An *et al. Alzheimer's Research & Therapy* (2025) 17:215
https://doi.org/10.1186/s13195-025-01850-3

Alzheimer's
Research & Therapy

RESEARCH

Open Access

Dissecting real-world memory clinical cohort heterogeneity: analysis of neuroanatomical subtypes using HYDRA

Gordon Zhaoqi An[1,2], Yunchang Xie[1,2], Tammie L. S. Benzinger[2], Brian A. Gordon[2] and Aristeidis Sotiras[2,3*]

Abstract

Background There is significant evidence for neuroanatomical heterogeneity in neurodegenerative disorders, which has been demonstrated predominantly through analyses of well-characterized research cohorts. Despite the known diversity in clinical presentations among patients attending memory clinics, studies exploring neuroanatomical heterogeneity in such clinically diverse groups remain sparse.

Methods To address this gap, we applied the semi-supervised Heterogeneity through Discriminative Analysis (HYDRA) (Neuroimage 145:346-364 2017) machine learning method to magnetic resonance imaging (MRI) data from the Open Access Series of Imaging Studies (OASIS) (NeuroImage 26:102248 2020) to uncover patterns of neurostructural heterogeneity in memory clinic attendees. Cross-validation was used to assess clustering stability via the Adjusted Rand Index (ARI), Silhouette Score, and Calinski–Harabasz Index (CHI). We performed survival analyses using Kaplan–Meier curves and mixed-effects models for longitudinal cognitive data (e.g., memory, executive function, and language assessments) to examine differences in disease progression.

Results Cross-validation analyses indicated two highly stable subtypes of cognitively impaired individuals (ARI = 0.552), exhibiting significant neuroanatomical differences. Subtype 1, termed the Temporal-Sparing Atrophy (TSA) Subtype, was defined by relatively mild atrophy, especially in temporal areas, with slower cognitive decline and preserved Function across most domains. Subtype 2, termed the Temporal-Parietal Predominated Atrophy (TPPA) Subtype, was marked by notable alterations in areas critically affected in neurodegenerative disorders. These included key areas critical for executive function and memory, such as the frontal, temporal, and parietal cortices including the precuneus. Longitudinal analysis of neuroimaging and cognitive data revealed contrasting trajectories. The TSA Subtype demonstrated a gradual decline in cognitive functions over time, particularly in the assessments that are memory-focused tests. Conversely, the TPPA Subtype exhibited a more severe decline in these functions.

Conclusions This research illustrates that neurodegenerative diseases present a spectrum of structural brain changes rather than uniform pathology, suggesting that future research may benefit from stratified therapeutic approaches and targeted recruitment strategies for clinical trials. By leveraging detailed clinical assessments and longitudinal data, including uncertain diagnoses and Clinical Dementia Rating (CDR) scores, this study contributes to better understanding/characterizing memory clinic populations, which could help with optimizing interventions.

Keywords Neurodegeneration, Memory clinic, Dementia, Machine learning, Subtyping, Heterogeneity

*Correspondence:
Aristeidis Sotiras
aristeidis.sotiras@wustl.edu
Full list of author information is available at the end of the article

Background

Neurodegenerative disorders, including Alzheimer's Disease (AD), Parkinson's Disease (PD) [1] and others, exhibit significant heterogeneity, challenging the notion of a "typical" disease presentation [2]. Importantly, clinicians encounter a spectrum of cognitive and neuroanatomical variations among patients, which are often influenced by multiple risk factors and comorbidities. This heterogeneity is critical to understand, as it impacts diagnostic accuracy, disease prognosis, and therapeutic response [3]. Characterizing disease heterogeneity within real-world clinical cohorts by identifying neurobiologically homogeneous clusters of patients can lead to more precise diagnoses and targeted treatments [4]. To this end, neuroimaging provides a reliable and reproducible noninvasive in vivo method to assess the neurobiology of brain disorders, and dementia in particular. However, the high dimensionality and complexity of neuroimaging data present significant analytical challenges.

Over the past few decades, machine learning methods have emerged as an effective way to address these challenges because of their ability to capture complex relationships of imaging signals among brain regions [5]. This capability enhances the characterization of group differences, potentially leading to improved personalized diagnosis and prognosis. Consequently, machine learning has been increasingly successful in driving highly sensitive and specific disease biomarkers at the individual level [6]. For example, Support Vector Machines (SVMs) effectively classified patients based on neuroimaging features [7, 8], while Random Forests predicted disease progression and identified biomarkers [9, 10]. However, a common assumption behind commonly used machine learning methods is that there is a single pattern that differentiates the two contrasted groups. While this approach enables the detection of hallmark pathology, it ignores disease heterogeneity, potentially overlooking distinct neuroanatomical subtypes.

Recent advancements address this limitation by leveraging unsupervised or semi-supervised methods [11–21]. Prior research studies have primarily investigated heterogeneity within well-characterized research cohorts. However, these cohorts do not represent well the variability seen in clinical practice. This limitation is particularly evident in memory clinic populations, where patients present with diverse cognitive profiles and often have uncertain diagnoses. Despite this, investigations into the neurobiological heterogeneity of dementia in clinical settings are scarce. Additionally, while longitudinal cognitive data may offer invaluable insights into disease progression, these are often overlooked in heterogeneity analyses of research or real-world clinical cohorts [22].

In response to these gaps in the field, we investigated the heterogeneity of neurodegeneration using data from a large publicly available sample of patients and their follow-ups seen at the Washington University Memory Diagnostic Center outpatient clinic in Saint Louis, MO by applying the HYDRA (heterogeneity through discriminative analysis) algorithm [23] to the Open Access Series of Imaging Studies (OASIS-4) dataset [24], a clinically representative cohort with neuroimaging and longitudinal clinical data. Specifically, we sought to identify neuroanatomical subtypes using a semi-supervised machine learning method termed HYDRA. HYDRA is a well-established and validated method that has been successfully applied in various neurological and psychiatric disorders, consistently uncovering biologically meaningful and homogeneous clusters within heterogeneous patient populations [25–29]. We hypothesized that HYDRA would objectively reveal distinct neuroanatomical subtypes by analyzing volumetric data. We further hypothesized that the identified subtypes would exhibit distinct clinical and cognitive longitudinal trajectories.

Methods

Dataset description

To implement HYDRA, a patient group and a reference group of controls are needed to identify structural subtypes. Two datasets were used for patients and controls separately for this purpose.

The OASIS-4 dataset was employed to extract patients specifically, sourced from patients seen at the Washington University in St. Louis Memory Diagnostic Center outpatient clinic. This memory clinic cohort provides comprehensive neuroimaging and longitudinal clinical data from individuals with memory disorders and dementia. Patients at various stages of impairment underwent magnetic resonance imaging (MRI) scans and clinical assessments. From this dataset, we included 580 individuals aged 50 to 94 who passed FreeSurfer [30] quality checks. Longitudinal clinical assessments were available for a subset of participants. Though the number and timing of follow-up visits varied, it is a hallmark of real-world clinical data, where patients are assessed as needed rather than at fixed intervals.

We utilized OASIS-3 dataset [31] coming from the Washington University Knight Alzheimer's Disease Research Center (ADRC) – a research-based cohort, which is used to derive a robust reference group of cognitively unimpaired (CU) controls. We identified 536 CU individuals based on a global Clinical Dementia Rating (CDR) [32] of 0 and amyloid negative. Amyloid negativity was defined as the absence of substantial preclinical cognitive impairment pathology, assessed using either cerebrospinal fluid (CSF) biomarkers or amyloid positron

emission tomography (PET) imaging. For amyloid PET, the Centiloid scale [33–35] was used to standardize quantitative amyloid measurements across different tracers and methods. A Centiloid value of ≤ 20 was used as the threshold for amyloid negativity, consistent with the widely accepted cutoff for distinguishing amyloid-negative individuals from those with amyloid-positive pathology [36]. For CSF, amyloid negativity was determined based on the Aβ42/40 ratio that minimizes variability caused by preanalytical factors (e.g., tube adsorption) and interindividual differences in total Aβ production. A threshold of Aβ42/40 < 0.067 was applied to define amyloid negativity, consistent with established protocols for detecting preclinical AD pathology in CSF [37]. In this study, we prioritized amyloid PET data when available, as it provides a direct measure of fibrillar amyloid deposition in the brain. When PET data were unavailable, CSF biomarkers were used as a secondary measure, given their high concordance with amyloid PET and their ability to detect early amyloid changes.

After combining OASIS-3 and OASIS-4, the final dataset comprised 1,116 Subjects with 86 cortical and subcortical volumes from both the left and right hemispheres (i.e., the dimensionality of the data matrix was 1116 × 86).

Analysis of the demographic data (see Table 1) revealed significant differences between the clinical group from OASIS-4 and CU controls from OASIS-3 across several key variables. The CU control group was younger and consisted of a higher proportion of females, while the clinical group was older on average and had a more balanced gender distribution. As expected, due to the way the groups were constructed, significant disparities were evident in clinical assessments: CDR scores range from 0.5 to greater than 1 in the clinical group, indicating varying degrees of cognitive impairment. Similarly, the Mini-Mental State Examination (MMSE) [38] scores differed markedly, with the clinical group showing an average score of 22.9, indicative of cognitive decline, compared to 28.9 in CU controls.

Image acquisition and processing

Our analysis focused on 3 T T1-weighted volumetric MRI data to investigate brain atrophy patterns characteristic of AD. Regional volumes were processed with FreeSurfer [30] version 5.3, based on the Desikan-Killiany atlas [39]. To account for head size variations, total intracranial volume adjustment [40] was applied to normalize regional volumes across participants. Lastly, regional volumes were z-scored using the CU control group as the reference for normalization.

