Analysing heterogeneity in Alzheimer's Disease using multimodal normative modelling on ATN biomarkers

Sayantan Kumar[a,b*], Thomas Earnest[c], Braden Yang[c], Deydeep Kothapalli[c], Tammie L. S. Benzinger[c], Brian A. Gordon[c], Philip Payne[a,b], Aristeidis Sotiras[b,c], for the Alzheimer's Disease Neuroimaging Initiative

[a] Department of Computer Science and Engineering, Washington University in St Louis; 1 Brookings Drive, Saint Louis, MO 63130

[b] Institute for Informatics, Data Science & Biostatistics, Washington University School of Medicine in St Louis; 660 S. Euclid Ave, Campus Box 8132, Saint Louis, MO 63110

[c] Mallinckrodt Institute of Radiology, Washington University School of Medicine in St Louis; 4525 Scott Ave, Saint Louis, MO 63110

*Corresponding author: sayantan.kumar@wustl.edu, 660 S. Euclid Ave, Campus Box 8132, Saint Louis, MO 63110

Keywords: Normative modelling; ADNI; Alzheimer Disease; heterogeneity; ATN; multimodal variational autoencoders

Structured Abstract

Background and Objectives

Previous approaches pursuing normative modelling for analyzing heterogeneity in Alzheimer's Disease (AD) have relied on a single neuroimaging modality. However, AD is a multi-faceted disorder, with each modality providing unique and complementary info about AD. In this study, we used a deep-learning based multimodal normative model to assess the heterogeneity in regional brain patterns for ATN (amyloid-tau-neurodegeneration) biomarkers.

Methods

We selected discovery (n = 665) and replication (n = 430) cohorts with simultaneous availability of ATN biomarkers: Florbetapir amyloid, Flortaucipir tau and T1-weighted MRI (magnetic resonance imaging) imaging. A multimodal variational autoencoder (conditioned on age and sex) was used as a normative model to learn the multimodal regional brain patterns of a cognitively unimpaired (CU) control group. The trained model was applied on individuals on the ADS (AD Spectrum) to estimate their deviations (Z-scores) from the normative distribution, resulting in a Z-score regional deviation map per ADS individual per modality. Regions with Z-scores < -1.96 for MRI and Z-scores > 1.96 for amyloid and tau were labelled as outliers. Hamming distance was used to quantify the dissimilarity between individual based on their outlier deviations across each modality. We also calculated a disease severity index (DSI) for each ADS individual which was estimated by averaging the deviations across all outlier regions corresponding to each modality.

Results

ADS individuals with moderate or severe dementia showed higher proportion of regional outliers for each modality as well as more dissimilarity in modality-specific regional outlier patterns compared to ADS individuals with early or mild dementia. DSI was associated with the progressive stages of dementia, (ii) showed significant associations with neuropsychological composite scores and (iii) related to the longitudinal risk of CDR progression. Findings were reproducible in both discovery and replication cohorts.

Discussion

Our is the first study to examine the heterogeneity in AD through the lens of multiple neuroimaging modalities (ATN), based on distinct or overlapping patterns of regional outlier deviations. Regional MRI and tau outliers were more heterogenous than regional amyloid outliers. DSI has the potential to be an individual patient metric of neurodegeneration that can help in clinical decision making and monitoring patient response for anti-amyloid treatments.

1. Introduction

Alzheimer Disease (AD) is the most common cause of dementia, which is a syndrome characterized by impairment of memory and/or thinking severe enough to interfere with activities of daily life.[1,2] AD-related brain pathology, which includes the accumulation and deposition of amyloid-β peptide and tau protein, begins almost 10-20 years before the onset of dementia symptoms.[3] Therefore, many individuals with early AD brain pathology are cognitively normal but at higher risk for developing dementia in the future. AD is highly heterogeneous, with significant variations in clinical presentations, cognitive impairments, disease progression rates, and underlying neuropathological changes, even within a specific diagnostic group.[4] However, the most common statistical designs in AD research overlook this heterogeneity and instead rely on comparisons of group averages (e.g. patient versus control), implicitly assuming within-group homogeneity.[5] To advance towards the goal of precision medicine in AD, it is imperative to move beyond the average or "one-size-fits-all" approach and focus on analysing heterogeneity at the individual subject level.[6,7]

Recent studies have used data-driven clustering approaches to parse heterogeneity and group subjects based on distinctive neurobiological and cognitive characteristics.[8–12] However, clustering assumes homogeneity of subjects within each cluster, limiting the extent of analysing individual-level variability in the sample.[13] Normative modelling is an emerging statistical technique, alternative and complementary to clustering, which shifts the focus from group averages to subject-level variation. The procedure of normative modelling in the context of AD is a two-fold approach: (i) machine learning models are estimated using data from a reference cohort of cognitively unimpaired (CU) participants and (ii) the estimated model is applied to a target clinical cohort of subjects diagnosed with AD.[14–20] A diseased

individual can then be located within the normative distribution to assess what extent they deviate from the norm at any given brain region, providing a map of individual variability.

Recent studies using normative modelling on subjects with AD have used only single modality neuroimaging data.[20–22] However, AD is a multi-factorial disease, showing deviations from the norm in features across multiple imaging modalities. Modalities like magnetic resonance imaging (MRI) and positron emission tomography (PET) are generally used to measure the neuroanatomical (atrophy) and pathophysiological changes (amyloid and tau deposition) in the brain respectively.[23–25] Each modality provides unique and complementary information about the disease, allowing for a more comprehensive understanding of AD pathology, progression, and individual differences. This multimodal approach has the potential to improve diagnostic accuracy, enhance predictive models, identify subtypes or clusters of subjects with distinct disease profiles, and guide personalized treatment strategies.[26–29] However, no study to date has investigated normative modelling across multiple modalities.

Ours is the first study to examine individual patterns of variation in brain structure in individuals with AD using normative modelling across multiple neuroimaging modalities. We applied our previous implementation[30,31] of normative modelling framework based on multimodal variational autoencoders (mmVAE), which used the ATN (amyloid-tau-neurodegeneration)[32] biomarker data as multimodal input. First, we trained mmVAE on a reference control group (i.e., cognitively unimpaired (CU) subjects) from the Alzheimer's Disease Neuroimaging Initiative (ADNI) dataset to learn the healthy (normative) brain patterns across all modalities. Next, we evaluated model performance on ADNI individuals spanning the AD spectrum (ADS) and estimated how their modality-specific regional brain patterns deviate from the normative healthy distribution. Our main objective was to assess the

regional patterns of neuroanatomical (MRI atrophy) and neuropathological (amyloid and tau deposition) variability in ADS individuals based on overlapping or distinct pattern of outlier deviations. Next, we used Hamming distance to quantify the dissimilarity between ADS individuals based on their outlier deviations across each modality. We also calculated a disease severity index (DSI) for each ADS individual which was estimated by averaging the deviations across all outlier regions corresponding to each modality. Lastly, we estimated the correlation between DSI and cognitive performance and examined if DSI related to disease progression. The results were validated in an independent dataset, part of the Charles F. and Joanne Knight Alzheimer's Disease Research Center (ADRC) dataset at Washington University in St. Louis.

2. Materials and Methods

2.1 Overview

Our analysis included a discovery dataset consisting of individuals from ADNI[33] (Section 2.2) and a replication dataset consisting of individuals from Charles F. and Joanne Knight Alzheimer's Disease Research Center (ADRC) dataset at Washington University in St. Louis (section 2.3). From both datasets, we selected a reference control cohort (i.e. cognitively unimpaired (CU) with Clinical Dementia Rating (CDR®) = 0 and amyloid negative) and a disease cohort with individuals on the ADS. Note that we included individuals in the disease cohort only based on amyloid positivity and not clinical or cognitive status. Our analysis consisted of four major steps. (1) First, we trained a multimodal variational autoencoder (mmVAE) on regional gray matter volumes, amyloid and tau signals of the control cohort to learn the healthy brain patterns (normative distribution) for each modality (Section 2.4.1). (2) We applied the trained mmVAE on the ADS group to estimate Z-scores, indicating how their modality-specific regional brain deviate from the normative healthy distribution (Sections 2.4.2, 2.4.3). (3) We identified regional deviations for each modality which were statistical outliers (Section 2.5.1) and analyzed the spatial patterns of modality-specific outliers per CDR-stratified diagnostic group within the ADS cohort (Sections 2.5.2, 2.5.3). Next, we used Hamming distance to quantify the dissimilarity in regional outlier patterns between individual within the same diagnostic group (Section 2.5.4). (4) Finally, we calculated a disease severity index (DSI) for everyone on the ADS cohort which was estimated by averaging the deviations across all outlier regions corresponding to each modality (Section 2.5.5). We estimated the correlation between DSI and cognitive performance (Section 2.5.6) and examined if DSI related to CDR progression (Section 2.5.7). Note that steps (1-4) were first performed on the discovery dataset. mmVAE, pretrained on the discovery dataset was

subsequently fine-tuned on the reference control cohort in the replication dataset by transfer learning. Finally, we used the replication dataset to repeat steps (1) - (4) above and validate the results observed in the discovery dataset. Additional methodological details can be found in the Supplementary Methods (SM).

2.2 Discovery Dataset - ADNI

2.2.1 Participants

Data used in the preparation of this article were obtained from the ADNI database (adni.loni.usc.edu). ADNI started in 2003 as a public-private partnership, led by Principal Investigator Michael W. Weiner, MD. Its goal was to verify whether different neuroimaging biomarkers and neuropsychological assessments can be combined to measure the progression of Mild Cognitive Impairment (MCI) and early AD. The criteria for inclusion from ADNI was the simultaneous availability of baseline scans of T1-weighted MRI, amyloid PET and tau PET respectively, all of which should be dated within 1 year of one another. We selected two cohorts based on amyloid status (Section 2.2.3): a reference control dataset (ADNI-CU) for model training and a target disease cohort (ADNI-ADS) for model evaluation (eFigure 2). 434 CU subjects with CDR® = 0 and without amyloid pathology were included in the reference control cohort. The target disease cohort consisted of 231 amyloid positive individuals on the ADS. ADNI-ADS individuals were assigned into 3 diagnostic groups based on CDR: CDR = 0 (n = 121), CDR = 0.5 (n = 80) and CDR >= 1 (n = 30) (eFigure 2). The ADNI-CU cohort was further divided into training (n = 326), holdout validation (n = 65) and test sets (n = 43) respectively in a ratio of 75:15:10. We used the training set for training mmVAE, and the holdout validation set for normalizing the deviations of ADNI-ADS with reference to ADNI-CU (see Section 2.4.3). The test set was used to compare deviations

across CU and ADNI-ADS groups. We refer to these three sets as ADNI-CU-train, ADNI-CU-holdout and ADNI-CU-test respectively (eFigure 2).

