https://doi.org/10.1093/brain/awae118 BRAIN 2024: 147; 2400–2413 | 2400

BRAIN
ORIGINAL ARTICLE

A generalizable data-driven model of atrophy heterogeneity and progression in memory clinic settings

Hannah Baumeister,[1] Jacob W. Vogel,[2] Philip S. Insel,[3] Luca Kleineidam,[4,5] Steffen Wolfsgruber,[4,5] Melina Stark,[5] Helena M. Gellersen,[1] Renat Yakupov,[1,6] Matthias C. Schmid,[4,7] Falk Lüsebrink,[1] Frederic Brosseron,[4] Gabriel Ziegler,[6] Silka D. Freiesleben,[8,9] Lukas Preis,[9] Luisa-Sophie Schneider,[9] Eike J. Spruth,[8,9] Slawek Altenstein,[8,9] Andrea Lohse,[9] Klaus Fliessbach,[4,5] Ina R. Vogt,[4] Claudia Bartels,[10] Björn H. Schott,[10,11,12] Ayda Rostamzadeh,[13] Wenzel Glanz,[1] Enise I. Incesoy,[1,6,14] Michaela Butryn,[1] Daniel Janowitz,[15] Boris-Stephan Rauchmann,[16,17,18] Ingo Kilimann,[19,20] Doreen Goerss,[19,20] Matthias H. Munk,[21,22] Stefan Hetzer,[23] Peter Dechent,[24] Michael Ewers,[15,25] Klaus Scheffler,[26] Anika Wuestefeld,[2] Olof Strandberg,[2] Danielle van Westen,[27,28] Niklas Mattsson-Carlgren,[2,29,30] Shorena Janelidze,[2] Erik Stomrud,[2,31] Sebastian Palmqvist,[2,31] Annika Spottke,[4,32] Christoph Laske,[21,22,33] Stefan Teipel,[19,20] Robert Perneczky,[16,25,34,35] Katharina Buerger,[15,25] Anja Schneider,[4,5] Josef Priller,[8,9,36,37] Oliver Peters,[8,9] Alfredo Ramirez,[4,5,38,39,40] Jens Wiltfang,[10,11,41] Michael T. Heneka,[42] Michael Wagner,[4,5] Emrah Düzel,[1,6,43] Frank Jessen,[4,13,38] Oskar Hansson[2,31] and David Berron[1,2,43]

Memory clinic patients are a heterogeneous population representing various aetiologies of pathological ageing. It is not known whether divergent spatiotemporal progression patterns of brain atrophy, as previously described in Alzheimer's disease patients, are prevalent and clinically meaningful in this group of older adults.

To uncover distinct atrophy subtypes, we applied the Subtype and Stage Inference (SuStaIn) algorithm to baseline structural MRI data from 813 participants enrolled in the DELCODE cohort (mean ± standard deviation, age = 70.67 ± 6.07 years, 52% females). Participants were cognitively unimpaired ($n = 285$) or fulfilled diagnostic criteria for subjective cognitive decline ($n = 342$), mild cognitive impairment ($n = 118$) or dementia of the Alzheimer's type ($n = 68$). Atrophy subtypes were compared in baseline demographics, fluid Alzheimer's disease biomarker levels, the Preclinical Alzheimer Cognitive Composite (PACC-5) as well as episodic memory and executive functioning. PACC-5 trajectories over up to 240 weeks were examined. To test whether baseline atrophy subtype and stage predicted clinical trajectories before manifest cognitive impairment, we analysed PACC-5 trajectories and mild cognitive impairment conversion rates of cognitively unimpaired participants and those with subjective cognitive decline.

