Revealing Individual Neuroanatomical Heterogeneity in Alzheimer Disease Using Neuroanatomical Normative Modeling

Author(s):
Serena Verdi, MPhil¹,²; Seyed Mostafa Kia, PhD³,⁴; Keir X. X Yong, PhD⁵; Duygu Tosun, PhD⁵; Jonathan M. Schott²; Andre F Marquand, PhD⁶,⁷; James H Cole, PhD²,³

Corresponding Author:
James H Cole, james.cole@ucl.ac.uk

Affiliation Information for All Authors: 1. Centre for Medical Image Computing, Medical Physics and Biomedical Engineering, University College London, London, UK; 2. Dementia Research Centre, UCL Queen Square Institute of Neurology, London, UK; 3. Donders Centre for Cognitive Neuroimaging, Donders Institute for Brain, Cognition and Behaviour, Radboud University, Nijmegen, 6525EN, the Netherlands; 4. Department of Psychiatry, University Medical Centre Utrecht, Utrecht, 3584CX, the Netherlands; 5. Department of Radiology and Biomedical Imaging, University of California San Francisco, San Francisco, CA, USA; 6. Department of Cognitive Neuroscience, Radboud University Medical Centre, Nijmegen, 6525EN, the Netherlands

Equal Author Contribution:
Andre F. Marquand & James H. Cole contributed equally to this work

This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (CC BY), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Neurology® Published Ahead of Print articles have been peer reviewed and accepted for publication.
This manuscript will be published in its final form after copyediting, page composition, and review of proofs. Errors that could affect the content may be corrected during these processes. Videos, if applicable, will be available when the article is published in its final form.
Abstract

Background and Objectives: Alzheimer’s Disease (AD) is highly heterogeneous, with marked individual differences in clinical presentation and neurobiology. To explore this, we employed neuroanatomical normative modelling to index regional patterns of variability in cortical thickness. We aimed to characterise individual differences and outliers in cortical thickness in patients with AD, people with mild cognitive impairment (MCI) and controls. Furthermore, we assessed the relationships between cortical thickness heterogeneity and cognitive function, amyloid-beta, phosphor-tau, ApoE genotype. Finally, we examined whether cortical thickness heterogeneity was predictive of conversion from MCI to AD.

Methods: Cortical thickness measurements across 148 brain regions were obtained from T1-weighted MRI scans from 62 sites of the Alzheimer’s Disease Neuroimaging Initiative. AD was determined by clinical and neuropsychological examination with no comorbidities present. MCI participants had reported memory complaints, and controls were cognitively normal. A neuroanatomical normative model indexed cortical thickness distributions using a separate healthy reference dataset (n=33,072), employing hierarchical Bayesian regression to predict cortical thickness per region using age and sex, whilst adjusting for site noise. Z-scores per region were calculated, resulting in a z-score ‘brain map’ per participant. Regions with z-scores <=1.96 were classified as outliers.

Results: Patients with AD (n=206) had a median of 12 outlier regions (out of a possible 148), with the highest proportion of outliers (47%) in the parahippocampal gyrus. For 62 regions, over 90% of these patients had cortical thicknesses within the normal range. Patients with AD had more outlier regions than people with MCI (n=662) or controls (n=159) [F(2, 1022) = 95.39, P = 2.0×10^{-16}]. They were also more dissimilar to each other than people with MCI or controls [F(2, 1024) = 209.42, P = 2.2×10^{-16}]. A greater number of outlier regions was
associated with worse cognitive function, CSF protein concentrations and an increased risk of converting from MCI to AD within three years (HR = 1.028, 95% CI[1.016,1.039], P =1.8×10^{-16}).

**Discussion:** Individualised normative maps of cortical thickness highlight the heterogeneous impact of AD on the brain. Regional outlier estimates have the potential to be a marker of disease and could be used to track an individual’s disease progression or treatment response in clinical trials.

**Main Text**

**Introduction**

Alzheimer’s Disease (AD) is the commonest cause of dementia, being characterised by a progressive deterioration in cognitive functioning and independence. The Alzheimer’s disease spectrum comprises substantial clinical and biological differences between patients recognised in clinical and research criteria. These differences include variations in genetic basis, symptom profile, age-at-onset, trajectory and severity, biomarker readouts (e.g., CSF amyloid-beta levels), comorbidities, and in atrophy patterns. Despite this, conventional statistical analyses focus on group averages. This fundamental statistical assumption posits that AD will affect different patients in similar ways, characterising the ‘average’ patient. To reach the goal of precision medicine for Alzheimer’s we need to look beyond the average and design statistical approaches that reflect the patient heterogeneity at the individual level.