Parsing heterogeneity using HYDRA

To identify neuroanatomical subtypes within memory clinic patients, we used a non-linear machine learning method developed for simultaneous binary classification and subtype identification, termed HYDRA. Unlike traditional SVMs [41], which rely on a single hyperplane to separate cases from controls and cannot distinguish between subtypes of cases, HYDRA clusters cases based on their differences from controls (see Fig. 1).This is achieved by identifying multiple linear hyperplanes that collectively form a convex polytope. This enables

Table 1 Demographics of Memory Clinic Patients (OASIS-4) and Cognitively Unimpaired Controls (OASIS-3)

Characteristic	Clinical Group (OASIS-4), N = 580[1]	Healthy Controls (OASIS-3), N = 536[1]	p-value[2]
Age	72 (SD = 10)	64 (SD = 9)	<0.001
Sex			<0.001
Female	297 (51%)	335 (62%)	
Male	283 (49%)	201 (38%)	
Race			<0.001
African American	56 (9.7%)	86 (16%)	
Other	22 (3.8%)	5 (0.9%)	
White	502 (87%)	445 (83%)	
CDR			<0.001
0.0	66 (11%)	536 (100%)	
0.5	317 (55%)	0 (0%)	
1 or greater	197 (34%)	0 (0%)	
CDR_SB	3.61 (SD = 3.27)	0.00 (SD = 0.00)	<0.001
MMSE	22.9 (SD = 6.2)	28.9 (SD = 3.1)	<0.001

[1] Mean(SD = Standard Deviation); n (%)

[2] Wilcoxon rank sum test; Pearson's Chi-squared test

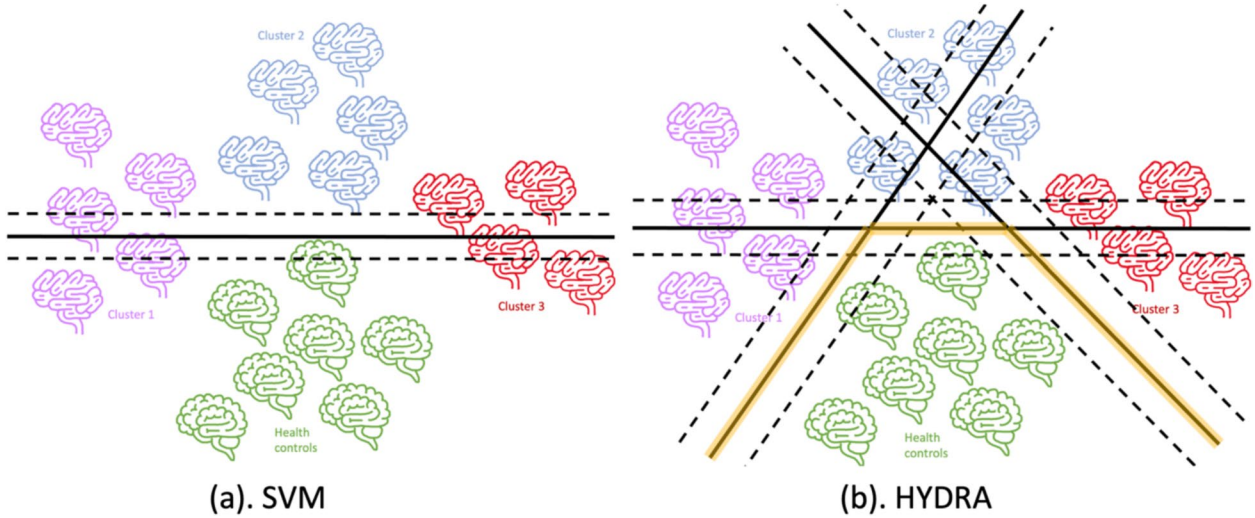

Fig. 1 Schematic representation of the utility of HYDRA over SVMs for unraveling heterogeneity: HYDRA algorithm (**b**) aims to find a maximum margin convex polytope that separates two groups (e.g., healthy controls and pathologic samples). This polytope is formed by combining multiple linear hyperplanes, with each face defining a cluster. Whereas traditional SVMs (**a**), can only define two groups based on a single hyperplane

HYDRA to not only differentiate healthy controls from pathological samples but also identify neuroanatomical subtypes within the pathological group by assigning cases to the different sides of the polytope. As a result, HYDRA uncovers distinct neuroimaging patterns that characterize subtypes and clusters cases by focusing on their differences from control subjects. This is contrast to unsupervised clustering techniques (e.g., k-means), which group participants based on their similarities. Lastly, HYDRA adjusts for important covariates such as age and gender, ensuring robust clustering. Note that HYDRA only uses a binary classification objective, distinguishing healthy controls from patients, to guide the discovery of multiple disease subtypes. Importantly, it does not utilize detailed diagnostic labels (e.g., mild cognitive impairment or AD) or clinical outcomes during training, allowing for the identification of data-driven neuroanatomical patterns.

HYDRA defined neurostructural Subtypes using the z-scored regional volumetric values as input while adjusting for covariates such as age, Gender, and race. Following previous studies that used this approach, we evaluated multiple clustering solutions to identify the optimal number of clusters. Specifically, we tested clusters ranging from 2 to 8, with the goal of selecting the most stable solution. Stability was quantified primarily using the Adjusted Rand Index (ARI) [42], calculated through eightfold cross-validation. A high ARI value indicates a robust and reproducible clustering solution. The solution with the highest ARI value was selected for subsequent analyses. To further evaluate cluster structure, we supplemented ARI with two widely used internal validation metrics which are the Silhouette coefficient and the Calinski–Harabasz Index (CHI). A higher CHI indicates better-defined cluster structure with greater between-cluster dispersion relative to within-cluster compactness. Similarly, the Silhouette coefficient reflects the degree of separation between clusters, with higher values indicating greater distinctiveness.

For more detailed information on the specific parameters and steps used in implementing the HYDRA algorithm, please refer to the supplementary document.

Neuropsychological and cognitive assessments

Cognitive functioning in OASIS-4 participants was assessed using six tests: the Boston Naming Test, the Logical Memory Immediate Recall, the Word Memory Test, the Word Recall Test, the Short Blessed Test, and the Verbal Fluency Test [43–45]. These tests evaluated cognitive domains commonly affected by dementia, such as memory, language, and executive function.

The CDR scale provides a global measure of cognitive and Functional impairment, ranging from 0 (no impairment) to 3 (severe dementia). The MMSE is a brief, standardized measure of global cognitive Function. It assesses orientation, registration, attention and calculation, recall, language, and visuospatial skills, yielding scores from 0 to 30 (higher scores indicate better cognitive performance). We examined baseline MMSE scores to evaluate overall cognitive status and used them as an outcome to compare between the identified subtypes.

Statistical analyses

Differences in demographic and clinical characteristics

To validate the neuroanatomical subtypes identified by the HYDRA algorithm, statistical comparisons of demographic and clinical characteristics between the subtypes were conducted. Categorical variables, such as sex and CDR categories, were analyzed using Pearson's Chi-square test [46]. Continuous variables, including age and MMSE scores, were evaluated using Mann–Whitney U Test, also known as Wilcoxon rank-sum test [47].

Differences in spatial neuroanatomy

Spatial neuroanatomical atrophy patterns were analyzed using effect size measures and visualization techniques. To quantify the extent of atrophy for each subtype, we calculated Cohen's d [48], a standardized measure of effect size that represents the magnitude of atrophy in brain regions relative to CU controls. Higher Cohen's d values indicate greater reductions in regional brain volumes, signifying more pronounced atrophy. To visualize the spatial distribution of brain atrophy, we employed the ggseg [49] R package, which allows for cortical parcellation, and the creation of brain atlases based on standardized regions. This tool enabled us to map the neuroanatomical patterns of atrophy for Subtypes relative to CU controls. All visualizations were produced in the ggseg framework, ensuring consistent and interpretable representations of the spatial data.

Differences in vascular burden

To evaluate vascular burden across subtypes, we analyzed white matter hypo-intensities (WMH) derived from T1-weighted MRI. While Fluid-Attenuated Inversion Recovery (FLAIR) imaging offers a more direct assessment of vascular burden through white matter hyperintensities, it was not utilized in the current analysis due to resource constraints. Therefore, T1-derived WMH volumes were used as a proxy, which have shown reasonable sensitivity to vascular pathology in prior studies of aging and dementia [50, 51]. All WMH volumes were log-transformed to account for skewness and compared between subtypes, controlling for age, sex, and total intracranial volume.

Differences in cognitive and clinical decline over time

To gain deeper insights into the cognitive impact of the neurostructural subtypes, we leveraged the available longitudinal component of our study to assess cognitive decline over time. A comprehensive longitudinal analysis was performed evaluating memory, language, verbal, and executive function to compare trajectories over time. For each neuropsychological assessment and clinical test (e.g., MMSE and CDR Sum of Boxes), we employed mixed-effects models [52] to account for individual variability and the non-independence of repeated measurements within participants, allowing us to track cognitive impairment progression. Two models were used:

Model 1: Included fixed effects for baseline age, sex, race, subtypes, time interval, baseline MMSE and the interaction between interval and subtypes, with a random intercept to account for participant-level variability:

$$Test = \beta_0+\beta_1\times age\ at baseline+\beta_2\times sex+\beta_3\times race+\beta_4\times subtypes+\beta_5\times baseline MMSE+\beta_6\times time interval+\beta_7\times(time interval\times subtypes)+u_i+\epsilon_i$$

Model 2: Extended Model 1 by adding a random slope for the time interval variable to account for individual differences in the rate of cognitive decline:

$$Test = \beta_0+\beta_1\times age\ at\ baseline+\beta_2\times sex+\beta_3\times race+\beta_4\times subtypes+\beta_5\times baseline\ MMSE+\beta_6\times time\ interval+\beta_7\times(time\ interval\ \times subtypes)+u_i+b_i\times time\ interval+\epsilon_i$$

, where:

$u_i \sim N(0, \sigma_u^2)$ is the random intercept for participant i, representing the participant-specific baseline level;$b_i \sim N(0, \sigma_b^2)$ is the random slope for participant i, representing the participant-specific rate of change over time; ϵ_i is the residual error.

Initially, we implemented Model 1 to examine cognitive trajectories. Baseline MMSE has been included as a covariate in the model to address the potential confounding effect of initial cognitive severity. Model 2 extended this framework to analyze temporal progression in longitudinal observations. We then used Akaike information criterion (AIC) [53] and Analysis of Variance (ANOVA) [54] to compare the models for cognitive tests, with p-values adjusted by the False Discovery Rate (FDR) [55].

Differences in clinical progression

Survival analysis was performed to evaluate the time to cognitive impairment progression. We included participants with a baseline CDR of 0.5 and defined clinical progression as the earliest visit with CDR ≥ 1. Survival curves were generated to compare the probability of remaining below this threshold between subtypes and assessed using the log-rank test, and hazard ratios were estimated using Cox proportional hazards models [56], adjusting for relevant covariates such as age, and sex. Confidence intervals were reported to support interpretability.

