International Conference (AAIC)*. Toronto, Canada, 2025.
- [115] Cheng Fang et al. "Buntanetap, a novel translational inhibitor of multiple neurotoxic proteins, proves to be safe and promising in both Alzheimer's and Parkinson's patients". In: *The journal of prevention of Alzheimer's disease* 10.1 (2023), pp. 25–33.
- [116] Stephen Macfarlane et al. "Blarcamesine for the treatment of Early Alzheimer's Disease: Results from the ANAVEX2-73-AD-004 Phase IIB/III trial". In: *The Journal of Prevention of Alzheimer's Disease* 12.1 (2025), p. 100016.
- [117] Bruno Dubois et al. "Donepezil decreases annual rate of hippocampal atrophy in suspected prodromal Alzheimer's disease". en. In: *Alzheimer's & Dementia* 11.9 (2015), pp. 1041–1049. ISSN: 1552-5279. DOI: 10.1016/j.jalz.2014.10.003.
- [118] Paul S. Aisen et al. "Tramiprosate in mild-to-moderate Alzheimer's disease - a randomized, double-blind, placebo-controlled, multi-centre study (the

- Alphase Study)". eng. In: *Archives of medical science: AMS* 7.1 (Feb. 2011), pp. 102–111. ISSN: 1896-9151. DOI: 10.5114/aoms.2011.20612.
- [119] Gordon K Wilcock et al. "Potential of low dose leuco-methylthionium bis (hydromethanesulphonate)(LMTM) monotherapy for treatment of mild Alzheimer's disease: Cohort analysis as modified primary outcome in a phase III clinical trial". In: *Journal of Alzheimer's Disease* 61.1 (2017), pp. 435–457.
- [120] Stephen Salloway et al. "Amyloid-Related Imaging Abnormalities in 2 Phase 3 Studies Evaluating Aducanumab in Patients With Early Alzheimer Disease". In: *JAMA Neurology* 79.1 (Jan. 2022), pp. 13–21. ISSN: 2168-6149. DOI: 10.1001/jamaneuro1.2021.4161.
- [121] Colin Groot et al. "Tau PET imaging in neurodegenerative disorders". In: *Journal of Nuclear Medicine* 63.Supplement 1 (2022), 20S–26S.
- [122] Sudeepti Southekal et al. "Flortaucipir F 18 quantitation using parametric estimation of reference signal intensity". In: *Journal of Nuclear Medicine* 59.6 (2018), pp. 944–951.
- [123] Christina B. Young et al. "Influence of common reference regions on regional tau patterns in cross-sectional and longitudinal [18F]-AV-1451 PET data". In: *NeuroImage* 243 (Nov. 2021), p. 118553. ISSN: 1053-8119. DOI: 10.1016/j.neuroimage.2021.118553.
- [124] Susan M. Landau et al. "Measurement of Longitudinal β -Amyloid Change with 18F-Florbetapir PET and Standardized Uptake Value Ratios". en. In: *Journal of Nuclear Medicine* 56.4 (Apr. 2015), pp. 567–574. ISSN: 0161-5505, 2159-662X. DOI: 10.2967/jnumed.114.148981.
- [125] Michael J Pontecorvo et al. "A multicentre longitudinal study of flortaucipir (18F) in normal ageing, mild cognitive impairment and Alzheimer's disease dementia". en. In: *Brain* 142.6 (June 2019), pp. 1723–1735. ISSN: 0006-8950, 1460-2156. DOI: 10.1093/brain/awz090.
- [126] Olivier G Rousset, Yilong Ma, and Alan C Evans. "Correction for partial volume effects in PET: principle and validation". In: *Journal of nuclear medicine* 39.5 (1998), pp. 904–911.

- [127] H. Braak and E. Braak. "Neuropathological stageing of Alzheimer-related changes". en. In: *Acta Neuropathologica* 82.4 (Sept. 1991), pp. 239–259. ISSN: 1432-0533. DOI: 10.1007/BF00308809.