Neuroimaging has revealed that differences in brain structure are very common in AD patients. Neuroimaging methods are the gold standard of understanding the *in vivo* brain, specifically, structural imaging has been described as the ‘imaging workhorse of neurodegeneration’, being commonly recommended in Alzheimer’s disease diagnostic guidelines. With this in mind, large structural neuroimaging datasets are increasingly available for dementia, such as Alzheimer’s Disease Neuroimaging Initiative (ADNI), Open Access Series of Imaging Studies (OASIS) and National Alzheimer’s Coordinating Center (NACC) as well as in the general population (e.g., UK Biobank and the Human Connectome...
Project). These datasets provide the ability to chart variation across cohorts and facilitate individual prediction.

Furthermore, large neuroimaging datasets have supported the development and application of data-driven methods in Alzheimer’s disease research. This has revealed that differences in brain structure are very common in patients\textsuperscript{8,13}. Moreover, they have enabled the estimation of disease subtypes from neuroimaging data, as a way to disentangle heterogeneity by grouping patients by distinctive neurobiological and cognitive characteristics\textsuperscript{8,10,13,14} and disease progression\textsuperscript{15}. Such subtypes have the potential to stratify patient groups for clinical decision-making, such as regarding treatment strategy, services and therapies tailored to clinico-radiological phenotype and/or trial enrolment\textsuperscript{16,17}.

Nevertheless, there are challenges associated with the clinical translation of neuroimaging-derived subtypes\textsuperscript{10}. These include the validity of subtypes, how distinct subtypes are from each other and how stable subtypes are over the disease course\textsuperscript{13,18}. Moreover, by design, clustering assumes homogeneity within each cluster, clouding the individual-level variation present, therefore limiting the representation of heterogeneity in the sample\textsuperscript{19}. For instance, individual-level variation is seen in atypical, non-amnestic AD (who comprise up to a third of young-onset AD) which results in challenges to diagnosis and appropriate care\textsuperscript{17}. Arguably, assessing the neurobiology of AD at the individual patient level will provide a precise understanding of their disease, likely outcomes and facilitate tailored treatment strategies. However, whilst this concept of patient-centred, individualised precision medicine for AD is well established, current research efforts are limited.

Neuroanatomical normative modelling is an emerging technique which captures individual-level variability in the brain. This can provide individual statistical inferences with respect to an expected ‘normative’ distribution or trajectory over time. Specifically, by modelling the relationship between neurobiological variables (e.g., neuroimaging features) and covariates (e.g., demographic variables such as age and sex) to map centiles of variation across a cohort (i.e., Z-scores). An individual can then be located within the normative distribution to establish to what extent they deviate from the expected pattern in each measure, and a map can be generated of where and to what extent an individual’s brain differs from the norm\textsuperscript{20,21}.  

Copyright © 2023 The Author(s). Published by Wolters Kluwer Health, Inc. on behalf of the American Academy of Neurology.
This technique has shown to be suitable for precise mapping of individual patterns of variation in brain structure across multiple psychiatric and neurodevelopmental disorders\textsuperscript{20,22–24}. Such findings motivate the first application of neuroanatomical normative modelling to AD\textsuperscript{2}.

Here, we examine individual patterns of variation in brain structure in patients with AD using neuroanatomical normative modelling. Using the well-characterised, muti-site, Alzheimer’s Disease Neuroimaging Initiative (ADNI) dataset, we applied a recent implementation of the normative modelling framework, hierarchical Bayesian regression. This technique has been shown to efficiently accommodate inter-site variation and provides computational scaling, which is useful when using large studies, or combining smaller studies together, that are acquired across multiple sites in a federated learning framework\textsuperscript{25–27}. Our main objective was to quantify spatial patterns of neuroanatomical heterogeneity using cortical thickness measures in patients with AD, people with mild cognitive impairment (MCI) and cognitively normal controls, by calculating deviations from normative ranges for each brain region and then identifying statistical outliers. Specifically, we aimed to 1) assess the extent of neuroanatomical variability between individual patients based on overlapping or distinct patterns of outliers, 2) quantify group differences in between-participant dissimilarity, 3) relate the quantity of neuroanatomical outliers to cognitive performance and AD biomarkers, and 4) examine whether the number of outliers related to subsequent disease progression from MCI to AD.