Sensitivity analyses

To address the potential imbalance in sample sizes between the identified Subtypes, we conducted a sensitivity analysis using a 1:1 matched Subsample. We employed an anti-clustering approach to match the Subtypes on both baseline MMSE and age for participants with longitudinal follow-up. Given the inherent randomness in anti-clustering, we repeated the matching and modeling process across 100 bootstrap iterations. For each iteration, we fit the linear mixed-effects models for six cognitive assessments and computed the difference in the median rate of change between subtypes.

In addition, to examine the impact of irregular follow-up length on longitudinal model estimates, we conducted a secondary sensitivity analysis including only participants with at least three timepoints. The anti-clustering matching procedure on baseline age and MMSE was repeated for this Subset, and the Full 500-iteration bootstrapping pipeline was rerun.

Results

HYDRA identifies two reproducible atrophy subtypes

HYDRA clustered patients based on regional brain volumes after adjusting for covariates (e.g. age, sex etc.). Evaluation of cluster stability using eightfold cross-validation exhibited a well-defined peak at $K=2$ (see Fig. 2, the peak ARI value), indicating the existence of two highly reproducible subtypes (ARI = 0.552) within memory clinic patients. In addition to the ARI, we examined internal cluster validity using the Silhouette coefficient and CHI (see Supplementary Figure S2.). Both metrics supported the selection of two clusters as optimal. Specifically, the CH index peaked at $K=2$ and declined steadily with additional clusters, indicating that the two-cluster solution provided the best separation relative to within-cluster compactness. Similarly, the Silhouette coefficient was highest at $K=2$, suggesting strong clustering quality in terms of compactness and distinctness.

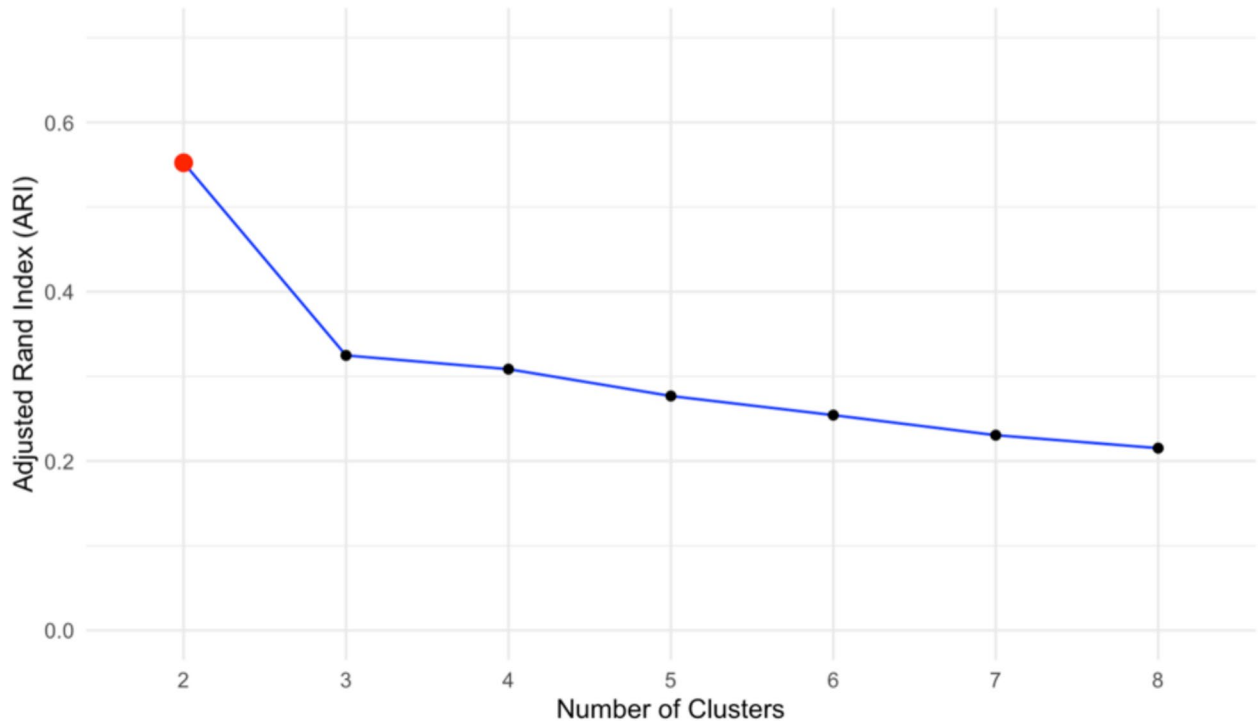

Fig. 2 A high ARI suggests a high-quality clustering result where the algorithm successfully identifies distinct and meaningful subtypes within the data. The figure presents ARI values corresponding to different numbers of clusters (K), obtained with HYDRA. The peak at K = 2 suggests the highest ARI, indicating that the two-cluster solution has the best agreement

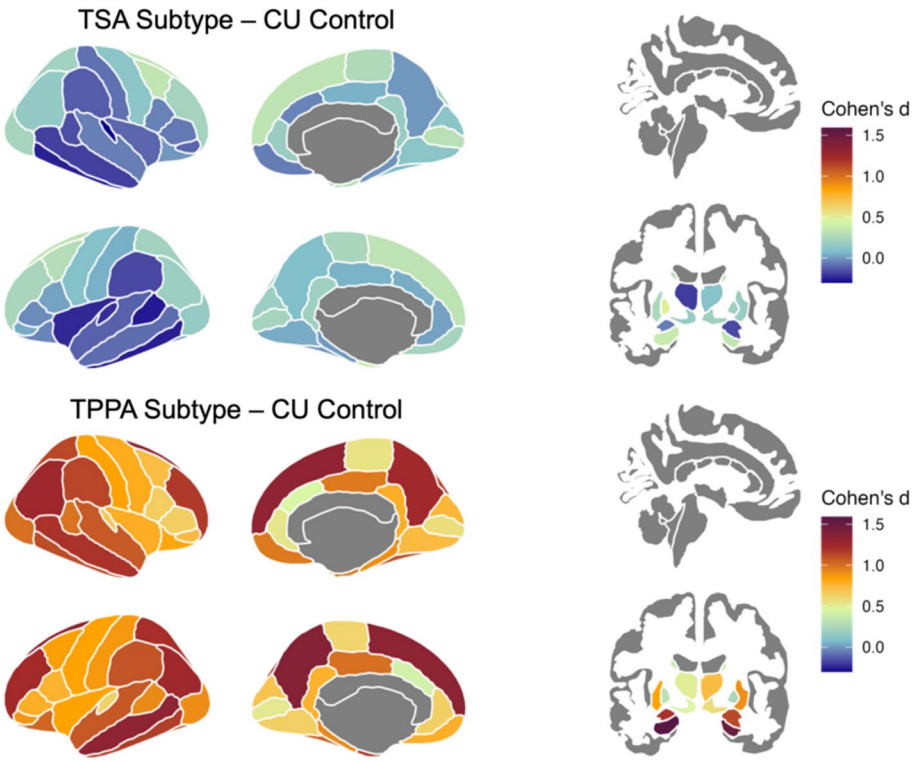

Fig. 3 Neuroanatomical patterns of atrophy: subtype-specific atrophy maps highlight the distinct patterns between the TSA Subtype and the TPPA Subtype compared to CU controls. illustrate regions of significant atrophy, with warmer colors denoting higher effect sizes, which indicates reduced regional volumes when compared with controls. In the TSA Subtype's patterns, the temporal cortex (including the superior, middle, inferior, and transverse temporal gyri), supramarginal gyrus, and insula show minimal atrophy (Cohen's d typically < 0.5). Mild-to-moderate reductions are visible in a few parietal and frontal regions, but the overall pattern remains relatively subtle (blues/greens). Conversely, the patterns of the TPPA Subtype show more extensive atrophy across temporal and parietal cortices, as well as frontal regions, evidenced by a predominance of warmer colors (Cohen's d above 1), which highlights this subtype's more diffuse and severe neurodegenerative changes

Subtypes exhibit divergent neuroanatomical deficits

Upon identifying two stable subtypes of memory clinic patient, we sought to understand which regional volumetric features drove this clustering result. The results revealed distinct neuroanatomical patterns of atrophy for the two subtypes compared to the CU controls (see Fig. 3). Subtype 1 exhibited relatively subtle atrophy, with low effect sizes throughout much of the temporal lobe and precuneus and was therefore termed the Temporal-Sparing Atrophy Subtype (TSA) Subtype. In contrast, Subtype 2 demonstrated more widespread atrophy, particularly in the frontal, parietal, and temporal lobes, and was therefore termed the Temporal-Parietal Predominated Atrophy (TPPA) Subtype.

Subtype demographics

Next, we evaluated the demographics of the neurostructural subtypes (see Table 2). The TSA Subtype comprised 150 participants with a mean age of 71 (SD = 11), while the TPPA Subtype included 430 participants with a mean age of 73 (SD = 9). While there was substantial overlap in age ranges among both subtypes, the TPPA Subtype had a slightly higher mean and median age than the TSA Subtype. Consequently, the two atrophy subtypes exhibited statistically significant differences in age distribution. When comparing with control group, both subtypes are generally older than the control group, with the TPPA Subtype showing the highest median age (see Supplementary Figure S1).