- [128] Richard Baumgartner et al. "Statistical evaluation of test-retest studies in PET brain imaging". en. In: *EJNMMI Research* 8.1 (Feb. 2018), p. 13. ISSN: 2191-219X. DOI: 10.1186/s13550-018-0366-8.
- [129] Abhinay D Joshi et al. "A semiautomated method for quantification of F 18 florbetapir PET images". In: *Journal of Nuclear Medicine* 56.11 (2015), pp. 1736–1741.
- [130] Christopher G Schwarz et al. "A comparison of partial volume correction techniques for measuring change in serial amyloid PET SUVR". In: *Journal of Alzheimer's Disease* 67.1 (2019), pp. 181–195.
- [131] Isaac Llorente-Saguer and Neil P Oxtoby. "A data-driven framework for biomarker discovery applied to optimizing modern clinical and preclinical trials on Alzheimer's disease". In: *Brain Communications* 6.6 (2024), fcae438.
- [132] Reisa A. Sperling et al. "Trial of Solanezumab in Preclinical Alzheimer's Disease". In: *New England Journal of Medicine* 389.12 (Sept. 2023), pp. 1096–1107. ISSN: 0028-4793. DOI: 10.1056/NEJMoa2305032.
- [133] JiaQie Lee et al. "Amyloid PET Processing Methods". en. In: *adni.loni.usc.edu* (2023).
- [134] Sarah K. Royse et al. "Validation of amyloid PET positivity thresholds in centiloids: a multisite PET study approach". In: *Alzheimer's Research & Therapy* 13.1 (May 2021), p. 99. ISSN: 1758-9193. DOI: 10.1186/s13195-021-00836-1.
- [135] Rahul S. Desikan et al. "An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest". In: *NeuroImage* 31.3 (July 2006), pp. 968–980. ISSN: 1053-8119. DOI: 10.1016/j.neuroimage.2006.01.021.
- [136] JiaQie Lee et al. "Tau PET Processing Methods". en. In: *adni.loni.usc.edu* (2023).

- [137] Martin Reuter et al. “Within-subject template estimation for unbiased longitudinal image analysis”. en. In: *NeuroImage* 61.4 (July 2012), pp. 1402–1418. ISSN: 10538119. DOI: 10.1016/j.neuroimage.2012.02.084.
- [138] Edmund T. Rolls, Marc Joliot, and Nathalie Tzourio-Mazoyer. “Implementation of a new parcellation of the orbitofrontal cortex in the automated anatomical labeling atlas”. In: *NeuroImage* 122 (Nov. 2015), pp. 1–5. ISSN: 1053-8119. DOI: 10.1016/j.neuroimage.2015.07.075.
- [139] Luca Scrucca. *On some extensions to GA package: hybrid optimisation, parallelisation and islands evolution*. en. May 2016.
- [140] Clifford R Jack et al. “Age-specific and sex-specific prevalence of cerebral β -amyloidosis, tauopathy, and neurodegeneration in cognitively unimpaired individuals aged 50–95 years: a cross-sectional study”. en. In: *The Lancet Neurology* 16.6 (June 2017), pp. 435–444. ISSN: 14744422. DOI: 10.1016/S1474-4422(17)30077-7.
- [141] Antoine Leuzy et al. “Comparison of Group-Level and Individualized Brain Regions for Measuring Change in Longitudinal Tau Positron Emission Tomography in Alzheimer Disease”. en. In: *JAMA Neurology* 80.6 (June 2023), p. 614. ISSN: 2168-6149. DOI: 10.1001/jamaneurol.2023.1067.
- [142] Natasha Krishnadas et al. “Rates of regional tau accumulation in ageing and across the Alzheimer’s disease continuum: an AIBL 18F-MK6240 PET study”. English. In: *eBioMedicine* 88 (Feb. 2023). ISSN: 2352-3964. DOI: 10.1016/j.ebiom.2023.104450.
- [143] Team R.c. “R A language and environment for statistical computing, R Foundation for Statistical”. In: *Computing* (2020).
- [144] Douglas Bates et al. *Fitting Linear Mixed-Effects Models using lme4*. en. June 2014.