All ADNI participants provided written informed consent, and study protocols were approved by each local site's institutional review board. ADNI data used in this study are publicly available and can be requested following ADNI Data Sharing and Publications Committee guidelines: https://adni.loni.usc.edu/data-samples/access-data/. Data included in this manuscript were downloaded on October, 2023.

2.2.2 MRI Imaging

All participants included in our analysis underwent T1-weighted MRI imaging using 3T MRI scanners. The T1-weighted sequences were processed into Region-of-Interest (ROI) segmentation with FreeSurfer version 6. The cortical surface of each hemisphere was parcellated according to the Desikan–Killiany atlas[34] and anatomical ROI measures were obtained via a whole-brain segmentation procedure (Aseg atlas).[35] The final data included regional grey matter volumes of 66 cortical ROIs (33 per hemisphere) and 24 subcortical ROIs for each participant. All ROI volumes were normalized by the intracranial volume (ICV). More information about T1-weighted MRI acquisition and preprocessing in ADNI are available online (https://adni.loni.usc.edu/methods/mri-tool/mri-analysis/).

2.2.3 Amyloid and tau PET Imaging

All participants included in our analysis underwent amyloid-PET imaging with either Florbetapir (AV45) or Florbetaben (FBB) as tracers and tau-PET imaging with Flortaucipir (AV1451) as the tracer. In our study, we only considered AV45 for amyloid-PET. PET images (AV45 and AV1451) were registered to the nearest T1-weighted image and processed with FreeSurfer (version 6) to derive average standardized uptake value ratios (SUVRs) in ROIs. Similar to MRI ROI volumes, we extracted both amyloid and tau SUVR for 66 cortical

ROIs (33 per hemisphere) and 24 subcortical ROIs. More information about PET acquisition and preprocessing in ADNI can be found available online (https://adni.loni.usc.edu/methods/pet-analysis-method/pet-analysis/).

Amyloid positivity for AV45-PET was determined using established cut-off procedures recommended within ADNI documentation.[36–39] Specifically, the average AV45 SUVR in a cortical meta-ROI (spanning frontal, anterior/posterior cingulate, lateral parietal, and lateral temporal regions) was calculated and normalized with a whole cerebellum reference region. In our study, individuals with meta-ROI SUVR uptake greater than 1.11 cutoff were considered amyloid positive.

2.3 Replication Cohort - Alzheimer's Disease Research Center (ADRC)

2.3.1 Participants

We included participants from the Knight ADRC dataset who underwent T1-weigted MRI imaging, amyloid-PET and tau-PET scans, all of which should be dated within 1 year of each other. Like the discovery dataset, the replication dataset was also divided into two cohorts based on amyloid status (Section 2.3.2): reference control cohort for transfer learning and a disease cohort for external validation (eFigure 3). The reference cohort (ADRC-CU) consisted of 301 CU individuals (CDR = 0, amyloid negative) and the disease cohort (ADRC-ADS) consisted of 129 amyloid positive individuals on the ADS (eFigure 3). ADRC-ADS individuals were assigned into 3 diagnostic groups based on CDR: CDR = 0 (n = 98), CDR = 0.5 (n = 24) and CDR >= 1 (n = 7). Like the discovery dataset, the ADRC-CU cohort was divided into training (n = 225), holdout validation (n = 45) and test sets (n = 31). We refer to these three sets as ADRC-CU-train, ADRC-CU-holdout and ADRC-CU-test respectively (eFigure 3).

All protocols were approved by the Institutional Review Board at Washington University in St. Louis, and all participants provided informed consent before all procedures. Knight ADRC data can be obtained by submitting a data request through https://knightadrc.wustl.edu/data-request-form/.

2.3.2 MRI Imaging

Participants from Knight ADRC underwent T1-weighted MRI imaging using the Siemens Biograph mMR 3T scanner. Freesurfer (version 5.3) was used for volumetric segmentation of the T1-weighted scans into Region-of-Interest (ROI). Similar to the discovery dataset, cortical surface of each hemisphere was parcellated according to the Desikan–Killiany atlas[34] and anatomical ROI measures were obtained via a whole-brain segmentation procedure (Aseg atlas).[35] For our analysis, we extracted regional grey matter volumes of 66 cortical ROIs (33 per hemisphere) and 24 subcortical ROIs for each participant (same regions as in ADNI). The ROI volumes were normalized by the Intracranial volume. More detailed information about MRI image acquisition and processing for the ADRC dataset can be found in the Supplementary Methods (SM1).

2.3.3 Amyloid and tau PET Imaging

ADRC participants in our study underwent tau-PET imaging with Flortaucipir (AV1451) and amyloid-PET imaging with either AV45 or Pittsburgh Compound B (PIB) as the tracer respectively. In our study, we only considered AV45 for amyloid-PET. All PET images from ADRC (AV45 and AV1451) were processed using the PET Unified Pipeline (https://github.com/ysu001/PUP)[40,41]. AV45 and AV1451 PET images were registered to the nearest T1-weighted image and processed with FreeSurfer (version 5.3) to derive average standardized uptake value ratios (SUVRs) in ROIs. Similar to MRI ROI volumes, we extracted both AV45 and AV1451 SUVR for 66 cortical ROIs (33 per hemisphere) and 24

subcortical ROIs (same regions as in ADNI). More detailed information about PET image acquisition and processing for ADRC can be found in the Supplementary Methods (SM2).

Amyloid positivity for AV45-PET was determined using well-established cut-off procedures as recommended in previous works.[42,43] Specifically, a mean cortical SUVR was calculated in a cortical meta-ROI consisting of precuneus, prefrontal, gyrus rectus, and lateral temporal cortices, using the cerebellum cortex as the reference region. We converted the cortical meta-ROI SUVR to Centiloid values using equations developed by Su et al.[42] In our study, ADRC individuals with Centiloid greater than the 20.6 cut-off were considered amyloid positive.

2.4 Multimodal Normative Modelling

Normative modelling is an emerging statistical technique, alternative and complementary to clustering which have been used for analyzing heterogeneity in neurodegenerative diseases like AD. Briefly, normative modelling learns healthy brain patterns and quantifies how each individual diagnosed with AD deviate from the normative distribution, providing a map of individual level variability.[17,20,21] In our work, we adopted a multimodal variational autoencoder (mmVAE) as the normative model, validated in previous works.[21,30,31]

2.4.1 Model training

The mmVAE framework used regional gray matter volumes, amyloid AV45 SUVR and tau AV1451 SUVR as multimodal input data. The model had separate encoders to learn brain patterns for each modality which were aggregated in a joint latent distribution. The shared information in the aggregated latent space were fed through modality-specific decoders to reconstruct the input for each modality. A detailed description of the mmVAE architecture is provided in the Supplementary methods (SM3.1, SM3.2).

In our work, mmVAE was first applied on the ADNI-CU-train cohort. The input matrix for ADNI-CU-train had a shape of 326 x 270 (subjects x ROIs, 90 ROIs for each modality). During training, mmVAE learns to reconstruct the multimodal input data as closely as possible to the original. The joint latent distribution allows the model to learn the healthy brain patterns across all modalities. mmVAE was conditioned on the age and sex of participants to remove the effect of covariates (see Supplementary Methods SM3.3). For replication in the ADRC dataset, mmVAE, pretrained on ADNI-CU-train was fine-tuned on ADRC-CU-train by transfer learning (Section 2.3.1). Further information about training and hyperparameter details can be found in the Supplementary Methods (SM3.3).

2.4.2 Calculating regional deviations for each modality.

The main idea of the normative approach is that since mmVAE learns only to reconstruct the data of CU individuals, it will be less precise in reconstructing the data of individuals on the ADS. As a result, the difference between the reconstructed and input data will be larger in the ADS cohort compared to CU individuals. For each participant in ADNI-ADS and ADRC-ADS (including ADNI-CU-test and ADRC-CU-test respectively), deviations for each region and for each modality can be calculated by the squared error between input and reconstructed data.

2.4.3 Normalizing deviations into Z-scores

Considering that the brain contains highly variable complex structure such as the cortex, we expect that mmVAE will not be able to fully capture normative variation within healthy subjects, resulting in a non-trivial reconstruction error. Therefore, it is important to standardize the deviations of ADNI-ADS with reference to the deviations of the CU population. We used ADNI-CU-holdout (Section 2.2.1) as a representative of CU reference population and used the mean and variance (calculated independently for each region) to

normalize the regional deviations of the ADNI-ADS cohort for each modality. After fine-tuning mmVAE on the ADRC-CU-train cohort, ADRC-CU-holdout (Section 2.3.1) was used similarly to normalize the deviations on ADRC-ADS. This process generated regional Z-score deviations for each individual and for each modality in the ADS cohort relative to the normative range of the reference control cohort.

2.5 Statistical Analysis

2.5.1 Outlier definition

In our study, we are only interested in loss in gray matter volumes (MRI atrophy) and increase in amyloid/tau SUVR uptake associated with neurodegeneration. MRI outliers in terms of low gray matter volumes (atrophy) were identified for each region with a lower bound for z-scores ($Z < -1.96$), corresponding to the bottom 2.5% of the normative distribution of ROI volumes. Likewise, amyloid and tau outliers in terms of high amyloid and tau loading were identified for each region with an upper bound for z-scores ($Z > 1.96$), corresponding to the top 2.5% of the normative distribution of amyloid/tau SUVR uptake. All outlier deviations were corrected for multiple comparisons using the False Discovery Rate (FDR).[44]

2.5.2 Group differences of regional outlier deviations

We estimated the pairwise group differences in magnitude of outlier deviations at each region between CU individuals (ADNI-CU-test and ADRC-CU-test) and each of the CDR groups in ADS, namely (i) CDR = 0, (ii) CDR = 0.5 and (iii) CDR >= 1 (Section 2.2.1 and 2.3.1). The group differences at each region were quantified using Cohen's d-statistic effect size [45,46] (FDR corrected), adjusted for age and sex and calculated separately for each modality. Higher effect size for a MRI region indicated low gray matter volume (more MRI atrophy) for that region compared to CU individuals. Likewise, high effect size for amyloid or tau

region indicated high SUVR uptake (more amyloid/tau deposition) for that region related to the CU group. Effect size maps for both ADNI and ADRC were visualized using the Desikan-Killiany atlas[34] (cortical ROIs) and Aseg atlas[35] (subcortical ROIs) via the python package ggseg.[47]

2.5.3 Spatial patterns of outliers for each region

We also aimed to analyse the spatial patterns of outlier deviations for each region and for each modality. For each region, we calculated outlier frequency or the proportion of individuals on the ADS (ADNI-ADS and ADRC-ADS) and a CU group (ADNI-CU-test and ADRC-CU-test; see Section 2.2.1 and 2.3.1) whose deviations were outliers (FDR corrected). The outlier frequency for each region was calculated for separately across all modalities. This enabled visualization of the extent to which patterns of outlier regions overlap or are distinct across modalities. Outlier frequency maps for both ADNI and ADRC were visualized using the Desikan-Killiany atlas[34] (cortical ROIs) and Aseg atlas[35] (subcortical ROIs) via the python package ggseg.[47]

2.5.4 Dissimilarity in outlier patterns between individuals within same CDR group

Our goal was to assess the extent to which individuals on the ADS are different from each other with respect to outlier patterns even within the same CDR group. For each individual, we created a binary thresholded outlier vector with the regions having outlier deviations marked as 1 and 0 otherwise. Note that binary thresholded vectors were calculated both using regions (n = 90) corresponding to a single modality (mri, amyloid or tau) and also total number of regions across all modalities (n = 270). We used Hamming distance between the binary thresholded outlier vectors to measure the level of dissimilarity between individuals within each group. Hamming distances were estimated corresponding to all modalities and

also separately for individual modalities for both ADNI-ADS and ADRC-ADS. We refer to these distances as hamming_all, hamming_mri, hamming_amyloid and hamming_tau respectively.