Atrophy subtypes demonstrate distinct clinical and cognitive profiles at baseline

We then evaluated whether the subtypes differed in baseline cognitive performance. Both subtypes exhibited notable differences in CDR and cognitive function. As depicted in Fig. 4, the TSA Subtype was primarily Characterized by very mild dementia, with most participants having CDR scores of 0.5. However, this subtype also included a smaller proportion (16%) of individuals with a CDR score of 1 or greater. In contrast, participants in the TPPA Subtype were likely to have mild to moderate

Table 2 Subtype differences on demographics

Characteristic	TSA Subtype, N = 150[1]	TPPA Subtype, N = 430[1]	p-value[2]
Age	71 (SD = 11)	73 (SD = 9)	0.016
Sex			0.3
Female	71 (47%)	226 (53%)	
Male	79 (53%)	204 (47%)	
Race			<0.001
African American	8 (5.3%)	48 (11%)	
Other	16 (11%)	6 (1.4%)	
White	126 (84%)	376 (87%)	
CDR_SB	2.3 (SD = 2.4)	4.1 (SD = 3.4)	<0.001
Verbal Fluency test	16.1 (SD = 4.6)	13.0 (SD = 5.2)	<0.001
Boston Naming test	13.88 (SD = 1.72)	13.09 (SD = 2.28)	<0.001
Word Memory test	17.0 (SD = 5.3)	12.9 (SD = 5.2)	<0.001
Logical Memory test	7.5 (SD = 4.4)	5.5 (SD = 3.9)	<0.001
Word Recall test	5.19 (SD = 2.75)	3.05 (SD = 2.66)	<0.001
Short Blessed test	4 (SD = 4)	9 (SD = 8)	<0.001

[1] Mean(SD = Standard Deviation); n (%)

[2] Wilcoxon rank sum test; Pearson's Chi-squared test

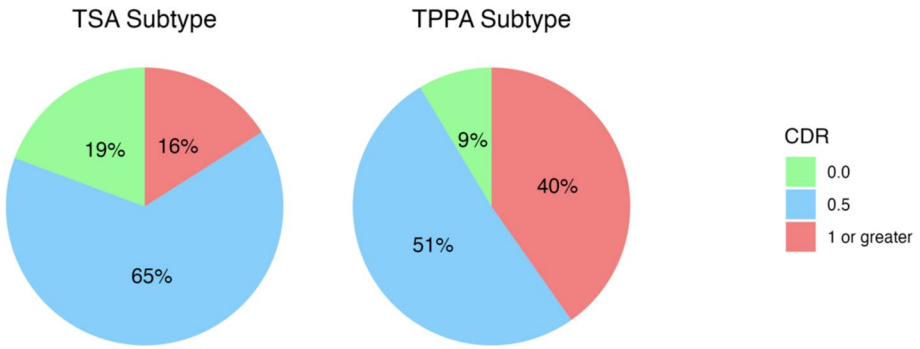

Fig. 4 Subtype characterization using CDR: the TSA Subtype patients mostly exhibited CDR scores of 0.5 (65%), Suggesting very mild dementia. In contrast, the TPPA Subtype patients were Characterized by a higher prevalence of CDR scores of 1 or greater (40%), indicative of mild to moderate dementia

dementia, with 40% having CDR scores of 1 or greater. Both subtypes consisted of individuals across the spectrum of CDR scores, further reflecting the heterogeneity within each group, indicating that the subtypes are not exclusively tied to a single clinical stage of dementia.

The violin plot in Fig. 5 illustrates the distribution of MMSE scores for the two subtypes, revealing distinct cognitive profiles. The TSA Subtype was characterized by a more concentrated range of MMSE scores, indicating consistent cognitive impairment within this group. In contrast, the TPPA Subtype was marked by a wider distribution of MMSE scores, reflecting greater variability in cognitive function. Notably, some individuals in the TPPA Subtype retained relatively better cognitive function, while others experienced more severe impairments. The significant difference in MMSE scores between the two subtypes reinforces the idea that these subtypes represent distinct severities of dementia, with the TPPA Subtype displaying more pronounced cognitive deficits.

Vascular distribution difference between the subtypes

To examine potential vascular contributions to subtype differentiation, WMH volumes have been utilized. After

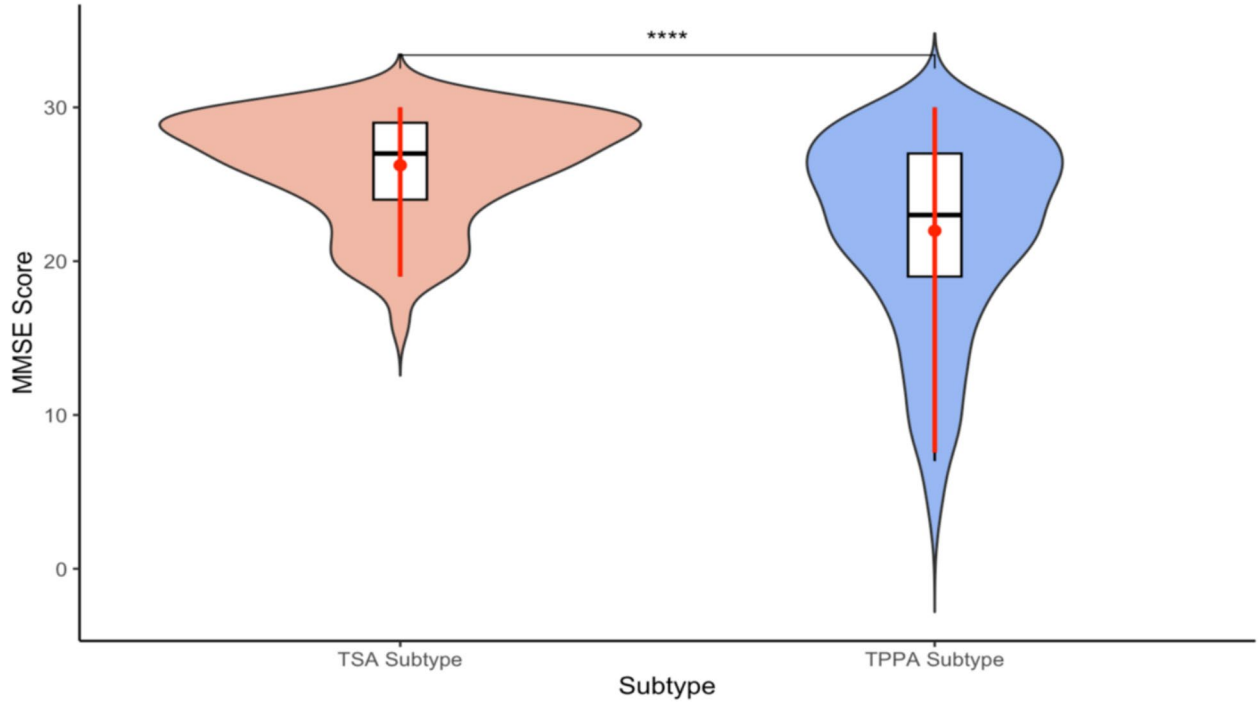

Fig. 5 Comparison of MMSE scores between subtypes: the distribution of MMSE scores across the two atrophy subtypes is shown. The TSA Subtype (red) exhibits a more concentrated distribution of MMSE scores. The TPPA Subtype (blue) is characterized by a wider spread of MMSE scores, suggesting greater variability in cognitive function. The boxplot within each violin represents the interquartile range and median MMSE score, with a significant difference between the two subtypes (***$p < 0.001$). Confidence interval is reported (red line)

log transformation and adjustment for age, sex, and intracranial volume, significant differences are revealed in WMH burden between subtypes. The TPPA Subtype exhibited higher WMH volume compared to the TSA Subtype (see Supplementary Figure S3), suggesting a greater vascular burden in this group. These findings indicate that vascular pathology may disproportionately contribute to the neurodegenerative profile observed in the TPPA group, potentially accelerating cognitive decline or compounding the effects of neuroanatomical atrophy.

Longitudinal cognitive and clinical decline differentiates atrophy subtypes

To assess cognitive differences across subtypes over time, we analyzed longitudinal trajectories using standardized tests of memory, language, verbal fluency, and executive function. The number of follow-up visits and the average time interval from baseline to follow-up were well-balanced between subtypes, as summarized in Supplementary Table S1.

A formal comparison of two models (random intercept only vs. random intercept with random slope) is provided in Supplementary Table S2., indicating that the inclusion of random slopes offers an enhanced ability to capture individual trajectories of decline ($p < 0.05$). Furthermore, the significant effects of baseline age, subtype classification, and their interactions on the progression of cognitive decline have been shown in Supplementary Table S3., highlighting the importance of demographic factors and neuroanatomical subtype in shaping long-term outcomes.

Turning to the rate of change in clinical assessment scores from the latter model, statistical analysis revealed significant differences in rate of change between subtypes across all tests (****$p < 0.0001$).

In general, the TPPA Subtype exhibited a wider distribution and steeper rates of change (more positive for the Short Blessed test and more negative for the other tests), signifying faster cognitive deterioration (see Fig. 6(a)). However, it is notable that the Verbal Fluency test showed worse performance in the TSA Subtype compared to the TPPA Subtype. To contextualize these rates within a temporal framework, the longitudinal trajectories of clinical test scores over a 10-year span for both subtypes are visualized (see Fig. 6(b)). The TPPA Subtype demonstrated a steeper decline across most tests, particularly those assessing memory (Word Memory, Word Recall), language function (Boston Naming), and executive function (Logical Memory). Conversely, the TSA

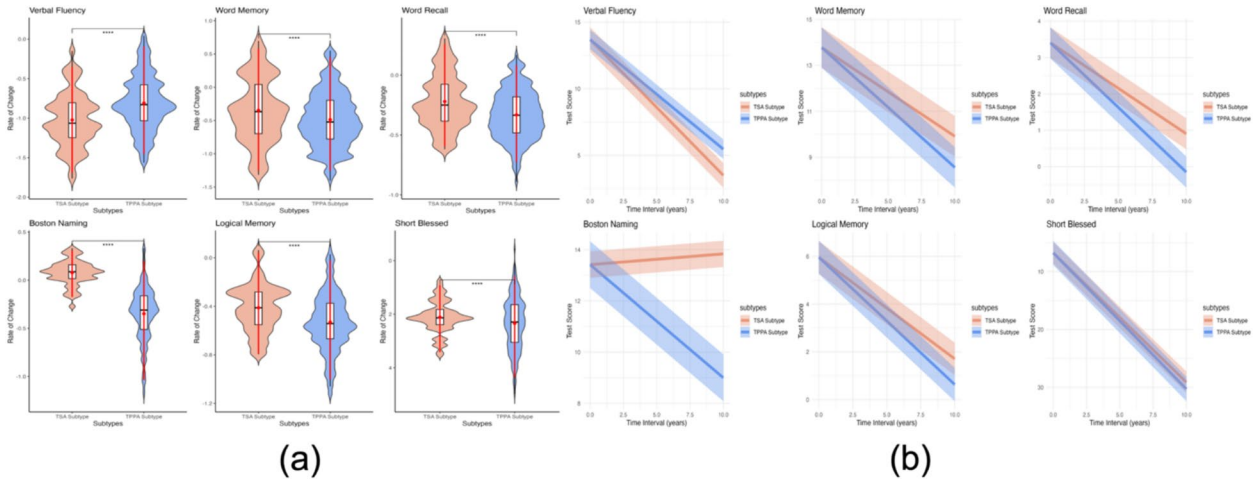

Fig. 6 Visualizations of cognitive decline rates: **a**: Violin plots highlight the variability within subtypes, with the TPPA Subtype showing broader distributions and more pronounced rates of decline across most tests, emphasizing the severity of cognitive deterioration in this group. **b**: Longitudinal trajectories reveal consistent differences between subtypes, with the TPPA Subtype declining more rapidly. Notably, the divergence is particularly evident in memory and executive function tests, suggesting these domains are most impacted by the subtype-specific atrophy patterns

Subtype exhibited a slower, more gradual decline, potentially reflecting a less aggressive disease course or different diseases feeding into the subtype. Despite this overall trend, the TSA Subtype's poorer performance on the Verbal Fluency Test highlights a potential domain-specific deficit that warrants further investigation. In contrast, when analyzing longitudinal changes in global clinical assessments such as CDR Sum of Boxes and MMSE, we found no statistically significant difference in the rate of clinical progression between the two subtypes (see Supplementary Figure S4).