**Materials and methods**

**Participants**

Participants were derived from two datasets (1) a reference dataset comprised of healthy people across the human lifespan (2) a clinical target dataset which included people with AD or MCI in addition to age-matched cognitively normal controls. The reference dataset was made by combining data on healthy people from multiple publicly available sources\textsuperscript{27}, including Open Access Series of Imaging Studies (OASIS), Adolescent Brain Cognitive Development (ABCD) study and UK Biobank (UKB), detailed in eTable 1 in the Supplement. The clinical data used in the preparation of this article was obtained from the
ADNI database; http://adni.loni.usc.edu). Inclusion criteria were the availability of a baseline T1-weighted MRI, AD participants had to meet the National Institute of Neurological and Communicative Disorders and Stroke–AD and Related Disorders Association criteria for probable AD and were screened to exclude genetic risk for familial Alzheimer’s. MCI participants reported a subjective memory concern either autonomously or via an informant or clinician, participants had no significant levels of impairment in other cognitive domains.

**Standard Protocol Approvals, Registrations, and Patient Consents**

Written informed consent was obtained from all participants before experimental procedures were performed. Approval was received by an ethical standards committee for ADNI study data use.

**MRI acquisition**

For the clinical dataset, T1-weighted images were acquired at multiple sites using 3T MRI scanners. Detailed MRI protocols for T1-weighted sequences are available online (http://adni.loni.usc.edu/methods/documents/mri-protocols/). The quality of raw scans was evaluated by UCSF prior to our exclusion criteria. Scans were excluded based on technical problems and significant motion artefacts and clinical abnormalities (https://adni.bitbucket.io/reference/docs/UCSFFSX51/UCSF%20FreeSurfer%20Methods%20and%20QC_OFFICIAL.pdf).

**Estimation of cortical thickness**

T1-weighted scans from both the reference and ADNI datasets were processed using a mix of both FreeSurfer versions 5 and 6. Cortical thickness values were generated using the *recon-all* cross-sectional approach. This cortical thickness algorithm calculates the mean distance between vertices of a corrected, triangulated estimated grey/white matter surface and grey matter/CSF (pial) surface, which generated the cortical thickness of each region of the Destrieux atlas regions. This included the mean cortical thickness and 148 regions cortical thickness values for each participant.
Quality control of FreeSurfer processing for the reference dataset relied on automated filtering median-centred absolute Euler number higher than 25, as employed in prior work\textsuperscript{26,27}. The exclusion of outliers based on Euler numbers has shown to be a reliable quality control strategy in large neuroimaging cohorts\textsuperscript{31,32}. For ADNI, quality control was based on a visual review of each cortical region performed by UCSF. Only scans which passed this quality control were used.

**Neuroanatomical Normative modelling**

A hierarchical Bayesian regression model was trained on multi-site data to generate normative models per region using the covariates age and sex. This was based on the population variation in the reference dataset (training data), as per Kia and colleagues which adaptively pools parameter estimates across sites via a shared prior over regression parameters across sites\textsuperscript{27}. This simultaneously accounts for inter-site variation and allows sites to borrow strength from one another in a fully Bayesian framework. The advantage of training the models on the large independent dataset, compared to just using ADNI, is that ADNI consists of many sites with small sample sizes. This would result in unstable estimates of normative distributions that could be strongly influenced by outliers or sampling bias. Here, by training on over $n=33,000$ from only nine datasets (with 60 sites), the model produces a stable distribution of estimates across the entire lifespan. Next, these estimates were conditioned to our specific context, using an adapted transfer learning approach\textsuperscript{27}. The parameters of the reference normative model were recalibrated to the ADNI dataset using 70\% of healthy controls per ADNI site, where 70\% was used to give stable estimates of the transferred model parameters, given that many of the scan sites in ADNI have quite small sample sizes. The remaining 30\% of healthy controls plus MCI and patients with AD were used to assess the heterogeneity in neuroanatomical presentation. This process generated regional and mean cortical thickness z-scores for each participant in the clinical dataset, relative to the normative range of the reference dataset. All modelling steps are performed using PCNToolkit (v0.20) (https://github.com/amarquand/PCNtoolkit).
Statistical analysis

Group cortical thickness differences
Cortical thickness group comparisons were conducted using t-tests at each region and corrected for multiple comparisons using the False Discovery Rate (FDR). Significant p-values were mapped onto the Destrieux atlas using the R package ggseg.