2.5.5 Disease severity index (DSI) across CDR groups

We calculated a disease severity index (DSI) for everyone on the ADS (both ADNI-ADS and ADRC-ADS) and a group of the CU individuals (ADNI-CU-test and ADRC-CU-test; Section 2.2.1 and 2.3.1). DSI was designed to capture both the spatial extent of outliers and the magnitude of regional outlier deviations. Specifically, DSI for each individual was calculated by averaging the deviations across all outlier regions for a particular modality. Higher DSI values indicate more outliers which translates to more atrophy for MRI and higher burden for amyloid and tau uptake respectively. We compared DSI across each different CDR groups in and a fraction of the CU individuals included for evaluation (ADNI-CU-test and ADRC-CU-test; Section 2.2.1 and 2.3.1) to validate if increase in DSI is associated with high CDR values (progressive stages of dementia). FDR-corrected post hoc Tukey comparisons used to assess pairwise group differences. We calculated DSI across all modalities and separately for each of the three modalities. We refer to these values as DSI_all, DSI_mri, DSI_amyoid and DSI_tau respectively.

2.5.6 Relationship between DSI and neuropsychological composite scores

We assessed how DSI is related with cognitive performance for both ADNI-ADS and ADRC-ADS cohorts. Following previous work[48], we computed neuropsychological composites for memory, executive functioning and language. However, some modifications were added to the composites to account for availability of cognition data in ADNI and Knight ADRC. The specific tests from both datasets were calculated as follows: (i) **Memory**: The memory composite for ADNI consisted of the immediate and delayed recall items from

both the Logical Memory IIa test[49] and the Rey Auditory Verbal Learning Test.[50] For ADRC, the memory composite differed from ADNI and consisted of delayed item recall from the Logical Memory IIa test[49], the Free and Cued Selective Reminding Test[51], and Associate Memory[49] (ii) **Executive functioning**: the executive function composite consisted of Trails Making Test Parts A & B[52] for both ADNI and ADRC. (iii) **Language**: the language composite for both datasets included the category fluency for animals and vegetables[53], the Boston Naming Test [54], and the Multilingual Naming Test.[55] We converted all individual test scores into Z-scores relative to the respective control group (ADNI-CU and ADRC-CU with low scores indicating more impairment. Next, we averaged the rescaled scores to compute the neuropsychological composite scores.

For our analysis, we used linear regression, adjusted for age and sex to assess the relationship between DSI and the composites and also calculated the correlation between them using the Pearson Correlation coefficient.

2.5.7 Relationship between DSI and CDR progression

We analyzed associations between DSI_all and CDR progression in both ADNI-ADS and ADRC-ADS using survival analysis, adjusted for age and sex. Progression to CDR >= 1 was modelled as the event of interest. Subjects were only included in this analysis if they were CDR<1 at their baseline MRI scan and had available longitudinal CDR assessments. We used a Kaplan–Meier plot to illustrate how the 4 quantiles of DSI_all can contribute to the risk of disease progression. Log-ranked tests were performed to estimate pairwise differences between the progression risk for the DSI_all quantiles. Post-hoc comparisons were adjusted for multiple comparisons using FDR.[44]

3. Results

3.1 Dataset characteristics

Sample characteristics for the ADNI-ADS and ADRC-ADS datasets are shown in Table 1. The two cohorts had an unequal distribution of age and sex with relatively more females in ADRC-ADS ($p = 0.006$) and older individuals in ADNI-ADS ($p = 0.035$). The ADRC-ADS cohort had less AD pathology than the ADNI-ADS indicated by the generally high MMSE scores ($p < 0.001$) and higher proportion of individuals with CDR = 0 ($p < 0.001$). ADNI-ADS had higher amyloid burden (centiloid) than ADRC-ADS, but the difference between the two cohorts were not statistically significant ($p = 0.074$).

Sample characteristics for the ADNI-CU and ADRC-CU datasets are provided in eTable 1. The two cohorts differed only and sex distribution in terms of age with ADNI-CU individuals having higher age ($p < 0.001$) and ADRC-CU with more females ($p = 0.026$). ADNI-CU had a trend towards worse MMSE scores than ADRC-CU, but the difference was not statistically significant ($p = 0.067$). The difference in amyloid burden between ADNI-CU and ADRC-CU was also not statistically significant ($p = 0.145$). Comparing each CU cohort to their respective ADS counterpart (ADNI-CU vs ADNI-ADS; ADRC-CU vs ADRC-ADS), the CU cohorts exhibited lower age, higher MMSE scores and lower amyloid burden than the corresponding ADS cohorts, with all differences being statistically significant ($p < 0.001$).

	ADNI-ADS	**ADRC-ADS**	**p-value**
N	231	129	-
Sex, Male: Female	108:123	48:81	**p = 0.035***
Age (mean +/- SD)	73.6 +/- 6.9	71.5 +/- 8.3	**p = 0.006****
CDR (0/0.5/>=1)	121/80/30	98/24/7	**p < 0.001*****
MMSE (mean +/- SD)	24.5 +/- 3.2	26.5 +/ 3.7	**p < 0.001 *****
Amyloid centiloid (mean +/- SD)	68 +/- 28.2	62.5 +/- 25.5	p = 0.074

Table 1: Descriptive statistics for the ADNI-ADS and ADRC-ADS datasets. Statistical differences were assessed using two-sided ANOVA (continuous variables) and chi-squared tests (categorical. variables) respectively. Significant p-values are highlighted in bold with *: $0.01 < p < 0.05$, **: $0.005 < p < 0.01$, ***: $p < 0.001$. Abbreviations: SD = standard deviation, ANOVA = analysis of variance, CDR = Clinical Dementia Rating, MMSE = Mini-Mental State Examination.

3.2 ADS individuals with moderate or severe dementia showed more regional MRI atrophy, amyloid and tau loading compared ADS individuals with early or mild dementia.

We aimed to examine the pairwise group differences in magnitude of outlier deviations at each region between CU individuals (ADNI-CU-test and ADRC-CU-test) and each of the CDR groups in ADS (ADNI-ADS and ADRC-ADS), namely (i) CDR = 0, (ii) CDR = 0.5 and (iii) CDR >= 1 (Section 2.2.1 and 2.3.1). The pairwise group comparisons for each MRI region (total 90 regions) in ADNI-ADS showed significant MRI atrophy (outliers) in more regions for [CDR >= 1 vs CU; n = 56] compared to [CDR = 0.5 vs CU; n = 22] with no outlier regions for [CDR = 0 vs CU]. Region-level group differences in MRI atrophy were most evident within the temporal, parietal and hippocampal regions and to some extent within the frontal, occipital and amygdala regions (Figure 1A). Similar to MRI, regional pairwise group comparisons for tau (n = 90) in ADNI-ADS exhibited significant tau loading (outliers) in more regions for [CDR >= 1 vs CU; n = 80] compared to [CDR = 0.5 vs CU; n = 62] and no outlier regions for [CDR = 0 vs CU]. Differences in tau loading were most evident in the temporal, parietal and hippocampal regions with entorhinal, parahippocampal and temporalpole regions having the maximum effect size (Figure 1A). Significant differences in amyloid burden (ADNI-ADS) were observed in 84 (total 90 regions) regions for [CDR >= 1 vs CU], 75 regions for [CDR = 0.5 vs CU] and 85 regions for [CDR = 0 vs CU]. Region-level group differences in amyloid SUVR uptake were mostly observed in the frontal and temporal regions with regions like medialorbitofrontal, precuneus and frontalpole showing higher effect size (Figure 1A). In general, regions exhibited a trend of larger effect sizes

corresponding to amyloid and tau burden than gray matter volume atrophy which was observed for all the CDR groups in ADNI-ADS (Figure 1A).

We observed reasonable similarity between the regional group differences in ADRC-ADS and the effect size maps observed in ADRC-ADS (Figure 1B). Similar to ADNI-ADS, regional group differences in MRI atrophy were mostly observed in the temporal, parietal and hippocampal regions [CDR = 0 vs CU; n = 0, CDR = 0.5 vs CU; n = 25, CDR >= 1 vs CU; n = 50]. More regional group differences in outlier deviations were observed in the frontal and temporal regions for amyloid [CDR = 0 vs CU; n = 81, CDR = 0.5 vs CU; n = 77, CDR >= 1 vs CU; n = 80] and temporal and hippocampal regions for tau [CDR = 0 vs CU; n = 0, CDR = 0.5 vs CU; n = 52, CDR >= 1 vs CU; n = 74] respectively (Figure 1B). The trend of larger regional effect sizes for amyloid and tau loading compared to MRI atrophy was also replicated across the CDR groups in ADRC-ADS (Figure 1B).