The difference in the rates of clinical progression between the subtypes

In contrast to the clear subtype differences observed in neuropsychological test trajectories, the survival curves showed no statistically significant difference in time to clinical progression between the TSA and TPPA subtypes (see Supplementary Figure S5), aligned with the rate of clinical progression results in CDR Sum of Boxes and MMSE.

Sensitivity analyses for longitudinal studies

To evaluate whether the observed differences in cognitive decline between subtypes were robust to sampling bias, we conducted a sensitivity analysis using age and MMSE matched subsamples created with anti-clustering ($n=66$ per subtype at baseline visit). The process was repeated across 500 iterations, each time estimating the difference in rate of change across six cognitive domains using linear mixed-effects models (see Fig. 7). Consistent with the main findings from the full sample, the TPPA Subtype showed faster decline across most domains (i.e., negative difference in rate of change). The exception was Verbal Fluency, where the TSA Subtype showed slightly worse trajectories, replicating the paradox noted in the primary analysis. As expected, the Short Blessed score (where higher values indicate worse performance) exhibited a positive difference, indicating a steeper rate of decline in the TPPA Subtype. Notably, the Short Blessed Test showed considerable variability across bootstrap iterations. This widespread likely reflects multiple contributing factors, including the smaller number of repeated observations, the relatively coarse scoring structure of the instrument, and its limited sensitivity to subtle longitudinal change. Given its primary role as a screening tool rather than a nuanced progression measure, the broader distribution is expected. Nonetheless, the overall directionality of subtype differences remained preserved.

We also conducted a secondary sensitivity analysis focusing on participants with at least three longitudinal visits (see Supplementary Figure S6). The results from this subgroup were broadly consistent with those from the full sample. While the average differences in rate of change between subtypes were somewhat attenuated, the directionality of effects across all cognitive tests remained stable, supporting the robustness of our findings despite variation in follow-up duration.

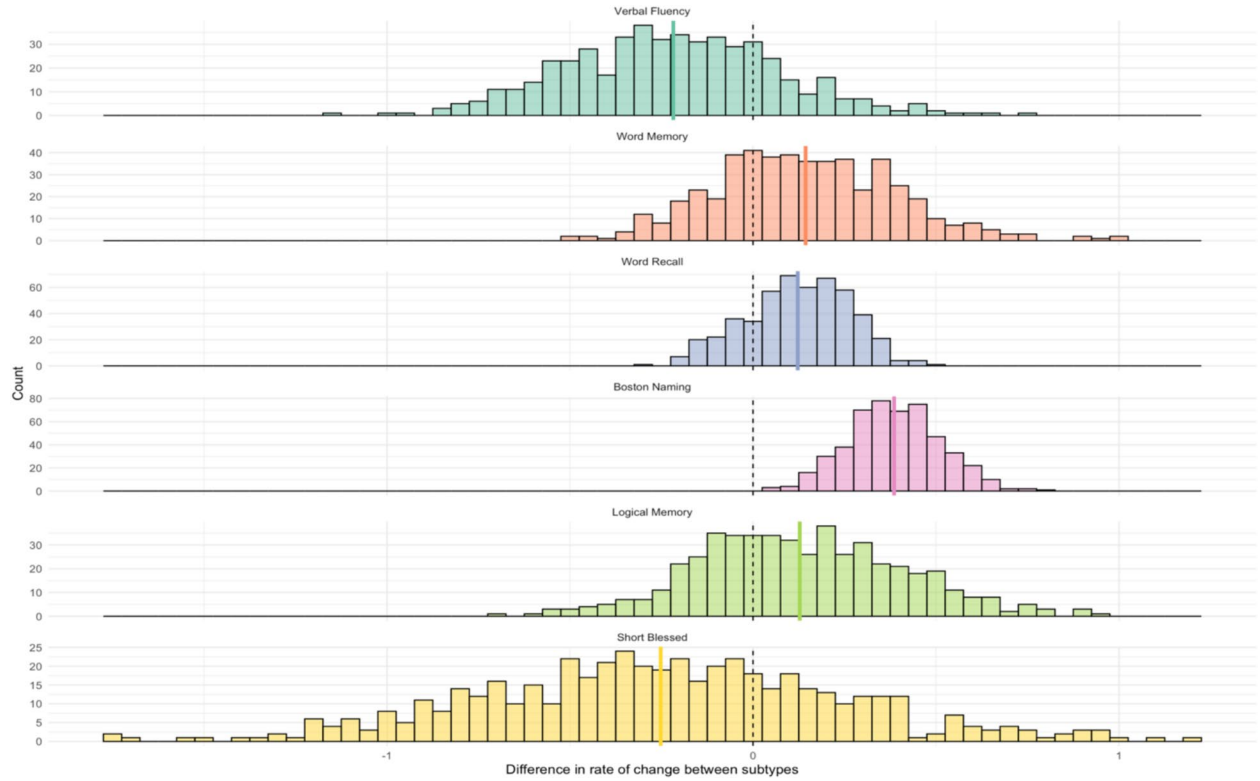

Fig. 7 Rate of decline on clinical assessments after matching the baseline MMSE or age: Each histogram reflects the distribution of the difference in median rate of cognitive change between the TSA and TPPA subtypes (TSA minus TPPA) across 100 iterations using matched subsamples. Dashed vertical lines indicate a null difference (x = 0), and the colored vertical lines marks the average difference in each test. For most domains, a negative value indicates steeper decline in the TPPA subtype. In contrast, positive values for Verbal Fluency and Short Blessed reflect the paradoxical pattern where the TSA subtype shows comparable or steeper decline in verbal fluency, and the TPPA subtype performs worse on Short Blessed (where higher scores reflect more impairment). These results confirm the reproducibility of our subtype findings under sampling-balanced conditions

Discussion

This study's application of the HYDRA algorithm to the OASIS-4 memory clinic cohort has successfully segmented patients into two neuroanatomically meaningful subtypes, each characterized by unique structural and cognitive patterns. It characterized heterogeneity of dementia in memory clinics, building upon previous studies by examining a larger sample and longitudinal cognitive and clinical data.

Neuroanatomical and cognitive variations between subtypes

The TSA Subtype, referred to as the Temporal-Sparing Atrophy Subtype, demonstrated relatively mild volumetric reductions overall, with notably lower effect sizes in temporal areas and precuneus. These are areas crucial for memory processing and especially for episodic memory [57]. In other words, within this subtype's atrophy pattern, the temporal lobes appear comparatively "spared," while subtler atrophy was observed in a few parietal or frontal regions. This pattern could explain the milder cognitive impairments observed in this group, as reflected in their relatively preserved CDR and MMSE scores (Table 2). Taken together, these results suggest that the TSA Subtype might encompass an early stage or a less aggressive form of dementia, potentially responsive to early intervention strategies that could delay cognitive progression. The milder cognitive impairments seen in the TSA Subtype suggest the possibility of extending the window for therapeutic intervention, especially for memory-related cognitive functions.

The TPPA Subtype, referred to as the Temporal-Parietal Predominated Atrophy Subtype, demonstrated widespread atrophy affecting the frontal, parietal, and temporal lobes. These regions are integral to higher cognitive functions, which correlates with the more severe cognitive deficits [57, 58], reflected in lower MMSE scores and higher CDR scores. The extensive

neurodegeneration in the TPPA Subtype suggests a more advanced or rapidly progressing form of dementia, highlighting the critical regions affected in these patients. The pronounced atrophy in multiple brain regions suggests a need for more intensive clinical management and highlights the potential for future research to focus on slowing the progression of degeneration in these critical areas.

Potential mechanisms

The differential decline rates between the subtypes may reflect underlying differences in neurobiological processes. The extensive atrophy observed in the TPPA Subtype suggests greater involvement of networks critical for executive function and memory, such as the frontoparietal and default mode networks. This may explain the faster cognitive deterioration and broader functional impairments in this group. Conversely, the TSA Subtype's slower progression might be attributed to the preservation of critical neural pathways, potentially delaying the onset of severe impairments.

In addition to structural and cognitive differences, we observed a significant disparity in vascular burden between subtypes. The TPPA Subtype exhibited higher T1-derived WMH volumes, suggesting greater vascular risk in this group. This vascular burden may compound the effects of neurodegeneration and contribute to the accelerated clinical progression seen in the TPPA Subtype. Importantly, because our clustering was based solely on normalized gray matter measures that directly capture neurodegeneration and can be widely obtained in clinical MRI, WMH differences were not part of the model input but instead represent an additional biological dimension that may interact with atrophy patterns. It supports the idea that subtype-specific disease trajectories may arise from a combination of neurodegenerative and vascular mechanisms. Collectively, these findings align with the hypothesis [59] that neurodegenerative processes can follow distinct pathways depending on the affected neural circuits, further highlighting the importance of stratified research into neurodegenerative mechanisms.

Addressing heterogeneity in clinical cohorts

A major strength of this study lies in its emphasis on real-world clinical cohorts, which inherently display significant heterogeneity compared to well-characterized research populations. By addressing the diversity in neuroanatomical presentations, this study bridges the gap between controlled research environments and clinical practice. This heterogeneity reflects the challenges faced by clinicians in diagnosing and managing patients with varying degrees and patterns of impairmentand may support future efforts to tailor research and care pathways to different subgroups.