Outlier definition and statistics
Outliers in terms of low cortical thickness were identified for each region, defined as \( Z < -1.96 \) (corresponding to the bottom 2.5% of the normative distribution of cortical thickness). We only used the lower bound threshold for outliers as we were interested in cortical thinning associated with neurodegeneration. The number of outliers were summed across 148 regions for each participant, to give a total outlier count (tOC) across regions. Linear regression tested for group differences in mean cortical thickness Z-score and tOC. Additionally, group comparisons at each region were conducted using Chi\(^2\) (FDR corrected). Hamming distance, a quantitative measure of similarity between binary thresholded cortical thickness outlier vectors was used to measure dissimilarity between individuals. Median Hamming distances were compared between groups. To explore spatial patterns of cortical thickness outliers per group, the proportion of participants within each group whose cortical thickness was an outlier (i.e., \( Z < -1.96 \)) was calculated for each region. This enabled visualisation of the extent to which patterns of outlier regions overlap or are distinct. This was mapped using the Destrieux atlas via the R package ggseg. All statistical analyses were implemented in R version 3.6.2.

Outlier associations with cognitive function and CSF markers
Linear regression adjusting for age, sex, years of education and Clinical Dementia Rating (sum of boxes) examined the relationship between tOC and cognitive composite scores (memory using ADNI MEM or executive function using ADNI EF). We assessed the interactional effects of diagnostic group within a subsequent regression. Furthermore, linear regression adjusting for age and sex only examined the relationship between tOC and CSF markers (amyloid-beta and phospho-tau). We also assessed the interactional effects of diagnostic group within a subsequent regression. To stratify outlier maps in both MCI and...
patients with AD groups, we used total scores from the Mini-Mental State Examination (MMSE).

**MCI to AD conversion analysis**

Follow-up diagnosis status data, up to three years from baseline scan, were obtained from 454 people with MCI. In total, 76 people with MCI at baseline had converted to AD within three years. We then ran a survival analysis using Cox proportional hazards regression to assess whether tOC related to the risk of converting from MCI to AD, controlling for age and sex. We use a Kaplan–Meier plot to illustrate how either a low or high tOC (split via median) can contribute to the risk of converting.

**Data availability Statement**

Statistical analysis scripts are available on GitHub (https://github.com/serenaverdi/ADNI_normative-modelling). The neuroanatomical normative model was generated using the PCNtoolkit software package (https://github.com/amarquand/PCNtoolkit). ADNI data used in this study are publicly available and can be requested following ADNI Data Sharing and Publications Committee guidelines: http://adni.loni.usc.edu/data-samples/access-data/

**Results**

**Participants**

In the reference dataset, a total of n=33,072 T1-weighted MRI scans were collated across 60 sites (This sample is described in detail in Kia et al 2021 and summarised in eTable 1 in the Supplement)\(^{27}\). The clinical ADNI dataset amounted to total 1,492 participants which were acquired across 62 sites (Table 1). Here 70% of controls were removed from the clinical dataset and were used as a calibration dataset to adapt the normative model to the new sites, these were randomly selected and stratified across sites and gender to make sure all sites and genders are present in the adaptation set. This left a total of 1,027 participants in the final clinical dataset.
Patients with AD have smaller cortical thicknesses than people with MCI or with normal cognition

Mean cortical thicknesses were compared across participant groups. Age- and sex-adjusted mean cortical thickness significantly differed between groups overall \( F(2,1487) = 137.9, \, P = 2.0 \times 10^{-16} \). Pairwise comparisons (Tukey post-hoc) were all significant [\( P<0.001 \)], with mean cortical thickness being lowest in AD (mean = 2.28 SD = 0.13, CI [0.161, -0.124]) and highest in controls (mean = 2.42 SD = 0.11, CI[2.415, 2.433]), with MCI being intermediate (mean = 2.38, SD = 0.12, CI[-0.054, -0.029]) (eFigure 1 in the Supplement).

Region-level pairwise group comparisons (total of 148 regions) provided evidence cortical thickness measures were on average lower in 133 regions in AD versus controls, in 111 regions in AD versus MCI and in 78 regions in MCI versus controls, after FDR correction (eFigure 1 in the Supplement).

Next, cortical thickness z-scores, derived from comparison to the normative model, were then compared across participant groups. In this way, we could compare the degree to which each group differed from the separate reference cohort, used to define the normative model. Consistent with comparisons of mean cortical thickness, age- and sex-adjusted z-scores differed between groups overall \( F(2, 1022) = 69.49), \, P = 2.0 \times 10^{-16} \). Pairwise comparisons (Tukey post-hoc) were all significant \( P<0.003 \), with Z-scores being lowest in AD (mean = -1.27, SD = 1.41, CI[-1.630, -1.130]), highest in controls (mean = 0.07, SD = 1.04, CI[-1.053, 0.374]), and intermediate in MCI (mean =-0.28, SD =1.17, CI[-0.600, -0.180]) (eFigure 2A in the Supplement).