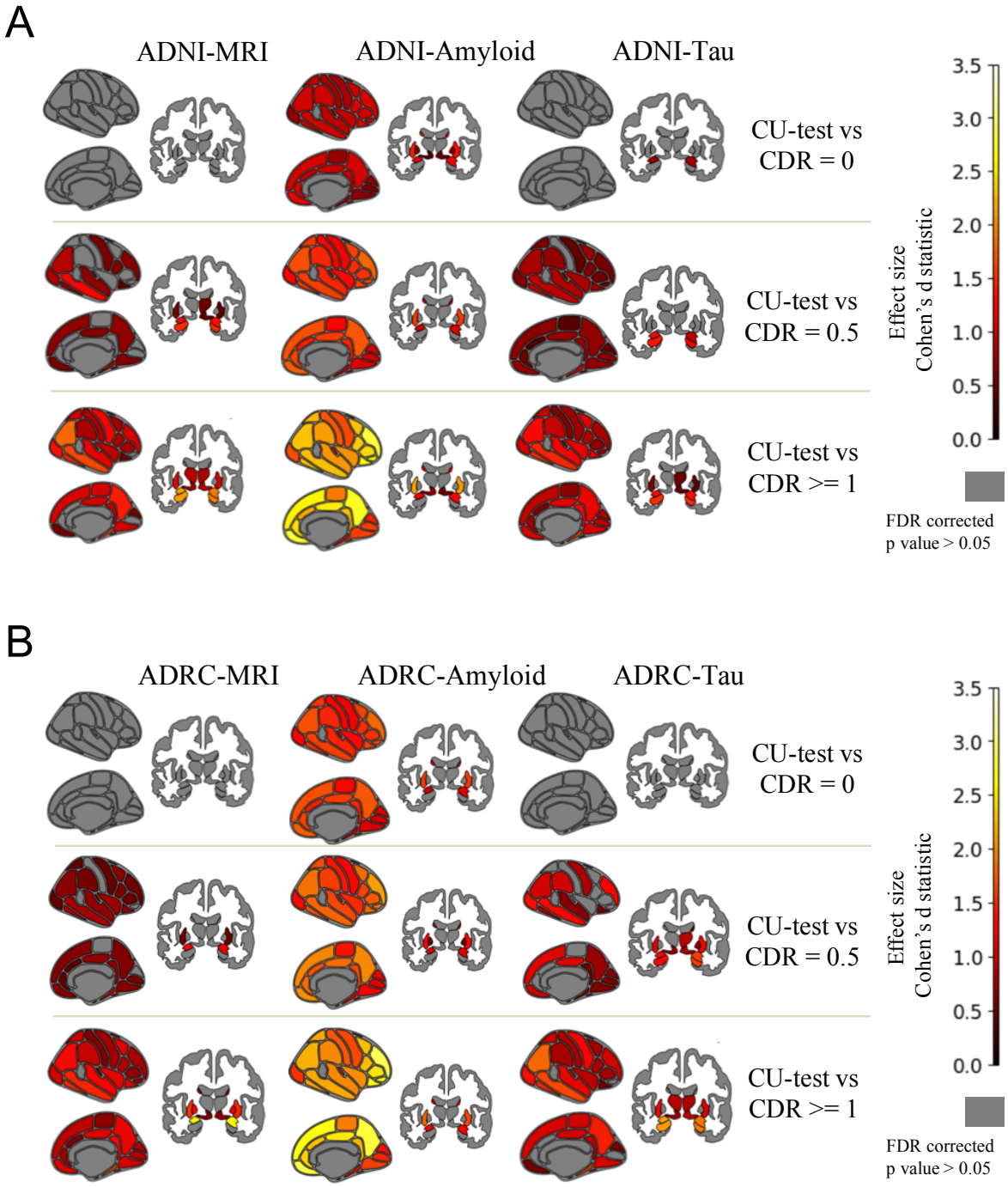

Figure 1: Brain atlas maps (Desikan-Killiany atlas for 66 cortical regions and Aseg atlas for 24 subcortical regions) showing the pairwise group differences in magnitude of deviations at each region between CU group and each of the CDR groups in ADNI-ADS (**1A**) and ADRC-ADS (**1B).** The figures from left to right indicate the brain maps corresponding to MRI, amyloid and tau respectively. The color bar represents the effect size (Cohen's d statistic). A commonly used interpretation is to refer to effect sizes as small (d = 0.2), medium (d = 0.5), and large (d = 0.8) based on benchmarks. Gray regions represent the regions with no statistically significant deviations after False Discovery Rate (FDR) correction.

3.3 Spatial patterns of outliers per region

Next, we examined the spatial patterns in outlier deviations for each region, calculated for each modality. Specifically, we estimated the outlier frequency of each region or the proportion of individuals whose deviations were outliers (FDR corrected). The outlier frequency was calculated for each CDR group in ADNI-ADS and ADRC-ADS and a CU group (ADNI-CU-test and ADRC-CU-test; see Section 2.2.1 and 2.3.1).

In general, regions exhibited an increasing number of outlier individuals within ADNI-ADS across the groups with increasing CDR (progressive stages of dementia), a trend that was observed for all the 3 modalities (Figure 2A) . For MRI, we observed more regions with outliers for [CDR >= 1; n = 40] and [CDR = 0.5; n = 26], compared to [CDR = 0; n = 14] and [CU-test; n = 0]. The hippocampal regions were the regions with the highest MRI outlier percentage in ADNI-ADS (47% of the [CDR>= 1] group, 25% of the [CDR = 0.5] group, 6% of [CDR = 0] and 3% of the CU-test group). Overall, more outlier frequency was observed in the hippocampal, frontal, temporal and parietal regions, which was also validated by the effect size maps in Figure 1A. Outliers in amyloid deviations were observed in 76 regions for [CDR >= 1], 85 regions for [CDR = 0.5], 83 regions for [CDR = 0] and 15 regions for the CU group respectively. The highest percentage of amyloid SUVR outliers in ADNI-ADS were observed in the precuneus and frontalpole regions (100% of the [CDR>= 1] group, 87% of the [CDR = 0.5] group, 71% of [CDR = 0] and 5% of the CU group). Outliers in tau SUVR uptake were observed in 82 regions for [CDR >= 1], 88 regions for [CDR = 0.5], 79 regions for [CDR = 0] and 54 regions for the CU-test group. The hippocampal and entorhinal regions were the regions with the highest tau outlier percentage in ADNI-ADS (84% of the [CDR>= 1] group, 65% of the [CDR = 0.5] group, 24% of [CDR = 0] and 14% of the CU-test group). As observed in the regional group differences for ADNI-ADS in Figure 1A, the hippocampal, frontal, temporal and parietal regions showed a general trend of

high outlier frequency for amyloid outliers. Likewise, the temporal, parietal and hippocampal regions exhibited high outlier frequency for tau outliers. Overall, we noticed that amyloid and tau corresponded to more regions with outliers and higher outlier frequency than MRI, a trend that was consistent across all the groups.

We observed a similar trend of regions exhibited an increasing number of outlier individuals within ADRC-ADS across the groups with increasing CDR values (Figure 2B). For all 3 modalities we observed similar regions with high outlier frequency as observed in Figure 2A. For MRI, we observed more regions with outliers for [CDR >= 1; n = 24] and [CDR = 0.5; n = 23], compared to [CDR = 0; n = 9] and [CU-test; n = 6]. Outliers in amyloid deviations were observed in 51 regions for [CDR >= 1], 75 regions for [CDR = 0.5], 82 regions for [CDR = 0] and 12 regions for the CU group respectively. For tau, outliers were observed in 54 regions for [CDR >= 1], 76 regions for [CDR = 0.5], 86 regions for [CDR = 0] and 68 regions for the CU group respectively. For each modality, the regions with highest proportion of outliers identified in ADNI-ADS were reproduced in ADRC-ADS too (hippocampal and temporal regions for MRI, frontal regions for amyloid and temporal and parietal regions for tau). One important observation was lower number of outlier regions for the [CDR = 1] group in ADRC-ADS compared to the [CDR = 1] in ADNI-ADS. This can be explained by the low sample size (n = 7) of the [CDR = 1] group in ADRC-ADS (Figure 2B).

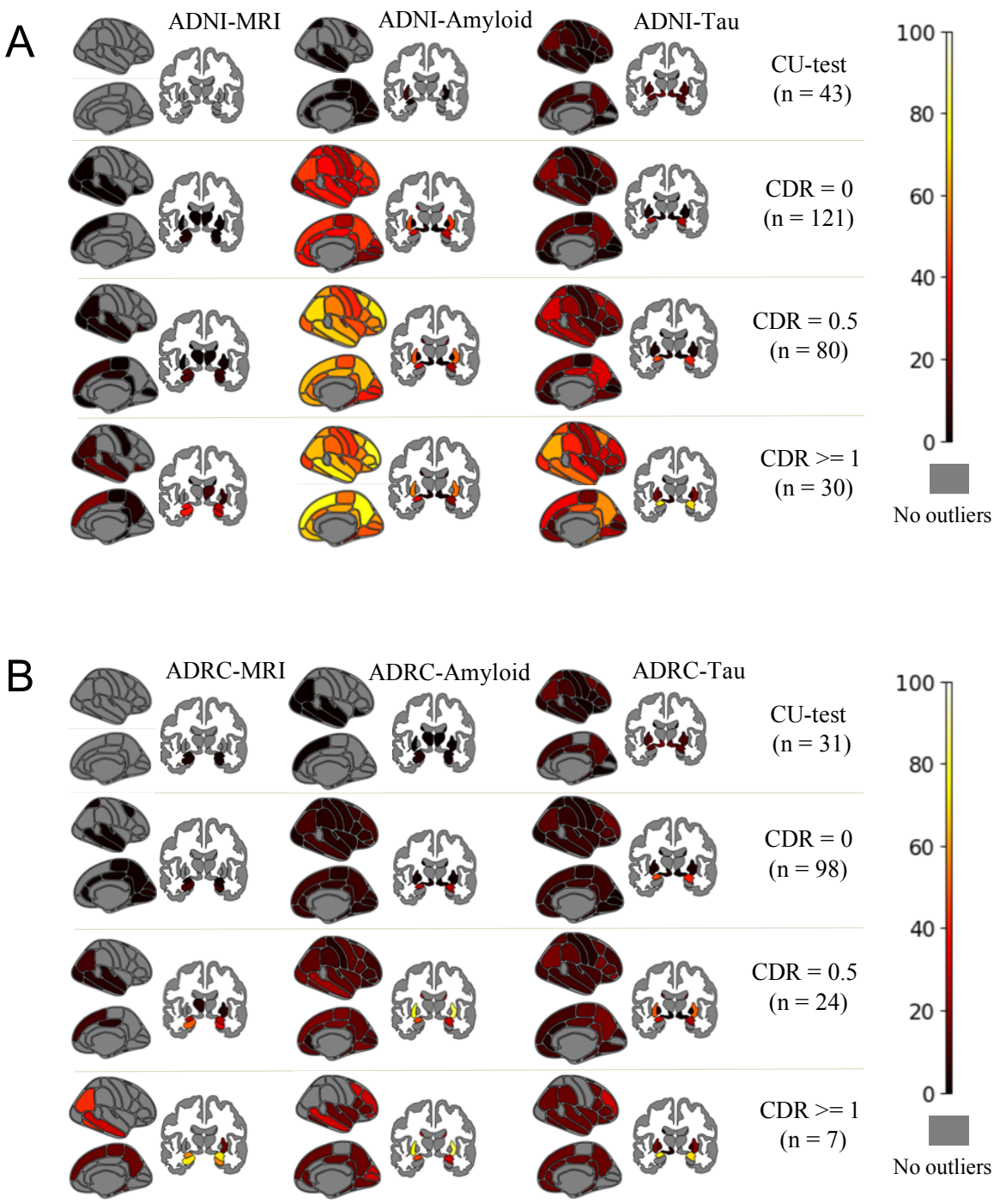

Figure 2: Brain atlas maps (Desikan-Killiany atlas for 66 cortical regions and Aseg atlas for 24 subcortical regions) showing the outlier frequency for each region in ADNI-ADS (**2A**) and ADRC-ADS (**2B).** The figures from left to right indicate the brain maps corresponding to MRI, amyloid and tau respectively. The color bar represents the outlier proportion of each region from 0 to 100%. Gray represents that no participants have outlier deviations for that region.