Insights from longitudinal analysis

The longitudinal analyses and the sensitivity tests revealed that cognitive decline does not follow a uniform trajectory across the dementia spectrum but varies significantly between the subtypes identified. Incorporating mixed-effects models allowed us to capture the nuances of how age and disease duration impact cognitive decline. These models indicated that the cognitive trajectory in the TPPA Subtype is steeper, aligning with its broader anatomical impact, while the TSA Subtype experienced a more gradual decline, consistent with its relatively low effect sizes in temporal regions and the precuneus, areas critical for memory processing and episodic recall. Interestingly, the Verbal Fluency Test results in the TSA Subtype, which were worse than those in the TPPA Subtype, point to a domain-specific vulnerability that stands apart from the otherwise milder pattern of global decline. This suggests that even "less aggressive" atrophy profiles may disproportionately affect certain cognitive circuits supporting lexical retrieval and language production, warranting hypothesis-driven investigation in future studies.

Complementing these cognitive findings, we also examined global clinical progression using longitudinal changes in CDR-Sum of Boxes and MMSE, as well as Cox proportional hazards survival models restricted to individuals with baseline CDR = 0.5. Notably, none of these global metrics revealed significant differences in progression rates between subtypes. This apparent dissociation between cognitive test decline and global clinical staging suggests that while the TPPA subtype exhibits more rapid domain-specific deterioration, such changes may not be fully captured by global staging tools like the CDR within the available follow-up window. These findings underscore the limitations of global clinical scales in reflecting the nuanced progression patterns of biologically distinct subtypes and highlight the need for more sensitive or domain-specific outcome measures in memory clinic cohorts.

Clinical implications and machine-learning perspectives

Recognizing heterogeneity in memory clinic populations may eventually contribute to improved diagnostic refinement and stratified care pathways. By identifying distinct subtypes based on neuroanatomical patterns, this study provides a framework for better understanding the underlying complexity of dementia, including variable clinical presentations and trajectories. The distinct rates of cognitive decline observed between subtypes reinforce the value of tailored approaches in both

routine assessments and longer-term monitoring of disease progression.

In parallel, our findings underscore the power of machine-learning methods, particularly semi-supervised algorithms like HYDRA, in parsing hidden heterogeneity. Unlike traditional classification models that assume a single disease pattern, HYDRA leverages multiple "margins" to identify unique atrophy profiles within a clinically diverse cohort. This approach is especially beneficial in real-world memory clinics, where patients may have a range of pathologies and disease trajectories beyond a single, canonical syndrome. Beyond simply detecting subtypes, machine learning (ML)-driven analyses can illuminate the specific neuroanatomical markers that differentiate these subgroups, guiding clinicians toward more nuanced and data-driven diagnostic insights. Nevertheless, the clinical utility of such clustering hinges upon robust validation, both biologically (e.g., with pathology biomarkers) and clinically (e.g., through longitudinal functional outcomes). As multimodal datasets become more common, ongoing advances in machine-learning algorithms will help integrate imaging, fluid, genetic, and functional data, further supporting more granular models of dementia progression and informing future translational research.

Limitations

While this study provides significant insights into the heterogeneity of dementia, several limitations must be acknowledged. First, because AD is the most common cause of dementia, having biomarker data for PET imaging, genetic markers, or fluid biomarkers would be critical for confirming or excluding AD pathology in our cohort [60]. However, the OASIS-4 dataset does not provide these biomarkers, which limits our ability to ascertain whether the identified subtypes map directly onto AD or other neurodegenerative processes. This constraint may affect the precision of our subtype classification and the overall robustness of our findings. Moreover, as our clustering approach relied on single modality design (i.e., T1-weighted structural MRI), the resulting subtypes capture dominant anatomical variation but may lack the sensitivity to differentiate Alzheimer's disease from other non-AD pathologies Such as frontotemporal lobar degeneration, vascular dementia, or mixed etiologies. Although HYDRA is designed to capture dominant neuroanatomical axes of variation, the lack of multimodal data restricts our ability to resolve overlapping atrophy patterns that may arise from distinct underlying mechanisms. Lastly, although approximately 10% of participants were African American, this remains a single-site dataset, potentially restricting generalizability. Studies with more diverse, multi-site cohorts are needed to confirm the external validity of these findings and refine these subtypes accordingly.

Future research

Future studies should incorporate multimodal data, including Amyloid and Tau biomarkers, to validate the subtypes identified and better understand the biological underpinnings of the distinct trajectories observed. Investigating molecular and genetic markers that define or influence these subtypes could help elucidate individual risk profiles and heterogeneity. Additionally, longitudinal analyses combining imaging, genetic, and cognitive data may provide deeper insights into the mechanisms driving subtype-specific progression, informing future efforts toward personalized care and stratified clinical trial designs.

Conclusions

This study underscores the critical importance of addressing heterogeneity in clinical cohorts for understanding neurodegenerative diseases. By applying the HYDRA algorithm to the OASIS-4 dataset, we identified two distinct subtypes of memory clinic populations based on neuroanatomical atrophy patterns and cognitive decline trajectories. These subtypes highlight the substantial variability in clinical presentations, reflecting the diversity seen in real-world clinical settings.

The TSA Subtype, characterized by more localized atrophy and slower cognitive decline, represents a less aggressive form of dementia, potentially offering a larger window for early intervention. In contrast, the TPPA Subtype, with widespread atrophy and faster cognitive decline, underscores the need for more intensive clinical management. These findings demonstrate the value of recognizing and analyzing heterogeneity within clinical cohorts, moving beyond traditional one-size-fits-all approaches.

Importantly, although clear subtype differences were observed in domain-specific cognitive trajectories, these did not consistently translate into differences in global clinical progression as captured by summary metrics such as CDR or MMSE. This dissociation highlights the limitations of relying solely on global staging tools to track disease progression in memory clinic populations. Future efforts should prioritize more sensitive and domain-specific measures to better capture the true extent and variability of cognitive decline across distinct subtypes.

The study's emphasis on real-world clinical populations ensures that its findings are directly relevant to clinical practice and may inform more representative sampling and stratification strategies in future clinical trials. By acknowledging and addressing heterogeneity, this study contributes to the evolving landscape of neurodegenerative disorder research, offering actionable insights for future studies seeking to optimize diagnosis, monitoring, and intervention development.

Abbreviations

AD	Alzheimer's Disease
PD	Parkinson's Disease
MRI	Magnetic Resonance Imaging
PET	Positron Emission Tomography
FLAIR	Fluid-Attenuated Inversion Recovery
CDR	Clinical Dementia Rating
MMSE	Mini-Mental State Examination
OASIS	Open Access Series of Imaging Studies
HYDRA	Heterogeneity through Discriminative Analysis
SVM	Support Vector Machine
ARI	Adjusted Rand Index
CSF	Cerebrospinal Fluid
CU	Cognitively Unimpaired
FDR	False Discovery Rate
TSA	Temporal-Sparing Atrophy (Subtype)
TPPA	Temporal-Parietal Predominated Atrophy (Subtype)
CHI	Calinski–Harabasz Index
LME model	Linear Mixed Effect model
ML	Machine Learning
WMH	White Matter Hypointensity
AIC	Akaike Information Criterion
ANOVA	Analysis of Variance

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s13195-025-01850-3.

Supplementary Material 1.

Acknowledgements
Data was provided by OASIS-3: Longitudinal Multimodal Neuroimaging: Principal Investigators: T. Benzinger, D. Marcus, J. Morris; NIH P30 AG066444, P50 AG00561, P30 NS09857781, P01 AG026276, P01 AG003991, R01 AG043434, UL1 TR000448, R01 EB009352. AV-45 doses were provided by Avid Radiopharmaceuticals, a wholly owned subsidiary of Eli Lilly, and OASIS-4: Clinical Cohort: Principal Investigators: T. Benzinger, L. Koenig, P. LaMontagne.

Authors' contributions
G.A. led the project and contributed to the majority of the implementation for the study and manuscript writing. Y.X. contributed to the longitudinal studies and manuscript writing. T.B. contributed to the acquisition and curation of the datasets (OASIS). B.G. provided clinical expertise and contributed to manuscript revision. A.S. conceived and supervised the project and contributed to manuscript revision. All authors read, reviewed, and approved the final manuscript.

Funding
This work was supported by the National Institutes of Health (R01-AG067103). We thank the staff of the Washington University Center for High Performance Computing for enabling the computational aspects of this project. Computations were performed in the Washington University Research Computing and Informatics Facility, which received funding from NIH S10 program grants: 1S10OD025200-01A1 and 1S10OD030477-01.

Data availability
The OASIS-3 dataset and the OASIS-4 dataset are publicly available. The link to the data access form is: https://sites.wustl.edu/oasisbrains/home/access/.

Declarations

Ethics approval and consent to participate
Ethics approvals were obtained by the OASIS-3 dataset and the OASIS-4 dataset. All participants were consented into Knight ADRC-related projects in accordance with the Declaration of Helsinki and following procedures approved by the Institutional Review Board of Washington University School of Medicine in St. Louis. For more details, we refer the reader to the OASIS-3 reference and the OASIS-4 reference.

Consent for publication
Not applicable.

Competing interests
Author A.S. holds equity in TheraPanacea and has received personal compensation for serving as a grant reviewer for BrightFocus Foundation. All other authors declare no conflicts of interest.

Author details
[1]Division of Computational and Data Sciences, Washington University in St. Louis, St. Louis, USA. [2]Mallinckrodt Institute of Radiology, Washington University School of Medicine, St. Louis, USA. [3]Institute for Informatics, Data Science and Biostatistics, Washington University School of Medicine, St. Louis, USA.