Furthermore, age- and sex-adjusted tOCs differed between groups overall \( F(2, 1022) = 95.39), \, P = 2.0 \times 10^{-16} \). Pairwise comparisons (Tukey post-hoc) were all significant \( P<0.003 \), with tOCs being highest in AD (median = 12, IQR = 28, CI[14.38, 19.88]), lowest in controls (median = 2, IQR = 6, CI[2.780, 18.494]), and intermediate in MCI (median = 4, IQR = 9, CI[1.56, 6.18]) (eFigure 2B in the Supplement).

Region-level pairwise group comparisons (total of 148 regions) showed higher numbers of outliers in cortical thickness in 79 regions in AD versus controls, in 63 regions in AD versus
MCI and 1 region in MCI versus controls, after FDR correction. Region-level group differences in outlier count were most evident within temporoparietal and to a lesser extent frontal and occipital regions (Figure 1A).

**Patients with AD are less similar to each other than people with MCI or with normal cognition**

Hamming distance matrices indicated greater within-group dissimilarity in patients with AD, relative to MCI or control participants, who were most similar to each other in spatial patterns of outliers (Figure 2).

Median hamming distance significantly differed between groups overall, \([F(2, 1024) = 209.42, P = 2.2 \times 10^{-16}]\). Pairwise comparisons (Tukey post-hoc) were all significant \([P<0.001]\), with being highest in AD (median = 32, IQR = 32, CI[26.29, 29.43]), and lowest in controls (median = 6, IQR = 8, CI[-24.37, -19.61]), with MCI being intermediate (median =10, IQR = 14, CI[-18.52, -14.92]).

**Patients with AD have spatially higher proportions of cortical thickness outliers**

The proportion of outliers defined within each group differed in regional patterns between Alzheimer's, MCI and control groups. This is illustrated in Figure 1B and in eFigure 3 in the Supplement. For a breakdown of proportions see eTable 2, for individual maps of outliers see Video 1. A greater number of regions and a higher proportion of the group were outliers in patients with AD, as expected. In fact, 145 regions in the AD group had over the expected 2.5% of patients with an outlier (based on the Z < -1.96 threshold). The left parahippocampal gyrus was the region with the highest outlier percentage (47% of the AD group). For the MCI group, 138 regions in the MCI group had outliers (over the expected 2.5% of the group). The left parahippocampal gyrus is the region with the highest outlier percentage (14% of the MCI group). For the control group, only 66 regions had outliers above the expected 2.5%. The left occipital temporal lateral sulcus was the region with the highest outlier percentage (6% of controls).
Outliers are associated with cognitive function and CSF amyloid-beta and phosphorylated tau

tOC across the whole sample was significantly associated with memory performance \( [\beta = -0.01, P = 2.2 \times 10^{-16}] \) and executive function \( [\beta = -0.02, P = 2.2 \times 10^{-16}] \) in a linear regression model. To check for the association between two variables within a sample, we also model a group by tOC interaction term, which was not significant for memory performance \( [F(2, 849) = 2.28, P = 0.103] \) and executive function \( [F(2, 849) = 2.534, P = 0.07] \) (Figure 3A, 3B). Lower MMSE scores showed different spatial patterns of outliers in both MCI and patients with AD (Figure 4A) groups. However, total MMSE score and age did explain some of the variance in tOC \( [\text{adjusted } R^2 = 0.1793, P = 2.2 \times 10^{-16}] \).”

In addition, tOC was significantly associated with amyloid-beta \( [\beta = 0.002, P = 0.022] \) and phospho-tau \( [\beta = 0.1301, P = 1.04 \times 10^{-8}] \), which was not influenced by either group amyloid-beta \( [F(2, 576) = 0.96, P = 0.38] \) or phospho-tau interaction \( [F(2, 576) = 1.362, P = 0.257] \) (Figure 3C, 3D).

Case studies suggest that variability in cortical thickness is not solely due to disease stage or other clinical factors

To explore whether individual differences in outlier maps were driven by disease-related characteristics (such as ApoE genotype and demographics) or by disease stage, we examined sets of participants closely matched for ApoE genotype status, age, sex and MMSE score. Figure 4B presents four individual female patients with AD all aged 71-72 years, heterozygous for ApoE ε4, with similar MMSE scores, all of whom were CSF amyloid-positive, with no underlying comorbidities. Furthermore, clinical impressions confirm these individuals all have mild dementia, with further confirmation of no depressive symptoms. These individual patients might be considered similar from biological or clinical perspectives, yet their patterns of outliers in cortical thickness are markedly variable; for example, variously suggesting lateralised (Patient 3) and occipital atrophy (Patient 1).”
Greater numbers of outliers are associated with risk of conversion from mild cognitive impairment to AD

A survival analysis indicated that for every 10 points of tOC, risk of converting from MCI to AD within 3 years increased by 31.4% (HR = 1.028, 95% CI [1.016,1.039], $P = 1.8 \times 10^{-16}$) (Figure 5A). This is illustrated within a Kaplan–Meier plot, which shows how a high tOC can contribute to the risk of converting in comparison to a low tOC (Figure 5B).