- [145] Samuel Iddi and Michael C. Donohue. “Power and Sample Size for Longitudinal Models in R – The longpower Package and Shiny App”. In: *The R Journal* 14.1 (July 2022), pp. 264–282. ISSN: 2073-4859.
- [146] Peter J Diggle. *Analysis of longitudinal data*. Oxford University Press, 2002.

- [147] Răzvan V. Marinescu et al. “BrainPainter: A Software for the Visualisation of Brain Structures, Biomarkers and Associated Pathological Processes”. en. In: *Multimodal Brain Image Analysis and Mathematical Foundations of Computational Anatomy*. Ed. by Dajiang Zhu et al. Cham: Springer International Publishing, 2019, pp. 112–120. ISBN: 978-3-030-33226-6. DOI: 10.1007/978-3-030-33226-6_13.
- [148] Kewei Chen et al. “Twelve-month metabolic declines in probable Alzheimer’s disease and amnesic mild cognitive impairment assessed using an empirically pre-defined statistical region-of-interest: Findings from the Alzheimer’s Disease Neuroimaging Initiative”. en. In: *NeuroImage* 51.2 (June 2010), pp. 654–664. ISSN: 10538119. DOI: 10.1016/j.neuroimage.2010.02.064.
- [149] Antoine Leuzy et al. “A multicenter comparison of [18F]flortaucipir, [18F]RO948, and [18F]MK6240 tau PET tracers to detect a common target ROI for differential diagnosis”. en. In: *European Journal of Nuclear Medicine and Molecular Imaging* 48.7 (July 2021), pp. 2295–2305. ISSN: 1619-7089. DOI: 10.1007/s00259-021-05401-4.
- [150] Michael Allen. “The huge carbon footprint of large-scale computing”. en. In: *Physics World* 35.3 (Aug. 2022), pp. 46–50. ISSN: 0953-8585, 2058-7058. DOI: 10.1088/2058-7058/35/03/32.
- [151] Reisa Sperling et al. “Amyloid-related imaging abnormalities in patients with Alzheimer’s disease treated with bapineuzumab: a retrospective analysis”. English. In: *The Lancet Neurology* 11.3 (Mar. 2012), pp. 241–249. ISSN: 1474-4422, 1474-4465. DOI: 10.1016/S1474-4422(12)70015-7.
- [152] P. M. Cogswell et al. “Amyloid-Related Imaging Abnormalities with Emerging Alzheimer Disease Therapeutics: Detection and Reporting Recommendations for Clinical Practice”. en. In: *American Journal of Neuroradiology* 43.9 (Sept. 2022), E19–E35. ISSN: 0195-6108, 1936-959X. DOI: 10.3174/ajnr.A7586.
- [153] Harald Hampel et al. “Amyloid-related imaging abnormalities (ARIA): radiological, biological and clinical characteristics”. In: *Brain* 146.11 (Nov. 2023), pp. 4414–4424. ISSN: 0006-8950. DOI: 10.1093/brain/awad188.

- [154] Reisa A. Sperling et al. "Association of Factors With Elevated Amyloid Burden in Clinically Normal Older Individuals". In: *JAMA Neurology* 77.6 (June 2020), pp. 735–745. ISSN: 2168-6149. DOI: 10.1001/jamaneuro.2020.0387.
- [155] Marina Ritchie, Daniel L. Gillen, and Joshua D. Grill. "Recruitment across two decades of NIH-funded Alzheimer's disease clinical trials". en. In: *Alzheimer's Research & Therapy* 15.1 (Feb. 2023), p. 28. ISSN: 1758-9193. DOI: 10.1186/s13195-023-01177-x.
- [156] Thomas J. Moore et al. "Variation in the estimated costs of pivotal clinical benefit trials supporting the US approval of new therapeutic agents, 2015–2017: a cross-sectional study". en. In: *BMJ Open* 10.6 (June 2020), e038863. ISSN: 2044-6055, 2044-6055. DOI: 10.1136/bmjopen-2020-038863.