Received: 25 March 2025 Accepted: 15 August 2025
Published online: 01 October 2025

References

1. Greenland JC, Williams-Gray CH, Barker RA. The clinical heterogeneity of Parkinson's disease and its therapeutic implications. Eur J Neurosci. 2019;49(3):328–38. https://doi.org/10.1111/ejn.14094.
2. Murray ME, Graff-Radford NR, Ross OA, Petersen RC, Duara R, Dickson DW. Neuropathologically defined subtypes of Alzheimer's disease with distinct clinical characteristics: a retrospective study. Lancet Neurol. 2011;10(9):785–96. https://doi.org/10.1016/S1474-4422(11)70156-9. Epub 2011 Jul 27. PMID: 21802369; PMCID: PMC3175379.
3. Balasa M, Gelpi E, Antonell A, Sánchez-Valle R, Molinuevo JL, Lladò A. Clinical features and APOE genotype of pathologically proven early-onset Alzheimer disease. Neurology. 2011;76:1720–5.
4. Insel TR, Cuthbert BN. Medicine. Brain disorders? Precisely. Science. 2015;348(6234):499–500. https://doi.org/10.1126/science.aab2358.
5. Davatzikos C. Machine learning in neuroimaging: progress and challenges. Neuroimage. 2019;197:652–6. https://doi.org/10.1016/j.neuroimage.2018.10.003. Epub 2018 Oct 6. PMID: 30296563; PMCID: PMC6499712.
6. Rathore S, Habes M, Iftikhar MA, Shacklett A, Davatzikos C. A review on neuroimaging-based classification studies and associated feature extraction methods for Alzheimer's disease and its prodromal stages. Neuroimage. 2017;155:530–48. https://doi.org/10.1016/j.neuroimage.2017.03.057. Epub 2017 Apr 13. PMID: 28414186; PMCID: PMC5511557.
7. Klöppel S, Stonnington CM, Chu C, Draganski B, Scahill RI, Rohrer JD, Fox NC, Jack CR Jr, Ashburner J, Frackowiak RS. Automatic classification of MR scans in Alzheimer's disease. Brain. 2008;131(Pt 3):681–9. https://doi.org/10.1093/brain/awm319. Epub 2008 Jan 17. PMID: 18202106; PMCID: PMC2579744.
8. Magnin B, Mesrob L, Kinkingnéhun S, Pélégrini-Issac M, Colliot O, Sarazin M, Dubois B, Lehéricy S, Benali H. Support vector machine-based classification of Alzheimer's disease from whole-brain anatomical MRI. Neuroradiology. 2009;51(2):73–83. https://doi.org/10.1007/s00234-008-0463-x.
9. Hothorn T, Jung HH. RandomForest4Life: a random forest for predicting ALS disease progression. Amyotroph Lateral Scler Frontotemporal Degener. 2014;15(5–6):444–52. https://doi.org/10.3109/21678421.2014.893361.

10. Khalilia M, Chakraborty S, Popescu M. Predicting disease risks from highly imbalanced data using random forest. BMC Med Inform Decis Mak. 2011;11: 51. https://doi.org/10.1186/1472-6947-11-51.
11. Dong A, Honnorat N, Gaonkar B, Davatzikos C. Chimera: clustering of heterogeneous disease effects via distribution matching of imaging patterns. IEEE Trans Med Imaging. 2016;35(2):612–21. https://doi.org/10.1109/TMI.2015.2487423.
12. Ten Kate M, Dicks E, Visser PJ, van der Flier WM, Teunissen CE, Barkhof F, Scheltens P, Tijms BM, Alzheimer's Disease Neuroimaging Initiative. Atrophy subtypes in prodromal Alzheimer's disease are associated with cognitive decline. Brain. 2018;141(12):3443–56. https://doi.org/10.1093/brain/awy264. PMID: 30351346; PMCID: PMC6669409.
13. Zhang X, Mormino EC, Sun N, Sperling RA, Sabuncu MR, Yeo BTT. Bayesian model reveals latent atrophy factors with dissociable cognitive trajectories in Alzheimer's disease. Proc Natl Acad Sci U S A. 2016;113(42):E6535–44. https://doi.org/10.1073/pnas.1611073113.
14. Yang Z, Nasrallah IM, Shou H, Wen J, Doshi J, Habes M, Erus G, Abdulkadir A, Resnick SM, Albert MS, Maruff P, Fripp J, Morris JC, Wolk DA, Davatzikos C, iSTAGING Consortium, Baltimore Longitudinal Study of Aging (BLSA), Alzheimer's Disease Neuroimaging Initiative (ADNI). A deep learning framework identifies dimensional representations of Alzheimer's disease from brain structure. Nat Commun. 2021;12(1):7065. https://doi.org/10.1038/s41467-021-26703-z.
15. Habes M, Grothe MJ, Tunc B, McMillan C, Wolk DA, Davatzikos C. Disentangling heterogeneity in Alzheimer's disease and related dementias using data-driven methods. Biol Psychiatry. 2020;88(1):70–82. https://doi.org/10.1016/j.biopsych.2020.01.016. Epub 2020 Jan 31. PMID: 32201044; PMCID: PMC7305953.
16. Wen J, Antoniades M, Yang Z, Hwang G, Skampardoni I, Wang R, Davatzikos C. Dimensional neuroimaging endophenotypes: neurobiological representations of disease heterogeneity through machine learning. Biol Psychiatry. 2024;96(7):564–84. https://doi.org/10.1016/j.biopsych.2024.04.017.
17. Verdi S, Marquand AF, Schott JM, Cole JH. Beyond the average patient: how neuroimaging models can address heterogeneity in dementia. Brain. 2021;144(10):2946–53. https://doi.org/10.1093/brain/awab165. PMID: 33892488; PMCID: PMC8634113.
18. Noh Y, Jeon S, Lee JM, Seo SW, Kim GH, Cho H, Ye BS, Yoon CW, Kim HJ, Chin J, Park KH, Heilman KM, Na DL. Anatomical heterogeneity of Alzheimer disease: based on cortical thickness on MRIs. Neurology. 2014;83(21):1936–44. https://doi.org/10.1212/WNL.0000000000001003. Epub 2014 Oct 24. PMID: 25344382; PMCID: PMC4248459.
19. Park JY, Na HK, Kim S, Kim H, Kim HJ, Seo SW, Na DL, Han CE, Seong JK, Alzheimer's Disease Neuroimaging Initiative. Robust identification of Alzheimer's disease subtypes based on cortical atrophy patterns. Sci Rep. 2017;7:43270. https://doi.org/10.1038/srep43270. PMID: 28276464; PMCID: PMC5343676.
20. Poulakis K, Pereira JB, Mecocci P, Vellas B, Tsolaki M, Kłoszewska I, Soininen H, Lovestone S, Simmons A, Wahlund LO, Westman E. Heterogeneous patterns of brain atrophy in Alzheimer's disease. Neurobiol Aging. 2018;65:98–108. https://doi.org/10.1016/j.neurobiolaging.2018.01.009. Epub 2018 Jan 31. PMID: 29455029.
21. Young AL, Marinescu RV, Oxtoby NP, Bocchetta M, Yong K, Firth NC, Cash DM, Thomas DL, Dick KM, Cardoso J, Van Swieten J, Borroni B, Galimberti D, Masellis M, Tartaglia MC, Rowe JB, Graff C, Tagliavini F, Frisoni GB, et al. Uncovering the heterogeneity and temporal complexity of neurodegenerative diseases with subtype and stage inference. Nat Commun. 2018;9(1):1–16. https://doi.org/10.1038/s41467-018-05892-0.
22. Loreto F, Verdi S, Kia SM, Duvnjak A, Hakeem H, Fitzgerald A, Patel N, Lilja J, Win Z, Perry R, Marquand AF, Cole JH, Malhotra P. Alzheimer's disease heterogeneity revealed by neuroanatomical normative modeling. Alzheimer's Dementia: Diagn Assessment Dis Monit. 2024;16(1). https://doi.org/10.1002/dad2.12559.
23. Varol E, Sotiras A, Davatzikos C. Alzheimer's disease neuroimaging initiative. HYDRA: revealing heterogeneity of imaging and genetic patterns through a multiple max-margin discriminative analysis framework. Neuroimage. 2017;145:346–364. https://doi.org/10.1016/j.neuroimage.2016.02.041.
24. Koenig LN, Day GS, Salter A, Keefe S, Marple LM, Long J, LaMontagne P, Massoumzadeh P, Snider BJ, Kanthamneni M, Raji CA, Ghoshal N, Gordon BA, Miller-Thomas M, Morris JC, Shimony JS, Benzinger TLS. Select atrophied regions in alzheimer disease (SARA): an improved volumetric model for identifying alzheimer disease dementia. NeuroImage: Clinical. 2020;26: 102248. https://doi.org/10.1016/j.nicl.2020.102248.
25. Kaczkurkin AN, Sotiras A, Baller EB, Barzilay R, Calkins ME, Chand GB, Cui Z, Erus G, Fan Y, Gur RE, Gur RC, Moore TM, Roalf DR, Rosen AFG, Ruparel K, Shinohara RT, Varol E, Wolf DH, Davatzikos C, Satterthwaite TD. Neurostructural heterogeneity in youths with internalizing symptoms. Biol Psychiatry. 2020;87(5):473–82. https://doi.org/10.1016/j.biopsych.2019.09.005. Epub 2019 Sep 18. PMID: 31690494; PMCID: PMC7007843.
26. Wen J, Fu CHY, Tosun D, Veturi Y, Yang Z, Abdulkadir A, Mamourian E, Srinivasan D, Skampardoni I, Singh A, Nawani H, Bao J, Erus G, Shou H, Habes M, Doshi J, Varol E, Mackin RS, Sotiras A, Fan Y, Saykin AJ, Sheline YI, Shen L, Ritchie MD, Wolk DA, Albert M, Resnick SM, Davatzikos C, iSTAGING consortium, ADNI, BIOCARD, BLSA. Characterizing heterogeneity in neuroimaging, cognition, clinical symptoms, and genetics among patients with late-life depression. JAMA Psychiatr. 2022;79(5):464–74. https://doi.org/10.1001/jamapsychiatry.2022.0020. PMID: 35262657; PMCID: PMC8908227.
27. Fu CHY, Antoniades M, Erus G, Garcia JA, Fan Y, Arnone D, Arnott SR, Chen T, Choi KS, Fatt CC, Frey BN, Frokjaer VG, Ganz M, Godlewska BR, Hassel S, Ho K, McIntosh AM, Qin K, Rotzinger S, Sacchet MD, Savitz J, Shou H, Singh A, Stolicyn A, Strigo I, Strother SC, Tosun D, Victor TA, Wei D, Wise T, Zahn R, Anderson IM, Craighead WE, Deakin JFW, Dunlop BW, Elliott R, Gong Q, Gotlib IH, Harmer CJ, Kennedy SH, Knudsen GM, Mayberg HS, Paulus MP, Qiu J, Trivedi MH, Whalley HC, Yan CG, Young AH, Davatzikos C. Neuroanatomical dimensions in medication-free individuals with major depressive disorder and treatment response to SSRI antidepressant medications or placebo. Nat Ment Health. 2024;2(2):164–76. https://doi.org/10.1038/s44220-023-00187-w. Epub 2024 Jan 12. PMID: 38948238; PMCID: PMC11211072.
28. Ge R, Sassi R, Yatham LN, Frangou S. Neuroimaging profiling identifies distinct brain maturational subtypes of youth with mood and anxiety disorders. Mol Psychiatry. 2023;28(3):1072–8. https://doi.org/10.1038/s41380-022-01925-9. Epub 2022 Dec 28. PMID: 36577839; PMCID: PMC10005933.
29. Chand GB, Dwyer DB, Erus G, Sotiras A, Varol E, Srinivasan D, Doshi J, Pomponio R, Pigoni A, Dazzan P, Kahn RS, Schnack HG, Zanetti MV, Meisenzahl E, Busatto GF, Crespo-Facorro B, Pantelis C, Wood SJ, Zhuo C, Shinohara RT, Shou H, Fan Y, Gur RC, Gur RE, Satterthwaite TD, Koutsouleris N, Wolf DH, Davatzikos C. Two distinct neuroanatomical subtypes of schizophrenia revealed using machine learning. Brain. 2020;143(3):1027–38. https://doi.org/10.1093/brain/awaa025. PMID: 32103250; PMCID: PMC7089665.
30. Fischl B. FreeSurfer. Neuroimage. 2012;62(2):774–81. https://doi.org/10.1016/j.neuroimage.2012.01.021.
31. PJ LaMontagne, TLS Benzinger, JC Morris, S Keefe, R Hornbeck, C Xiong, E Grant, J Hassenstab, K Moulder, AG Vlassenko, ME Raichle, C Cruchaga, D Marcus. OASIS-3: longitudinal neuroimaging, clinical, and cognitive dataset for normal aging and alzheimer disease. medRxiv. 2019;12:13.19014902. https://doi.org/10.1101/2019.12.13.19014902.
32. Berg L. Clinical Dementia Rating (CDR). Psychopharmacol Bull. 1989;24(3–4):637.
33. Su Y, Flores S, Hornbeck RC, Speidel B, Vlassenko AG, Gordon BA, Koeppe RA, Klunk WE, Xiong C, Morris JC, Benzinger TLS. Utilizing the centiloid scale in cross-sectional and longitudinal PiB PET studies. NeuroImage: Clinical. 2018;19:406–16. https://doi.org/10.1016/j.nicl.2018.04.022.
34. Su Y, D'Angelo GM, Vlassenko AG, Zhou G, Snyder AZ, Marcus DS, Blazey TM, Christensen JJ, Vora S, Morris JC, Mintun MA, Benzinger TL. Quantitative analysis of PiB-PET with FreeSurfer ROIs. PLoS ONE. 2013;8(11):e73377. https://doi.org/10.1371/journal.pone.0073377. PMID: 24223109; PMCID: PMC3819320.
35. Su Y, Blazey TM, Snyder AZ, Raichle ME, Marcus DS, Ances BM, Bateman RJ, Cairns NJ, Aldea P, Cash L, Christensen JJ, Friedrichsen K, Hornbeck RC, Farrar AM, Owen CJ, Mayeux R, Brickman AM, Klunk W, Price JC, Thompson PM, Ghetti B, Saykin AJ, Sperling RA, Johnson KA, Schofield PR, Buckles V, Morris JC, Benzinger TLS, Dominantly Inherited Alzheimer Network. Partial volume correction in quantitative amyloid imaging. Neuroimage. 2015;107:55–64. https://doi.org/10.1016/j.neuroimage.2014.11.058. Epub 2014 Dec 5. PMID: 25485714; PMCID: PMC4300252.
36. Klunk WE, Koeppe RA, Price JC, Benzinger TL, Devous MD Sr, Jagust WJ, Johnson KA, Mathis CA, Minhas D, Pontecorvo MJ, Rowe CC, Skovronsky DM, Mintun MA. The centiloid project: standardizing quantitative amyloid