3.4 Individuals on the ADS are more dissimilar in outlier patterns for MRI atrophy, amyloid and tau loading than CU individuals.

We next calculated hamming distances for individuals in each CDR group in ADNI-ADS and ADRC-ADS and a group of the CU individuals included for comparison (ADNI-CU-test and ADRC-CU-test; see Section 2.2.1 and 2.3.1). The hamming distance distribution indicated the level of heterogeneity in outlier patters even within the same group.

Hamming distance distribution (kernel density estimation plots) calculated for all modalities (hamming_all, see Section 2.5.4), indicated greater within-group dissimilarity for ADNI-ADS individuals compared to CU participants (ADNI-CU-test) (Figure 3A). The median hamming distance significantly differed between the different groups overall ($p < 0.001$). The within-group dissimilarity is highest for ADNI-ADS individuals with CDR >= 1 (median 62, IQR 39, 95% CI 60.3–63.8) and lowest in the CU group (median 5, IQR 6, 95% CI 8.4–10.7). The dissimilarity in outlier patterns showed an increasing trend for higher CDR values (progressive stages of dementia). The CDR = 0.5 group in ADNI-ADS (median 56, IQR 34, 95% CI 58.1– 59.3) demonstrated more heterogenous outlier patterns than the CDR = 0 group (median 47, IQR 30, 95% CI 49.2–50.1) (Figure 3A). For individual modalities (hamming_mri, hamming_amyloid and hamming_tau), we observed a consistent pattern of higher within-group dissimilarity for groups with more progressive dementia. When compared across modalities for a particular group, we observed higher heterogeneity in outlier patterns for amyloid and tau compared to MRI (eFigure 4A).

We repeated the above experiments in the ADRC-dataset and observed a similar trend of greater within-group dissimilarity for ADRC-ADS individuals compared to CU participants (ADRC-CU-test) (Figure 3B). Similar to ADNI, the within-group dissimilarity is more for the CDR >= 1 (median 39, IQR 40, 95% CI 28.8–43.5) and CDR = 0.5 group (median 30, IQR 33, 95% CI 30.7– 34.5) compared to CDR = 0 (median 14, IQR 20, 95% CI 21.6 – 22.9) and the CU group (median 4, IQR 8, 95% CI 6.9–8.5). Similar patterns were also

observed for hamming distances calculated using each modality (hamming_mri, hamming_amyloid and hamming_tau). When compared across modalities for a particular group, we observed higher heterogeneity in outlier patterns for amyloid compared to MRI and tau (eFigure 4B).

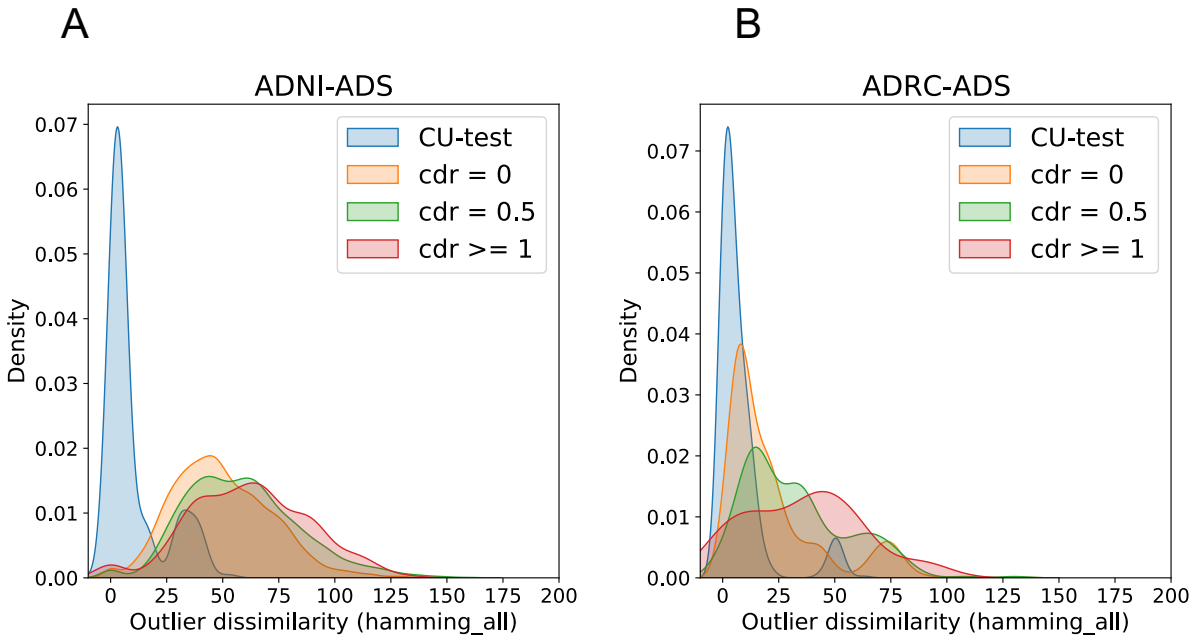

Figure 3: Hamming distance density (KDE plot) which illustrates the spread of outlier dissimilarity (calculated by the Hamming distance for all modalities or hamming_all; see Section 2.5.4) within each CDR group for ADNI-ADS (3A) and ADRC-ADS (3B). Higher hamming distance values indicated more heterogeneity in outlier patters even within a particular group.

3.5 DSI is associated with progressive stages of dementia.

Next, we compared DSI across each different CDR groups on the ADS (ADNI-ADS and ADRC-ADS) and a group of CU individuals included for comparison (ADNI-CU-test and ADRC-CU-test; Section 2.2.1 and 2.3.1). Higher DSI values indicate more outliers which translates to more atrophy for MRI and higher burden for amyloid and tau uptake respectively. Our goal was to validate if increase in DSI is associated with high CDR values (progressive stages of dementia).

DSI calculated across all modalities (DSI_all) show minimal values for ADNI-CU-test individuals (mean = 0.06, IQR = 0.03, 95% CI = [0.007-0.1]) and a monotonous increasing pattern across the dementia stages (Figure 4A). Specifically, maximum disease severity was observed for ADNI-ADS individuals with CDR >= 1 (mean = 1.8, IQR = 1.5, 95% CI = [1.31-2.26]). Comparatively lower DSI values were observed for ADNI-ADS individuals with no dementia (CDR = 0; mean = 0.45, IQR = 0.6, 95% CI = [0.39-0.52] and very mild dementia (CDR = 0.5; mean = 1.1, IQR = 1.3, 95% CI = [0.87-1.25]. Pairwise group differences (FDR-corrected post hoc Tukey comparisons) were statistically significant ($p < 0.05$) except between CDR = 0.5 and CDR >= 1 (Figure 4A). When compared across individual modalities, DSI values showed a similar monotonous increasing pattern across the CDR groups for all the 3 modalities (DSI_mri, DSI_amyloid, DSI_tau, eFigure 5A). Disease severity related to amyloid and tau loading (DSI_amyloid, DSI_tau) exhibited an overall trend of higher DSI across the CDR groups compared to DSI related to MRI atrophy (DSI_mri). Like DSI_all, pairwise group differences were statistically significant for the individual modalities except the following group pairs: [CDR = 0.5 and CDR >= 1] for DSI_mri, DSI_amyloid, DSI_tau, [CU-test and CDR = 0] for DSI_mri and DSI_tau (eFigure 5A).

We repeated the above experiments in the ADRC-ADS cohort and observed similar patterns of DSI increasing with the progressive stages of dementia (Figure 4 right). Similar to ADNI-ADS, DSI_all was highest for individuals with CDR >= 1 (mean = 0.46, IQR = 0.28, 95% CI = [0.25-0.6]) and lowest for ADRC-CU-test individuals (mean = 0.006, IQR = 0.005, 95% CI = [0.002-0.008]) (Figure 4B). ADRC-ADS individuals with CDR = 0 (mean = 0.21, IQR = 0.22, 95% CI = [0.12-0.37]) and CDR = 0.5 exhibited intermediate severity values (mean = 0.32, IQR = 0.32, 95% CI = [0.21-0.5]). Pairwise group differences were statistically significant ($p < 0.05$) except between CDR = 0.5 and CDR >= 1 (Figure 4B). An important

observation from our results is that higher disease severity was associated with ADNI-ADS in general compared to ADRC-ADS. This can be attributed to the fact that ADNI had more individuals on the ADS compared to the ADRC dataset. Modality wise DSI for ADRC-ADS exhibited similar trends as observed in ADNI (eFigure 5B). DSI Amyloid and tau loading were associated with higher disease severity (DSI_amyloid, DSI_tau) across the CDR groups compared to MRI atrophy.

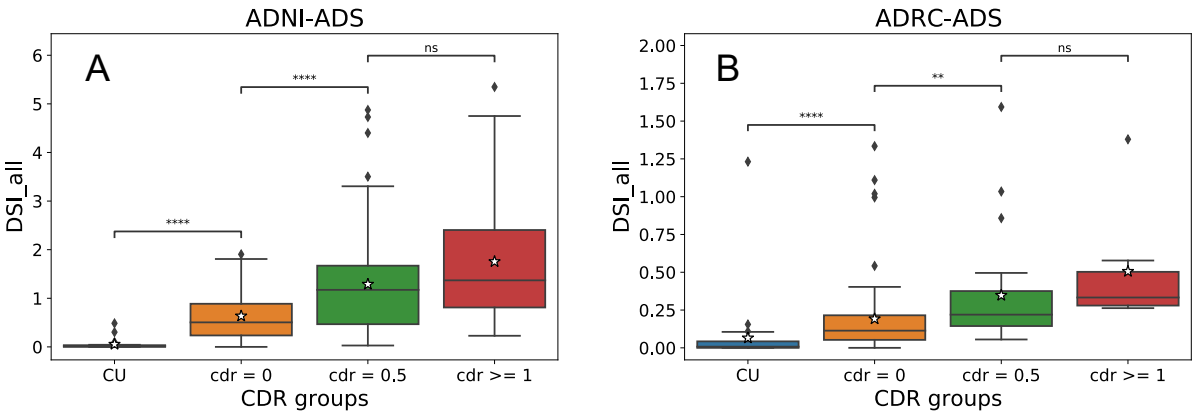

Figure 4: Box plot showing DSI_all (DSI across all modalities; see Section 2.5.5) for both ADNI-ADS (**4A**) and ADRC-ADS (**4B**). The x-axis shows the different CDR groups in the ADS and CU-test (Section 2.2.1 and 2.3.1). FDR-corrected post hoc Tukey comparisons used to assess pairwise group differences. Abbreviations: DSI: Disease Severity Index, CDR = Clinical Dementia Rating. Statistical annotations: ns: not significant $0.05 < p <= 1$, * $0.01 < p <= 0.05$, ** $0.001 < p < 0.01$, *** $p < 0.001$.