- [157] Yan Li et al. "Validation of Plasma Amyloid- β 42/40 for Detecting Alzheimer Disease Amyloid Plaques". In: *Neurology* 98.7 (Feb. 2022), e688–e699. DOI: 10.1212/WNL.0000000000013211.
- [158] Isaac Llorente Saguer, Marc A Busche, and Neil P Oxtoby. "Composite SUVR: a new method for boosting Alzheimer's disease monitoring and diagnostic performance, applied to tau PET". en. In: *Alzheimer's & Dementia* 18.S5 (2022), e063177. ISSN: 1552-5279. DOI: 10.1002/alz.063177.
- [159] R C Gur et al. "Gender differences in age effect on brain atrophy measured by magnetic resonance imaging." en. In: *Proceedings of the National Academy of Sciences* 88.7 (Apr. 1991), pp. 2845–2849. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.88.7.2845.
- [160] Susan M. Resnick et al. "One-year Age Changes in MRI Brain Volumes in Older Adults". In: *Cerebral Cortex* 10.5 (May 2000), pp. 464–472. ISSN: 1047-3211. DOI: 10.1093/cercor/10.5.464.
- [161] Rachael I. Scahill et al. "A Longitudinal Study of Brain Volume Changes in Normal Aging Using Serial Registered Magnetic Resonance Imaging". In: *Archives of Neurology* 60.7 (July 2003), pp. 989–994. ISSN: 0003-9942. DOI: 10.1001/archneur.60.7.989.

- [162] Nick C. Fox, Peter A. Freeborough, and Martin N. Rossor. "Visualisation and quantification of rates of atrophy in Alzheimer's disease". English. In: *The Lancet* 348.9020 (July 1996), pp. 94–97. ISSN: 0140-6736, 1474-547X. DOI: 10.1016/S0140-6736(96)05228-2.
- [163] Susanne G. Mueller et al. "Hippocampal atrophy patterns in mild cognitive impairment and Alzheimer's disease". eng. In: *Human Brain Mapping* 31.9 (Sept. 2010), pp. 1339–1347. ISSN: 1097-0193. DOI: 10.1002/hbm.20934.
- [164] Michael S. Rafii and Paul S. Aisen. "Detection and treatment of Alzheimer's disease in its preclinical stage". en. In: *Nature Aging* 3.5 (May 2023), pp. 520–531. ISSN: 2662-8465. DOI: 10.1038/s43587-023-00410-4.
- [165] G.B. Frisoni et al. "Hippocampal and entorhinal cortex atrophy in frontotemporal dementia and Alzheimer's disease". In: *Neurology* 52.1 (Jan. 1999), pp. 91–91. DOI: 10.1212/WNL.52.1.91.
- [166] Isaac Llorente Saguer and Neil P Oxtoby. "Applying CVR, composite value ratio, to brain MRI volume and thickness as an atrophy biomarker for Alzheimer's disease clinical trials". en. In: *Alzheimer's & Dementia* 19.S16 (2023), e078439. ISSN: 1552-5279. DOI: 10.1002/alz.078439.
- [167] Simon Sheather. *A Modern Approach to Regression with R*. Springer Science & Business Media, 2009.
- [168] Xiuling Nie et al. "Subregional structural alterations in hippocampus and nucleus accumbens correlate with the clinical impairment in patients with Alzheimer's disease clinical spectrum: parallel combining volume and vertex-based approach". In: *Frontiers in neurology* 8 (2017), p. 399.
- [169] Francesca Alves, Pawel Kalinowski, and Scott Ayton. "Accelerated brain volume loss caused by anti- β -amyloid drugs: a systematic review and meta-analysis". In: *Neurology* 100.20 (2023), e2114–e2124.
- [170] Michael W Weiner et al. "Increasing participant diversity in AD research: plans for digital screening, blood testing, and a community-engaged approach in the Alzheimer's Disease Neuroimaging Initiative 4". In: *Alzheimer's & Dementia* 19.1 (2023), pp. 307–317.