Discussion

In this study, we defined individual spatial patterns of cortical thickness outliers and illustrated that AD does not affect different people in a uniform way. Moreover, our analysis quantified and visualised these individual differences in patterns of cortical atrophy. Overall, the results of the present study provide evidence of 1) heterogeneous patterns of cortical thickness between patients with AD, 2) associations of cortical thickness heterogeneity with cognitive performance and CSF amyloid-beta and phospho-tau, and 3) the potential of individualised markers of cortical thickness heterogeneity to predict survival time before conversion from the MCI stage to diagnosed AD.

Our findings both complement and offer additional information to the established understanding of AD. We observed a high tOC in patients with AD, consistent with the evidence of cortical thinning as a consequence of AD neuropathology. Moreover, we also observe significant associations with cortical thinning and poor cognitive performance, a decrease in CSF amyloid-beta and an increase in CSF phospho-tau (Figure 3), which is also consistent with previous findings. Atrophy has also been associated with the risk of progression from MCI to AD (Figure 5), alongside a combination of other biomarkers. Importantly, these previous studies examined the correlates of common patterns of cortical atrophy – yet conversely, we considered individual variability in patterns of cortical thickness, as opposed to assessing group average relationships. This highlights that individualised measures of neuroanatomy are sensitive to both non-imaging disease markers and disease progression.
The tOC has the potential to be used as an individual patient metric of poor brain health to help inform clinical decisions. Indeed, similar measures have recently been adopted as a clinical measure, i.e., brain volume/thickness patient z-scores. However, these have been calculated using different normative modelling techniques\textsuperscript{40,41}, which base their normative population on smaller reference samples; limit modelling to just whole brain, or within specific regions; do not account for site related variation (i.e., site effects) and did not fully relate these to clinical outcomes and cognitive scores. Our tOC can provide an optimised measure here, and will translate within clinical applications for precision medicine.

When assessing regional heterogeneity of the ADNI sample, we observed more outliers in patients with AD in temporal regions such as the hippocampus and the cingulate cortex. These are areas known to be sensitive to neurodegeneration in AD\textsuperscript{42}, and are responsible for clinical symptoms in AD\textsuperscript{43}. However, looking beyond these group-average regional differences, we observe that the highest proportion of outliers in a single region was less than 50\% in the AD group (Figure 1). This suggests the individual spatial patterns of outliers in Alzheimer’s only partially overlap between patients; if atrophy were homogenous (as assumed within group averages), we might expect 100\% of participants to have outliers here.

The observed variation in atrophy in the temporal lobe is consistent with subtyping studies\textsuperscript{8,44}. Also, a recent study used normative modelling to estimate neuroanatomical heterogeneity within the ADNI cohort, which show similarities of variation in atrophy within the temporal regions\textsuperscript{45}. However, in comparison to these studies, our specific application of neuroanatomical normative modelling has enabled the creation of an individual metric of neuroanatomical heterogeneity, characterised the spatially distributed nature of alterations in MCI and AD, and assessed how neuroanatomical variability related to cognitive performance, CSF biomarkers, disease progression, or genetic factors. Furthermore, this present study is the first to account for the crucial confounding effects of multiple scanning sites when evaluating neuroanatomical heterogeneity in AD.

Going further, our study reveals that each patient not only differs in the number of outliers they have, but the regional patterns of outliers markedly differ (Video. 1). The latter is reflected in large levels of dissimilarity between AD individuals (Figure 2). Potentially, one
reason for the variable patterns of atrophy is simply disease stage, whereby more atrophy appears with greater disease progression. However, our results indicate that this is not the case, as when closely examining patients of very similar demographics and clinical characteristics, being at a comparable disease stage (e.g., based on MMSE score), heterogenous patterns of cortical atrophy were still present (Figure 4).

It is surprising to observe that cognitively normal controls also showed some outliers, suggesting a degree of within-group heterogeneity (Figure 1; Figure 2). Therefore, the assumption of homogeneity in case-control studies should be made with caution, even in control groups. Statistical designs for basic research and for clinical trials should better reflect this heterogeneity in brain structure.