plaque estimation by PET. Alzheimers Dement. 2015;11(1):1–15. https://doi.org/10.1016/j.jalz.2014.07.003. Epub 2014 Oct 28. PMID: 25443857; PMCID: PMC4300247.
37. Hansson O, Seibyl J, Stomrud E, Zetterberg H, Trojanowski JQ, Bittner T, Lifke V, Corradini V, Eichenlaub U, Batrla R, Buck K, Zink K, Rabe C, Blennow K, Shaw LM, Swedish BioFINDER study group, Alzheimer's Disease Neuroimaging Initiative. CSF biomarkers of Alzheimer's disease concord with amyloid-β PET and predict clinical progression: a study of fully automated immunoassays in BioFINDER and ADNI cohorts. Alzheimers Dement. 2018;14(11):1470–81. https://doi.org/10.1016/j.jalz.2018.01.010. Epub 2018 Mar 1. PMID: 29499171; PMCID: PMC6119541.
38. Folstein MF, Folstein SE, McHugh PR. Mini-mental state. J Psychiatr Res. 1975;12(3):189–98. https://doi.org/10.1016/0022-3956(75)90026-6.
39. Desikan RS, Ségonne F, Fischl B, Quinn BT, Dickerson BC, Blacker D, Buckner RL, Dale AM, Maguire RP, Hyman BT, Albert MS, Killiany RJ. An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. Neuroimage. 2006;31:968–80. https://doi.org/10.1016/j.neuroimage.2006.01.021.
40. Buckner RL, Head D, Parker J, Fotenos AF, Marcus DS, Morris JC, Snyder AZ. A unified approach for morphometric and functional data analysis in young, old, and demented adults using automated atlas-based head size normalization: reliability and validation against manual measurement of total intracranial volume. Neuroimage. 2004;23:724–38. https://doi.org/10.1016/j.neuroimage.2004.06.018.
41. Q Gu, J Han. Clustered support vector machines. Proceedings of the sixteenth international conference on artificial intelligence and statistics. 2013;307–315.
42. Hubert L, Arabie P. Comparing partitions. J Classif. 1985;2(1):193–218.
43. Iñesta C, Oltra-Cucarella J, Sitges-Maciá E. Regression-based normative data for independent and cognitively active Spanish older adults: verbal fluency tests and Boston naming test. Int J Environ Res Public Health. 2022;19(18):11445. https://doi.org/10.3390/ijerph191811445. PMID: 36141718; PMCID: PMC9517509.
44. Alverson WA, O'Rourke JJF, Soble JR. The word memory test genuine memory impairment profile discriminates genuine memory impairment from invalid performance in a mixed clinical sample with cognitive impairment. Clin Neuropsychol. 2019;33(8):1420–35. https://doi.org/10.1080/13854046.2019.1599071.
45. Carpenter CR, Bassett ER, Fischer GM, Shirshekan J, Galvin JE, Morris JC. Four sensitive screening tools to detect cognitive dysfunction in geriatric emergency department patients: brief Alzheimer's screen, short blessed test, Ottawa 3DY, and the caregiver-completed AD8. Acad Emerg Med. 2011;18(4):374. https://doi.org/10.1111/j.1553-2712.2011.01040.x.
46. Plackett RL. Karl Pearson and the chi-squared test. Int Stat Rev. 1983;51(1):59–72. https://doi.org/10.2307/1402731.
47. McKnight PE, Najab J. Mann-Whitney U test. 1. https://doi.org/10.1002/9780470479216.corpsy0524.
48. Lakens D. Calculating and reporting effect sizes to facilitate cumulative science: a practical primer for t-tests and ANOVAs. Front Psychol. 2013. https://doi.org/10.3389/fpsyg.2013.00863.
49. Mowinckel A, Vidal-Piñeiro D. Visualization of brain statistics with R packages ggseg and ggseg3d. Adv Methods Pract Psychol Sci. 2020;3:466–83. https://doi.org/10.1177/2515245920928009.
50. Wei K, Tran T, Chu K, Borzage MT, Braskie MN, Harrington MG, King KS. White matter hypointensities and hyperintensities have equivalent correlations with age and CSF β-amyloid in the nondemented elderly. Brain Behav. 2019;9(12): e01457. https://doi.org/10.1002/brb3.1457. Epub 2019 Nov 6. PMID: 31692294; PMCID: PMC6908861.
51. Kamal F, Moqadam R, Morrison C, Dadar M. Racial and ethnic differences in white matter hypointensities: the role of vascular risk factors. Alzheimers Dement. 2025;21(3): e70105. https://doi.org/10.1002/alz.70105. PMID: 40145319; PMCID: PMC11947760.
52. Bernal-Rusiel JL, Greve DN, Reuter M, Fischl B, Sabuncu MR. Statistical analysis of longitudinal neuroimage data with linear mixed effects models. Neuroimage. 2013;66:249–60. https://doi.org/10.1016/j.neuroimage.2012.10.065.
53. Akaike H. A new look at the statistical model identification. IEEE Trans Automat Control. 1974;19(6):716–23. https://doi.org/10.1109/TAC.1974.1100705.
54. Ross A, Willson VL. One-way anova. In: basic and advanced statistical tests. SensePublishers, Rotterdam. 2017. https://doi.org/10.1007/978-94-6351-086-8_5.
55. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc. Ser B (Methodological). 1995;57(1):289–300. https://doi.org/10.2307/2346101.
56. Abd ElHafeez S, D'Arrigo G, Leonardis D, Fusaro M, Tripepi G, Roumeliotis S. Methods to analyze time-to-event data: the Cox regression analysis. Oxid Med Cell Longev. 2021;2021:1302811. https://doi.org/10.1155/2021/1302811.
57. Foudil A, Macaluso E. The influence of the precuneus on the medial temporal cortex determines the subjective quality of memory during the retrieval of naturalistic episodes. Sci Rep. 2024. https://doi.org/10.1038/s41598-024-58298-y.
58. Liu Y, Li Z, Bai Y. Frontal and parietal lobes play crucial roles in understanding the disorder of consciousness: a perspective from electroencephalogram studies. Front Neurosci. 2022;16. https://doi.org/10.3389/fnins.2022.1024278.
59. Ferreira D, Verhagen C, Hernández-Cabrera JA, Cavallin L, Guo CJ, Ekman U, Muehlboeck JS, Simmons A, Barroso J, Wahlund LO, Westman E. Distinct subtypes of Alzheimer's disease based on patterns of brain atrophy: longitudinal trajectories and clinical applications. Sci Rep. 2017;7(1):46263. https://doi.org/10.1038/srep46263. PMID: 28417965; PMCID: PMC5394684.
60. Vermeiren MR, Calandri IL, van der Flier WM, van de Giessen E, Ossenkoppele R. Survey among experts on the future role of tau-PET in clinical practice and trials. Alzheimer's Dementia: Diagn Assessment Dis Monit. 2024;16(4):e70033. https://doi.org/10.1002/dad2.70033. PMID: 39583643; PMCID: PMC11582687.

Publisher's Note