- [171] Dieter Petit et al. "A β profiles generated by Alzheimer's disease causing PSEN1 variants determine the pathogenicity of the mutation and predict age at disease onset". en. In: *Molecular Psychiatry* 27.6 (June 2022), pp. 2821–2832. ISSN: 1476-5578. DOI: 10.1038/s41380-022-01518-6.
- [172] Lei Liu et al. "Identification of the A β 37/42 peptide ratio in CSF as an improved A β biomarker for Alzheimer's disease". In: *Alzheimer's & dementia : the journal of the Alzheimer's Association* 19.1 (Jan. 2023), pp. 79–96. ISSN: 1552-5260. DOI: 10.1002/alz.12646.
- [173] Stephanie A Schultz et al. " γ -Secretase activity, clinical features, and biomarkers of autosomal dominant Alzheimer's disease: cross-sectional and longitudinal analysis of the Dominantly Inherited Alzheimer Network observational study (DIAN-OBS)". In: *The Lancet Neurology* 23.9 (Sept. 2024), pp. 913–924. ISSN: 1474-4422. DOI: 10.1016/S1474-4422(24)00236-9.
- [174] Isaac Llorente Saguer, Charles Arber, and Neil P Oxtoby. "Theta, a multidimensional ratio biomarker applied to five amyloid beta peptides for investigations in familial Alzheimer's disease". In: *medRxiv* (2025), pp. 2025–08.
- [175] Isaac Llorente-Saguer et al. "Unbiased data-driven analysis of five amyloid-beta peptides for biomarker investigations in familial Alzheimer's disease". In: *bioRxiv* (2024), pp. 2024–11.
- [176] Antoinette O'Connor et al. "Plasma amyloid- β ratios in autosomal dominant Alzheimer's disease: the influence of genotype". In: *Brain* 144.10 (Apr. 2021), pp. 2964–2970. ISSN: 0006-8950. DOI: 10.1093/brain/awab166.
- [177] Nanet Willumsen et al. "The PSEN1 E280G mutation leads to increased amyloid- β 43 production in induced pluripotent stem cell neurons and deposition in brain tissue". eng. In: *Brain Communications* 5.1 (2023), fcac321. ISSN: 2632-1297. DOI: 10.1093/braincomms/fcac321.
- [178] Charles Arber et al. "The presenilin 1 mutation P436S causes familial Alzheimer's disease with elevated A β 43 and atypical clinical manifestations". eng. In: *Alzheimer's & Dementia: The Journal of the Alzheimer's Association* (June 2024). ISSN: 1552-5279. DOI: 10.1002/alz.13904.

- [179] Nobuhiro Suzuki et al. "An increased percentage of long amyloid β protein secreted by familial amyloid β protein precursor (β APP717) mutants". In: *Science* 264.5163 (1994), pp. 1336–1340.
- [180] Simon Hsu et al. "Systematic validation of variants of unknown significance in APP, PSEN1 and PSEN2". In: *Neurobiology of disease* 139 (2020), p. 104817.
- [181] AlzForum. *PSEN1 E318G — ALZFORUM*. <https://www.alzforum.org/mutations/psen1-e318g>.
- [182] AlzForum. *PSEN1 E69D — ALZFORUM*. <https://www.alzforum.org/mutations/psen1-e69d>.
- [183] Piotr Lewczuk et al. "Neurochemical diagnosis of Alzheimer's dementia by CSF A β 42, A β 42/A β 40 ratio and total tau". In: *Neurobiology of Aging* 25.3 (Mar. 2004), pp. 273–281. ISSN: 0197-4580. DOI: 10.1016/S0197-4580(03)00086-1.
- [184] Prashanthi Vemuri et al. "Tau-PET uptake: Regional variation in average SUVR and impact of amyloid deposition". In: *Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring* 6 (2017), pp. 21–30.
- [185] Alejandro Costoya-Sánchez et al. "Partial volume correction in longitudinal tau PET studies: is it really needed?" In: *NeuroImage* 289 (2024), p. 120537.
- [186] Marc Aurel Busche and Bradley T Hyman. "Synergy between amyloid- β and tau in Alzheimer's disease". In: *Nature neuroscience* 23.10 (2020), pp. 1183–1193.