3.6 DSI showed significant associations with neuropsychological composites scores.

We also assessed how DSI is related with neuropsychological composite scores (memory, executive functioning and language) for both ADNI-ADS and ADRC-ADS cohorts. Linear regression adjusted for age and sex showed that DSI across all modalities (DSI_all) in ADNI-ADS is significantly associated with memory ($\beta = -0.6$; $p < 0.001$; $r = -0.62$), executive functioning ($\beta = -0.46$; $p < 0.001$; $r = -0.54$) and language ($\beta = -0.39$; $p < 0.001$; $r = -0.47$), where r indicated the Pearson correlation coefficient (Table 2). The negative correlation between DSI_all and the scores can be validated by the fact that high DSI is related to low

scores which translates to more impairment. DSI for individual modalities (DSI_mri, DSI_amyloid, DSI_tau) also exhibited significant associations with each of the 3 neuropsychological composites (eTable 2). The correlation coefficients were in general higher for DSI_tau compared to DSI_mri and DSI_amyloid.

We observed similar trends of significant associations between DSI and the composites in ADRC-ADS, despite lower levels of DSI on average (Table 2). Like ADNI-ADS, DSI_all for ADRC-ADS was related to memory ($\beta = -0.71$; $p < 0.001$; $r = -0.68$), executive functioning ($\beta = -0.52$; $p < 0.001$; $r = -0.56$) and language ($\beta = -0.36$; $p < 0.001$; $r = -0.41$). As observed in ADNI, DSI for individual modalities were also related to each of the composites in ADRC-ADS. For ADRC-ADS, DSI_amyloid and DSI_tau showed higher correlation overall compared to DSI_mri (eTable 2).

	Composite scores	ADNI-ADS			ADRC-ADS		
		β	p	r	β	p	r
DSI_all	Memory	- 0.6	p < 0.001	- 0.62	- 0.71	p < 0.001	-0.68
	Executive	- 0.46	p < 0.001	- 0.54	- 0.52	p < 0.001	-0.56
	Language	- 0.39	p < 0.001	- 0.47	- 0.36	p < 0.001	-0.41

Table 2: Association between DSI across all modalities (DSI_all) and the neuropsychological composite scores (memory, executive functioning, and language) for both ADNI-ADS and ADRC-ADS. β and p represent the slope and p-value for linear regression, adjusted for age and sex. r represents the Pearson correlation coefficient.

3.7 High disease severity index is related to the longitudinal risk of dementia progression.

Finally, we used DSI_all (DSI across all modalities) to model risk of disease progression using survival analysis. Longitudinal CDR information was available for 175 individuals in ADNI-ADS and 85 individuals in ADRC-ADS who had none or very mild dementia at baseline (CDR = 0 or CDR = 0.5). In ADNI-ADS, DSI_all (DSI across all modalities) had significant associations with the conversion to mild or more severe dementia (CDR >= 1; p <

0.001; Figure 6A). Individuals in the higher DSI quantiles (higher DSI values), particularly q4 ($p < 0.001$) and q2 ($p < 0.01$) showed a higher progression risk compared to individuals in lower quantile q1 (less DSI), namely q1 and q2. (eTable 3). We repeated the survival analysis in ADRC-ADS and observed similar association between DSI and dementia conversion as observed in ADNI-ADS (Figure 6B). In ADRC-ADS, individuals in q4 ($p < 0.001$) and q3 ($p < 0.01$) exhibited a faster progression to dementia compared to individuals in q1 and q2 (eTable 3).

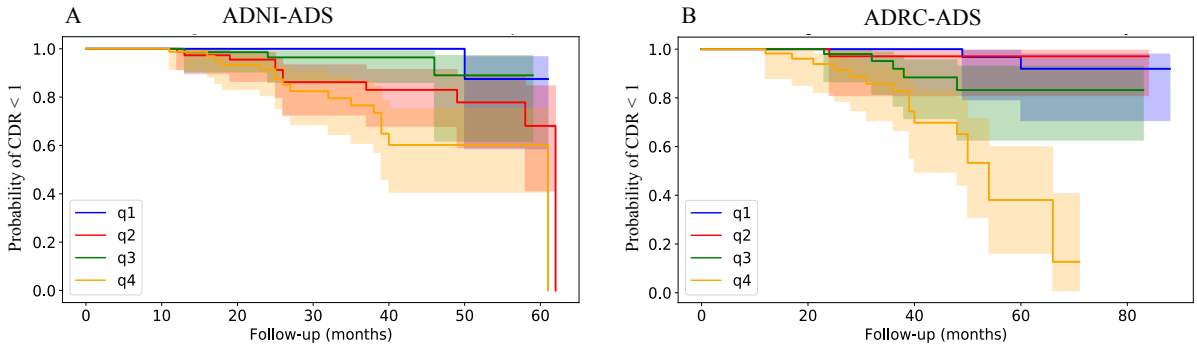

Figure 5: Kaplan-Meier plot of conversion from CDR < 1 to CDR >=1 for ADNI-ADS **(Figure 5A)** and ADRC-ADS **(Figure 5B)** participants. The x-axis and the y-axis represent the follow-up period (in months) and the progression probability of CDR <1 respectively. The 4 lines represent the 4 quantiles of DSI_all (DSI across all modalities), shown by blue, red, green and orange respectively. The filled color span represents the 95% confidence intervals.

4. Discussion

In this study, we applied data-driven methods to examine individual patterns of variability in brain structure across individuals on the ADS. The major contributions of our work are as follows. First, we used normative modelling across multiple modalities to quantify the extent to which individuals on the ADS deviate from the control CU group with respect to MRI gray matter volumes, amyloid and tau loading. Our results showed evidence in two independent datasets that ADS individuals with moderate or severe dementia were more heterogenous in regional outlier patterns for MRI atrophy, amyloid and tau loading compared to ADS individuals with early or mild dementia. Next, we used a disease severity index (DSI) to quantify both the spread and magnitude of outlier deviations across all 3 modalities. We observed that (i) DSI was associated with the progressive stages of dementia, (ii) showed significant associations with neuropsychological composite scores and (iii) related to the longitudinal risk of CDR progression. The observations were reproducible in both the discovery and replication datasets, which demonstrated the generalizability of our scientific findings.

Existing subtyping studies have analyzed heterogeneity in AD by assessing group average relationships and assuming homogeneity within each subtype. In contrast, we adopted a normative modelling technique which allowed us to learn the normative distribution with the healthy brain patterns of a CU group. An individual on the ADS were then located within the normative distribution to assess what extent they deviate from the norm at any given brain region, providing a map of individual variability. Recent normative modelling works have used Gaussian regression models as normative models for each brain region and used that to analyze the heterogeneity in MRI brain volume/cortical thickness outliers.[20,21,56] Our work differs from the above studies in two aspects. First, we adopted a

deep learning based normative model (like Pinaya et al.) using a variational autoencoder which learned a single normative model for all features instead of separate normative models for individual features. Second, we went beyond the idea of using only a single modality (MRI volumes/cortical thickness) as features to learn the normative distribution. Keeping in mind the multifaceted aspect of AD, we included multiple modalities in the form of MRI volumes, amyloid PET SUVR, and tau PET SUVR for learning the healthy brain patterns of CU individuals. Our is the first study to examine the heterogeneity in AD through the lens of multiple neuroimaging modalities.

Our findings were consistent the established understanding of the neurobiology of AD. We observed that ADS individuals with advanced stages of dementia (CDR = 0.5 and CDR >= 1) have higher DSI and a greater number of outlier regions compared to the preclinical ADS (CDR = 0) and the CU group. This is consistent with the evidence of loss in gray matter volumes (MRI atrophy) and increase in amyloid and tau SUVR uptake because of AD neuropathology.[57,58] The significant association between DSI and poor cognitive performance is also consistent with previous studies where atrophy, amyloid and tau binding were shown to be associated with cognitive decline.[59,60]. We observed more MRI outlier deviations in temporal regions such as the hippocampus and the medial temporal regions. These are areas known to be associated with neurodegeneration and the onset of clinical symptoms related to typical AD[61]. However, some atrophy was also observed in the frontal and parietal regions suggesting evidence of atypical AD variants like frontal variant AD (fvAD).[62] Similarly, more tau outliers were observed in the entorhinal cortex, parahippocampus, fusiform, posterior cingulate, temporal, parietal, and frontal brain regions, areas known to have specific binding to AV1451 tau PET.[63] Amyloid outlier deviations were observed in the precuneus, frontal and temporal regions, which are sensitive to ^{18}F AV45 amyloid PET accumulation.[64,65].

Spatial heterogeneity in AD refers to the non-uniform distribution of pathological changes and brain abnormalities across different brain regions. For example, if atrophy pattern is fully homogenous, it is consistent across individuals or brain regions, indicating that most ADS individuals are expected to have outliers in the same regions where MRI atrophy is observed. On the other hand, a heterogeneous pattern does not fully overlap between individuals, indicating that different individuals may exhibit outlier patterns in different brain regions for these markers. Looking beyond the group-average modality-specific regional differences, it was observed that the highest proportion of outliers in a single region was 54%, 75% and 94% for MRI, tau, and amyloid respectively. This highlighted the fact that amyloid was more homogenous in spatial patterns of outliers. However, MRI and tau were more heterogenous as individual spatial outlier patterns only partially overlap between ADS individuals; if the outlier pattern is homogenous for a region (as assumed within group averages), 100% of subjects are expected to have outliers in that region. The spatial heterogeneity of MRI atrophy was also consistent with the observations in Verdi et. al where highest proportion of cortical thickness outliers in a single region was less than 50% in the AD group.