A few considerations can be made to the datasets used within the study. Although the reference dataset includes over 30,000 individuals, we should be cautious to assume that it is representative of a healthy population. Also, patients who volunteer for research studies (i.e., ADNI) do not necessarily reflect the clinical population. Future neuroanatomical normative modelling studies could supplement the reference dataset with MRI scans acquired from routine clinical visits, community cohorts, or other less selective sources. Finally, the reference data were processed with a variety of FreeSurfer versions. While impractical to unify the processing retrospectively, these inconsistencies may potentially add noise to the normative models, so represents an important caveat.

As ADNI is comprised of more early-stage dementia participants, examining late-stage patients with AD may offer insights into the heterogeneity in spatial patterns of atrophy across the disease course. Clinical observations have suggested that late-stage patients with AD have widespread atrophy across the brain; therefore, we may hypothesise such patients will have less heterogeneous patterns of atrophy. However, regardless of the heterogeneous patterns of atrophy, the tOC can still provide information about the extent of cortical atrophy in a given individual.
Another limitation of ADNI dataset is the underrepresentation of cognitive domains beyond memory, executive function and language. Between a quarter to a third of the AD group exhibit parieto-occipital outliers, comparable to separate parieto-occipital predominant subtypes associated with prominent visuospatial dysfunction. Further characterisation of how outlier distribution relates to non-memory/executive symptoms may be of particular clinical relevance, for example, given the implications of visuospatial dysfunction for diminished autonomy, falls risk and appropriate services.

Future efforts when applying neuroanatomical normative modelling to AD data should incorporate serial neuroimaging across multiple time points. This will help define patient-level longitudinal trajectories. Mapping neuroanatomical variability using neuroanatomical normative modelling at different time points has the potential to improve predictions of disease progression or treatment response, at the level of the individual patient. Apart from our MCI to AD analysis, the sample taken from this present study is cross-sectional, reflecting a snapshot in time, yet heterogeneity has been shown to differ temporally. Potentially, data-driven staging methods here (e.g., SusStain) may also provide clinically useful information of longitudinal trends of individual heterogeneity whilst taking account of an individual’s disease stage.

Furthermore, it will also be valuable to map variation using other neuroanatomical metrics, such as subcortical volumes. Our methodology can be extended to include subcortical volumes by using a reference dataset which has such data available. Future efforts that adopt this could enrich our understanding of regional anatomical heterogeneity between patients.

**Conclusion**

We provide a quantitative approach to estimate this variability at the individual patient level based on spatial patterns of an individual's cortical thickness normative deviations. Individualised maps of neuroanatomical outliers were related to cognitive performance and CSF biomarkers. Furthermore, the number of outliers, based on individual patterns, helped predict conversion from MCI to AD. These individual neuroanatomical maps, derived from normative models, have the potential to be a marker of AD state. These could index disease...
progression or even evaluate the effectiveness of potential disease-modifying treatments, tailored to the individual patient.

References

2. Verdi S, Marquand AF, Schott JM, Cole JH. Beyond the average patient: how neuroimaging models can address heterogeneity in dementia. Brain. Published online April 23, 2021. doi:10.1093/brain/awab165
12. Dementia. Dementia: Assessment, management and support for people living with dementia and their carers. Published online 2018.


42. 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. *Hum Brain Mapp*. Published online 2020. doi:10.1002/hbm.25115


### Tables

<table>
<thead>
<tr>
<th></th>
<th>Controls</th>
<th>MCI</th>
<th>AD</th>
<th>Total</th>
<th>Statistical differences</th>
</tr>
</thead>
<tbody>
<tr>
<td>n</td>
<td>621</td>
<td>664</td>
<td>207</td>
<td>1492</td>
<td></td>
</tr>
<tr>
<td>X² = 5.001, P = 0.0241</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean ± sd Age (years)</td>
<td>72.2 ± 6.8</td>
<td>71.9 ± 7.7</td>
<td>74.0 ± 8.0</td>
<td>72.3 ± 7.4</td>
<td></td>
</tr>
<tr>
<td>F(2,1489) = 6.72, P = 0.001</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age range (years)</td>
<td>62.8 – 88.7</td>
<td>56.1 – 92.5</td>
<td>57.0 – 88.5</td>
<td>56.1-92.5</td>
<td></td>
</tr>
<tr>
<td>Mean ± sd total MMSE score</td>
<td>29.06 ± 1.18</td>
<td>27.82 ± 1.91</td>
<td>22.56 ± 3.19</td>
<td>26.94 ± 3.11</td>
<td>F(2,1400) = 867.97, P &lt; 2.2x10⁻¹⁶</td>
</tr>
<tr>
<td>Mean ± sd CSF amyloid-beta (pg/mL)</td>
<td>246.55 ± 305.02</td>
<td>231.98 ± 275.41</td>
<td>178.07 ± 250.44</td>
<td>223.00 ± 263.87</td>
<td>F(2, 765) = 3.39, P = 0.03</td>
</tr>
<tr>
<td>Mean ± sd CSF p-tau (pg/mL)</td>
<td>305.02 ± 275.41</td>
<td>312.40 ± 246.72</td>
<td>182.37 ± 267.32</td>
<td>263.87 ± 267.32</td>
<td>F(2, 765) = 37.11, P &lt; 4.1x10⁻¹⁶</td>
</tr>
<tr>
<td>ApoE ε4 negative (%) = proportion in group sample</td>
<td>385 (69%)</td>
<td>322 (53%)</td>
<td>58 (31%)</td>
<td>765 (56%)</td>
<td>X² = 85.92, P &lt; 2.2x10⁻¹⁶</td>
</tr>
</tbody>
</table>
Table 1.