The DSI values calculated for every individual and for each modality were essentially Z-scores estimating the extent to which their regional patterns deviated from the patterns of the healthy control distribution. Recent studies have also adopted similar patient-level deviation metrics as a clinical measure in the form of brain volume/thickness Z-scores. Specifically, they calculated either the average deviation in MRI gray matter volumes across all regions or the total number of MRI outlier deviations (TOC) across all regions.[20,21,30]. There are also studies quantifying the spatial spread of tau loading (TSS) for predicting cognitive impairment and disease progression.[66,67] In contrast to the existing approaches, DSI not only captured the spatial spread of outliers but also the magnitude of regional outlier

deviations both independently and aggregated across all modalities. Most importantly, DSI_all, calculated across all the modalities represented a marker of disease progression due to the cumulative effect of both neuroanatomical (MRI atrophy) and neuropathological (amyloid and tau loading) changes in the brain. Consistent with the findings in previous work[20], DSI was associated with the progressive stages of dementia, showed significant associations with neuropsychological composite scores and were also related to the longitudinal risk of CDR progression. We believe that DSI has the potential to be used as an individual patient metric of neurodegeneration that can help in clinical decision making. Particularly, in light of recent AD treatments using FDA-approved anti-amyloid drugs like Aducanumab and Lecanemab, DSI can be potentially used to monitor the amyloid burden and track patient response to treatments.[68,69]

Strengths, limitations, and scope for future work

Our study has the following points of strength. First, we specifically included only amyloid positive subjects for the ADS cohort, regardless of cognitive status. Thus, the normative modelling deviations and modality-specific outlier patterns we learned were specifically relevant to AD rather than other dementias like vascular dementia[70] or frontotemporal dementia.[71] Further, most of our findings were reproducible in two independent cohorts, both of which are freely available to scientific community following formal data use agreements. This supports the generalizability and the relevance of the patterns we identified in our investigation of AD heterogeneity across multiple neuroimaging modalities.

However, there are certain limitations that need to be considered regarding our analyses. First, we used cross-sectional imaging data for our normative modeling framework, providing a snapshot of the disease at a specific time. Due to our constraint that the MRI, amyloid and tau dates were within 1 year of one another, it is often challenging to obtain significant number of multimodal follow-up visits. However, heterogeneity in AD has been

demonstrated to vary over time.[72] In future applications of multimodal normative modeling on AD, it is important to incorporate serial neuroimaging data collected across multiple time points. Second, our results demonstrated that even CU individuals exhibited outlier deviations in our study, indicating a certain level of within-group heterogeneity (Figures 1,2 and 4). We believe that this was probably due to the smaller sample size of our CU cohort in ADNI and the Knight ADRC dataset. Non-zero deviations for the control group were also observed in the findings of Verdi et. al [20], which assumed individuals with the diagnosis label "CN" in ADNI as the control cohort to measure heterogeneity in cortical thickness atrophy. This observation suggests that caution should be exercised when assuming homogeneity even within control groups. Also, future normative modelling works should consider a more diverse population of the reference control group with a larger sample size for a more accurate representation of the healthy brain. Finally, another important caveat to consider is that the imaging scans in the discovery and replication cohorts were processed with a different version of FreeSurfer (FreeSurfer 6 for ADNI and FreeSurfer 5.3 for ADRC). Although we fine-tuned our pre-trained normative model on the replication cohort, the different versions of FreeSurfer may potentially add noise to the normative model. To address this issue, it is important to consider harmonization methods like COMBAT [73] for future multicenter studies on normative modelling.

Conclusions

Our is the first study to examine the heterogeneity in AD through the lens of multiple neuroimaging modalities (ATN), based on distinct or overlapping patterns of regional outlier deviations. We calculated a disease severity index (DSI) for each ADS individual using the spatial spread and the magnitude of outlier deviations across all modalities. DSI was associated with the progressive stages of dementia, (ii) showed significant associations with

neuropsychological composite scores and (iii) related to the longitudinal risk of CDR progression. DSI has the potential to be an individual patient metric of neurodegeneration that can help in clinical decision making and monitoring patient response for anti-amyloid treatments. Our findings were reproducible in both the discovery and replication datasets, which demonstrated the generalizability of our scientific findings.

Acknowledgements

We would like to thank the staff of the Washington University Center for High Performance Computing who helped enable this work.

Funding Sources

The preparation of this report was supported by the Centene Corporation contract (P19-00559) for the Washington University-Centene ARCH Personalized Medicine Initiative and the National Institutes of Health (NIH) (R01-AG067103). Computations were performed using the facilities of the Washington University Center for High-Performance Computing, which were partially funded by NIH grant S10 OD025200. Additional support is provided by The McDonnell Center for Systems Neuroscience.

Data collection and sharing for this project was funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U01 AG024904) and DOD ADNI (Department of Defense award number W81XWH-12-2-0012). ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: AbbVie, Alzheimer’s Association; Alzheimer’s Drug Discovery Foundation; Araclon Biotech; BioClinica, Inc.; Biogen; Bristol-Myers Squibb Company; CereSpir, Inc.; Cogstate; Eisai Inc.; Elan Pharmaceuticals, Inc.; Eli Lilly and Company; EuroImmun; F. Hoffmann-La

Roche Ltd and its affiliated company Genentech, Inc.; Fujirebio; GE Healthcare; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research & Development, LLC.; Johnson & Johnson Pharmaceutical Research & Development LLC.; Lumosity; Lundbeck; Merck & Co., Inc.; Meso Scale Diagnostics, LLC.; NeuroRx Research; Neurotrack Technologies; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal Imaging; Servier; Takeda Pharmaceutical Company; and Transition Therapeutics. The Canadian Institutes of Health Research is providing funds to support ADNI clinical sites in Canada. Private sector contributions are facilitated by the Foundation for the National Institutes of Health (www.fnih.org). The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer's Therapeutic Research Institute at the University of Southern California. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of Southern California.

Data were also provided (in part) by Knight Alzheimer Disease Research Center (ADRC), supported by grants P50-AG05681, P01- AG03991, and P01-AG26276 awarded to Dr Morris. AV-45 doses were provided by Avid Radiopharmaceuticals, a wholly owned subsidiary of Eli Lilly.

Competing Interests

Author AS has equity in TheraPanacea and has received personal compensation for serving as a grant reviewer for BrightFocus Foundation. The remaining authors have no conflicting interests to report.

Collaborators from ADNI

Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). As such, the investigators

within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. A complete listing of ADNI investigators can be found at: http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf

References

1. Kumar S, Oh I, Schindler S, Lai AM, Payne PR, Gupta A. Machine learning for modeling the progression of Alzheimer disease dementia using clinical data: a systematic literature review. JAMIA open. Oxford University Press; 2021;4:ooab052.

2. Richards M, Brayne C. What do we mean by Alzheimer's disease? BMJ. British Medical Journal Publishing Group; 2010;341.

3. Jack CR, Knopman DS, Jagust WJ, et al. Hypothetical model of dynamic biomarkers of the Alzheimer's pathological cascade. The Lancet Neurology. Elsevier; 2010;9:119–128.

4. Seeley WW. Mapping neurodegenerative disease onset and progression. Cold Spring Harbor Perspectives in Biology. Cold Spring Harbor Lab; 2017;9:a023622.

5. Mehta RI, Schneider JA. What is 'Alzheimer's disease'? The neuropathological heterogeneity of clinically defined Alzheimer's dementia. Current opinion in neurology. LWW; 2021;34:237–245.

6. Marquand AF, Wolfers T, Mennes M, Buitelaar J, Beckmann CF. Beyond lumping and splitting: a review of computational approaches for stratifying psychiatric disorders. Biological psychiatry: cognitive neuroscience and neuroimaging. Elsevier; 2016;1:433–447.

7. Graff-Radford J, Yong KX, Apostolova LG, et al. New insights into atypical Alzheimer's disease in the era of biomarkers. The Lancet Neurology. Elsevier; 2021;20:222–234.

8. 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. Biological psychiatry. Elsevier; 2020;88:70–82.

9. Dong A, Toledo JB, Honnorat N, et al. Heterogeneity of neuroanatomical patterns in prodromal Alzheimer's disease: links to cognition, progression and biomarkers. Brain. Oxford University Press; 2017;140:735–747.

10. Wen J, Varol E, Sotiras A, et al. Multi-scale semi-supervised clustering of brain images: Deriving disease subtypes. Medical image analysis. Elsevier; 2022;75:102304.

11. Ten Kate M, Dicks E, Visser PJ, et al. Atrophy subtypes in prodromal Alzheimer's disease are associated with cognitive decline. Brain. Oxford University Press; 2018;141:3443–3456.

12. Young PN, Estarellas M, Coomans E, et al. Imaging biomarkers in neurodegeneration: current and future practices. Alzheimer's research & therapy. BioMed Central; 2020;12:1–17.

13. Mohanty R, Mårtensson G, Poulakis K, et al. Comparison of subtyping methods for neuroimaging studies in Alzheimer's disease: a call for harmonization. Brain Communications. 2020;2:fcaa192.

14. Bethlehem RA, Seidlitz J, Romero-Garcia R, Trakoshis S, Dumas G, Lombardo MV. A normative modelling approach reveals age-atypical cortical thickness in a subgroup of males with autism spectrum disorder. Communications biology. Nature Publishing Group UK London; 2020;3:486.

15. Huizinga W, Poot DH, Vernooij MW, et al. A spatio-temporal reference model of the aging brain. NeuroImage. Elsevier; 2018;169:11–22.

16. Kia SM, Marquand AF. Neural processes mixed-effect models for deep normative modeling of clinical neuroimaging data. International Conference on Medical Imaging with Deep Learning. PMLR; 2019. p. 297–314.

17. Marquand AF, Rezek I, Buitelaar J, Beckmann CF. Understanding heterogeneity in clinical cohorts using normative models: beyond case-control studies. Biological psychiatry. Elsevier; 2016;80:552–561.

18. Ziegler G, Ridgway GR, Dahnke R, Gaser C, Initiative ADN. Individualized Gaussian process-based prediction and detection of local and global gray matter abnormalities in elderly subjects. Neuroimage. Elsevier; 2014;97:333–348.

19. Wolfers T, Doan NT, Kaufmann T, et al. Mapping the heterogeneous phenotype of schizophrenia and bipolar disorder using normative models. JAMA psychiatry. American Medical Association; 2018;75:1146–1155.

20. Verdi S, Kia SM, Yong KX, et al. Revealing individual neuroanatomical heterogeneity in Alzheimer disease using neuroanatomical normative modeling. Neurology. AAN Enterprises; 2023;100:e2442–e2453.

21. Pinaya WH, Scarpazza C, Garcia-Dias R, et al. Using normative modelling to detect disease progression in mild cognitive impairment and Alzheimer's disease in a cross-sectional multi-cohort study. Scientific reports. Springer; 2021;11:1–13.

22. Pinaya WH, Mechelli A, Sato JR. Using deep autoencoders to identify abnormal brain structural patterns in neuropsychiatric disorders: A large-scale multi-sample study. Human brain mapping. Wiley Online Library; 2019;40:944–954.

23. Johnson KA, Fox NC, Sperling RA, Klunk WE. Brain imaging in Alzheimer disease. Cold Spring Harbor perspectives in medicine. Cold Spring Harbor Laboratory Press; 2012;2:a006213.

24. Dubois B, Hampel H, Feldman HH, et al. Preclinical Alzheimer's disease: definition, natural history, and diagnostic criteria. Alzheimer's & Dementia. Elsevier; 2016;12:292–323.