**Figure/table legends**

**Table 1. Demographics of ADNI sample.** Statistical differences were assessed using ANOVA and chi-squared tests. *MMSE had maximum score of 30.

**Figure 1: Regional maps of heterogeneity**

**(A):** Mapped are significant group differences of outliers. Colour bar indicates effect size as Phi $\phi$ (0.1 is considered to be a small effect, 0.3 a medium effect, and 0.5 a large effect). **(B)** Mapped are the percentage of outliers present within each participant group. The colour bar reflects outlier proportion from 2.5% to 100% (thresholding of z-scores). Zero percent (grey) represents that no participants have outliers in those respective regions.
Figure 2: Outlier dissimilarity

(A) Outlier distance heatmaps: both x and y axes represent participants within each group. Yellow indicates higher hamming distance (greater dissimilarity between participants in this brain region), as opposed to if participants are identical in this brain region, the Hamming distance would be 0, represented by white in the colour bar. (B) Outlier distance density: illustrates the spread of outlier dissimilarity (calculated by Hamming distance) within each group.
Figure 3: Cognitive function and CSF markers association with tOC.

Fitted lines are from a linear regression model per diagnostics group for (A) Memory function, (B) Executive function, (C) CSF amyloid-beta, and (D) phospho-tau.
Figure 4: Regional maps of outliers, stratified by MMSE score.
Mapped are the percentage of region outliers proportional to MMSE scoring subgroup in (A) MCI participants and (B) patients with AD. The colour bar reflects outlier proportion from 2.5% to 100% (thresholding of z-scores). Zero percent (grey) represents that no participants have outliers in those respective regions.”
Figure 5: Conversion from mild cognitive impairment to AD

(A) Kaplan–Meier plot of MCI to AD conversion: the two lines represent a median split of tOC, with <4 classed as low tOC (blue), and ≥4 classed as high tOC (red). Crosses indicate censoring points (i.e., age at last diagnosis assessment). Filled colour represent the 95% confidence intervals. (B) Mapped are the proportion of regional outliers among people with MCI who converted to patients with AD.

Video 1: Individual outlier maps of each study participant

Personalised regional maps of outliers looping through each participant of this study.
Revealing Individual Neuroanatomical Heterogeneity in Alzheimer Disease Using Neuroanatomical Normative Modeling
Serena Verdi, Seyed Mostafa Kia, Keir X. X Yong, et al.

*Neurology* published online May 1, 2023
DOI 10.1212/WNL.0000000000207298

This information is current as of May 1, 2023

Updated Information & Services
including high resolution figures, can be found at:
http://n.neurology.org/content/early/2023/04/29/WNL.0000000000207298.full

Citations
This article has been cited by 1 HighWire-hosted articles:
http://n.neurology.org/content/early/2023/04/29/WNL.0000000000207298.full#otherarticles

Subspecialty Collections
This article, along with others on similar topics, appears in the following collection(s):
Alzheimer disease
http://n.neurology.org/cgi/collection/alzheimers_disease
MCI (mild cognitive impairment)
http://n.neurology.org/cgi/collection/mci_mild_cognitive_impairment
MRI
http://n.neurology.org/cgi/collection/mri

Permissions & Licensing
Information about reproducing this article in parts (figures, tables) or in its entirety can be found online at:
http://www.neurology.org/about/about_the_journal#permissions

Reprints
Information about ordering reprints can be found online:
http://n.neurology.org/subscribers/advertise

*Neurology* © is the official journal of the American Academy of Neurology. Published continuously since 1951, it is now a weekly with 48 issues per year. Copyright © 2023 The Author(s). Published by Wolters Kluwer Health, Inc. on behalf of the American Academy of Neurology. All rights reserved. Print ISSN: 0028-3878. Online ISSN: 1526-632X.