25. Lorenzi M, Simpson IJ, Mendelson AF, et al. Multimodal image analysis in Alzheimer's disease via statistical modelling of non-local intensity correlations. Scientific reports. Nature Publishing Group; 2016;6:1–8.

26. Perrin RJ, Fagan AM, Holtzman DM. Multimodal techniques for diagnosis and prognosis of Alzheimer's disease. Nature. Nature Publishing Group UK London; 2009;461:916–922.

27. Teipel S, Drzezga A, Grothe MJ, et al. Multimodal imaging in Alzheimer's disease: validity and usefulness for early detection. The Lancet Neurology. Elsevier; 2015;14:1037–1053.

28. Alberdi A, Aztiria A, Basarab A. On the early diagnosis of Alzheimer's Disease from multimodal signals: A survey. Artificial intelligence in medicine. Elsevier; 2016;71:1–29.

29. Zhang D, Wang Y, Zhou L, Yuan H, Shen D, Initiative ADN. Multimodal classification of Alzheimer's disease and mild cognitive impairment. Neuroimage. Elsevier; 2011;55:856–867.

30. Kumar S, Payne PR, Sotiras A. Normative modeling using multimodal variational autoencoders to identify abnormal brain volume deviations in Alzheimer's disease. Medical Imaging 2023: Computer-Aided Diagnosis. SPIE; 2023. p. 1246503.

31. Kumar S, Payne P, Sotiras A. Improving Normative Modeling for Multi-modal Neuroimaging Data using mixture-of-product-of-experts variational autoencoders. arXiv preprint arXiv:231200992. Epub 2023.

32. Van Der Flier WM, Scheltens P. The ATN framework—moving preclinical Alzheimer disease to clinical relevance. JAMA neurology. American Medical Association; 2022;79:968–970.

33. Mueller SG, Weiner MW, Thal LJ, et al. Ways toward an early diagnosis in Alzheimer's disease: the Alzheimer's Disease Neuroimaging Initiative (ADNI). Alzheimer's & Dementia. Elsevier; 2005;1:55–66.

34. Desikan RS, Ségonne F, Fischl B, et al. An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. Neuroimage. Elsevier; 2006;31:968–980.

35. Fischl B, Salat DH, Busa E, et al. Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain. Neuron. Elsevier; 2002;33:341–355.

36. Landau SM, Breault C, Joshi AD, et al. Amyloid-β imaging with Pittsburgh compound B and florbetapir: comparing radiotracers and quantification methods. Journal of Nuclear Medicine. Soc Nuclear Med; 2013;54:70–77.

37. Landau SM, Thomas BA, Thurfjell L, et al. Amyloid PET imaging in Alzheimer's disease: a comparison of three radiotracers. European journal of nuclear medicine and molecular imaging. Springer; 2014;41:1398–1407.

38. Clark CM, Schneider JA, Bedell BJ, et al. Use of florbetapir-PET for imaging β-amyloid pathology. Jama. American Medical Association; 2011;305:275–283.

39. Joshi AD, Pontecorvo MJ, Clark CM, et al. Performance characteristics of amyloid PET with florbetapir F 18 in patients with Alzheimer's disease and cognitively normal subjects. Journal of Nuclear Medicine. Soc Nuclear Med; 2012;53:378–384.

40. Su Y, Blazey TM, Snyder AZ, et al. Partial volume correction in quantitative amyloid imaging. Neuroimage. Elsevier; 2015;107:55–64.

41. Su Y, D'Angelo GM, Vlassenko AG, et al. Quantitative analysis of PiB-PET with freesurfer ROIs. PloS one. Public Library of Science San Francisco, USA; 2013;8:e73377.

42. Su Y, Flores S, Hornbeck RC, et al. Utilizing the Centiloid scale in cross-sectional and longitudinal PiB PET studies. NeuroImage: Clinical. Elsevier; 2018;19:406–416.

43. Aschenbrenner AJ, Gordon BA, Benzinger TL, Morris JC, Hassenstab JJ. Influence of tau PET, amyloid PET, and hippocampal volume on cognition in Alzheimer disease. Neurology. AAN Enterprises; 2018;91:e859–e866.

44. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. Journal of the Royal statistical society: series B (Methodological). Wiley Online Library; 1995;57:289–300.

45. Sawilowsky SS. New effect size rules of thumb. Journal of modern applied statistical methods. 2009;8:26.

46. Cohen J. Statistical power analysis. Current directions in psychological science. Sage Publications Sage CA: Los Angeles, CA; 1992;1:98–101.

47. Mowinckel AM, Vidal-Piñeiro D. Visualization of brain statistics with R packages ggseg and ggseg3d. Advances in Methods and Practices in Psychological Science. SAGE Publications Sage CA: Los Angeles, CA; 2020;3:466–483.

48. Initiative ADN. Four distinct trajectories of tau deposition identified in Alzheimer's disease. Nature medicine. Nature Publishing Group; 2021;27:871–881.

49. Wechsler D. Wechsler memory scale-revised. Psychological Corporation. Epub 1987.

50. Rey A. L'examen clinique en psychologie. Presses Universitaries De France; Epub 1958.

51. Grober E, Buschke H. Genuine memory deficits in dementia. Developmental neuropsychology. Taylor & Francis; 1987;3:13–36.

52. Reitan RM. Validity of the Trail Making Test as an indicator of organic brain damage. Perceptual and motor skills. SAGE Publications Sage CA: Los Angeles, CA; 1958;8:271–276.

53. Gladsjo JA, Schuman CC, Evans JD, Peavy GM, Miller SW, Heaton RK. Norms for letter and category fluency: demographic corrections for age, education, and ethnicity. Assessment. Sage Publications Sage CA: Thousand Oaks, CA; 1999;6:147–178.

54. Kaplan E, Goodglass H, Weintraub S. Boston naming test. Pro-ed Austin; Epub 2001.

55. Gollan TH, Weissberger GH, Runnqvist E, Montoya RI, Cera CM. Self-ratings of spoken language dominance: A Multilingual Naming Test (MINT) and preliminary norms for young and aging Spanish–English bilinguals. Bilingualism: language and cognition. Cambridge University Press; 2012;15:594–615.

56. Loreto F, Verdi S, Kia SM, et al. Examining real-world Alzheimer's disease heterogeneity using neuroanatomical normative modelling. medRxiv. Cold Spring Harbor Laboratory Press; Epub 2022.:2022.11. 02.22281597.

57. Vemuri P, Lowe VJ, Knopman DS, et al. Tau-PET uptake: Regional variation in average SUVR and impact of amyloid deposition. Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring. Elsevier; 2017;6:21–30.

58. Jack CR, Shiung MM, Gunter JL, et al. Comparison of different MRI brain atrophy rate measures with clinical disease progression in AD. Neurology. AAN Enterprises; 2004;62:591–600.

59. Koychev I, Gunn RN, Firouzian A, et al. PET tau and amyloid-β burden in mild Alzheimer's disease: divergent relationship with age, cognition, and cerebrospinal fluid biomarkers. Journal of Alzheimer's Disease. IOS Press; 2017;60:283–293.

60. McDonald CR, Gharapetian L, McEvoy LK, et al. Relationship between regional atrophy rates and cognitive decline in mild cognitive impairment. Neurobiology of aging. Elsevier; 2012;33:242–253.

61. Popuri K, Ma D, Wang L, Beg MF. Using machine learning to quantify structural MRI neurodegeneration patterns of Alzheimer's disease into dementia score: Independent validation on 8,834 images from ADNI, AIBL, OASIS, and MIRIAD databases. Human Brain Mapping. Wiley Online Library; 2020;41:4127–4147.

62. Blennerhassett R, Lillo P, Halliday GM, Hodges JR, Kril JJ. Distribution of pathology in frontal variant Alzheimer's disease. Journal of Alzheimer's Disease. IOS Press; 2014;39:63–70.

63. Zhao Q, Liu M, Ha L, Zhou Y, Initiative ADN. Quantitative 18F-AV1451 brain tau PET imaging in cognitively normal older adults, mild cognitive impairment, and Alzheimer's disease patients. Frontiers in Neurology. Frontiers Media SA; 2019;10:486.

64. Wong DF, Rosenberg PB, Zhou Y, et al. In vivo imaging of amyloid deposition in Alzheimer disease using the radioligand 18F-AV-45 (flobetapir F 18). Journal of nuclear medicine. Soc Nuclear Med; 2010;51:913–920.

65. Levitis E, Vogel JW, Funck T, et al. Differentiating amyloid beta spread in autosomal dominant and sporadic Alzheimer's disease. Brain Communications. Oxford University Press; 2022;4:fcac085.

66. McCullough AA, Gordon BA, Christensen J, et al. P3-401: EXAMINING THE ABILITY OF A TAU SPATIAL SPREAD METRIC TO INDICATE DISEASE PROGRESSION COMPARED TO AN INTENSITY-BASED APPROACH. Alzheimer's & Dementia. Wiley Online Library; 2018;14:P1255–P1256.

67. Doering S, McCullough AA, Gordon BA, et al. Evaluating Regional Importance for Tau Spatial Spread in Predicting Cognitive Impairment with Machine Learning. Alzheimer's & Dementia. Wiley Online Library; 2023;19:e082553.

68. Cummings J, Apostolova L, Rabinovici GD, et al. Lecanemab: appropriate use recommendations. The journal of prevention of Alzheimer's disease. Springer; 2023;10:362–377.

69. Tanzi RE. FDA approval of aduhelm paves a new path for Alzheimer's disease. ACS chemical neuroscience ACS Publications; 2021. p. 2714–2715.

70. T O'Brien J, Thomas A. Vascular dementia. The Lancet. Elsevier; 2015;386:1698–1706.

71. Bang J, Spina S, Miller BL. Frontotemporal dementia. The Lancet. Elsevier; 2015;386:1672–1682.

72. Ferreira D, Verhagen C, Hernández-Cabrera JA, et al. Distinct subtypes of Alzheimer's disease based on patterns of brain atrophy: longitudinal trajectories and clinical applications. Scientific reports. Nature Publishing Group UK London; 2017;7:46263.

73. Orlhac F, Eertink JJ, Cottereau A-S, et al. A guide to ComBat harmonization of imaging biomarkers in multicenter studies. Journal of Nuclear Medicine. Soc Nuclear Med; 2022;63:172–179.

74. Fischl B. FreeSurfer. Neuroimage. Elsevier; 2012;62:774–781.

75. Mishra S, Gordon BA, Su Y, et al. AV-1451 PET imaging of tau pathology in preclinical Alzheimer disease: defining a summary measure. Neuroimage. Elsevier; 2017;161:171–178.

76. Wu M, Goodman N. Multimodal generative models for scalable weakly-supervised learning. Advances in neural information processing systems. 2018;31.

77. Cao Y, Fleet DJ. Generalized product of experts for automatic and principled fusion of Gaussian process predictions. arXiv preprint arXiv:14107827. Epub 2014.