Limbic-predominant and hippocampal-sparing atrophy subtypes were identified. Limbic-predominant atrophy initially affected the medial temporal lobes, followed by further temporal regions and, finally, the remaining cortical regions. At baseline, this subtype was related to older age, more pathological Alzheimer's disease biomarker levels, *APOE* ε4 carriership and an amnestic cognitive impairment. Hippocampal-sparing atrophy initially occurred outside

Received September 07, 2023. Revised February 02, 2024. Accepted March 03, 2024. Advance access publication April 24, 2024

the temporal lobe, with the medial temporal lobe spared up to advanced atrophy stages. This atrophy pattern also affected individuals with positive Alzheimer's disease biomarkers and was associated with more generalized cognitive impairment. Limbic-predominant atrophy, in all participants and in only unimpaired participants, was linked to more negative longitudinal PACC-5 slopes than observed in participants without or with hippocampal-sparing atrophy and increased the risk of mild cognitive impairment conversion.
SuStaIn modelling was repeated in a sample from the Swedish BioFINDER-2 cohort. Highly similar atrophy progression patterns and associated cognitive profiles were identified. Cross-cohort model generalizability, at both the subject and the group level, was excellent, indicating reliable performance in previously unseen data.
The proposed model is a promising tool for capturing heterogeneity among older adults at early at-risk states for Alzheimer's disease in applied settings. The implementation of atrophy subtype- and stage-specific end points might increase the statistical power of pharmacological trials targeting early Alzheimer's disease.

1 German Center for Neurodegenerative Diseases (DZNE), 39120 Magdeburg, Germany
2 Clinical Memory Research Unit, Department of Clinical Sciences Malmö, Lund University, 222 42 Lund, Sweden
3 Department of Psychiatry and Behavioral Sciences, University of California, San Francisco, San Francisco, CA 94143, USA
4 German Center for Neurodegenerative Diseases (DZNE), 53127 Bonn, Germany
5 Department of Neurodegenerative Disease and Geriatric Psychiatry, University of Bonn Medical Center, 53127 Bonn, Germany
6 Institute of Cognitive Neurology and Dementia Research (IKND), Otto-von-Guericke University, 39120 Magdeburg, Germany
7 Institute for Medical Biometry, University Hospital Bonn, 53127 Bonn, Germany
8 German Center for Neurodegenerative Diseases (DZNE), 10117 Berlin, Germany
9 Department of Psychiatry and Neurosciences, Charité—Universitätsmedizin Berlin, corporate member of Freie Universität Berlin and Humboldt-Universität zu Berlin, 10117 Berlin, Germany
10 Department of Psychiatry and Psychotherapy, University Medical Center Göttingen, 37075 Göttingen, Germany
11 German Center for Neurodegenerative Diseases (DZNE), 37075 Göttingen, Germany
12 Leibniz Institute for Neurobiology, 39118 Magdeburg, Germany
13 Department of Psychiatry, Medical Faculty, University of Cologne, 50937 Cologne, Germany
14 Department of Psychiatry and Psychotherapy, Otto-von-Guericke University, 39120 Magdeburg, Germany
15 Institute for Stroke and Dementia Research (ISD), Ludwig-Maximilians-Universität, 81377 Munich, Germany
16 Department of Psychiatry and Psychotherapy, Ludwig-Maximilians-Universität, 80336 Munich, Germany
17 Sheffield Institute for Translational Neuroscience (SITraN), University of Sheffield, Sheffield S10 2HQ, UK
18 Department of Neuroradiology, Ludwig-Maximilians-Universität, 81377 Munich, Germany
19 German Center for Neurodegenerative Diseases (DZNE), 18147 Rostock, Germany
20 Department of Psychosomatic Medicine, Rostock University Medical Center, 18147 Rostock, Germany
21 German Center for Neurodegenerative Diseases (DZNE), 72076 Tübingen, Germany
22 Department of Psychiatry and Psychotherapy, University of Tübingen, 72076 Tübingen, Germany
23 Berlin Center for Advanced Neuroimaging, Charité—Universitätsmedizin Berlin, 10117 Berlin, Germany
24 MR-Research in Neurosciences, Department of Cognitive Neurology, Georg-August-University Göttingen, 37075 Göttingen, Germany
25 German Center for Neurodegenerative Diseases (DZNE), 81377 Munich, Germany
26 Department for Biomedical Magnetic Resonance, University of Tübingen, 72076 Tübingen, Germany
27 Diagnostic Radiology, Institution of Clinical Sciences Lund, Lund University, 211 84 Lund, Sweden
28 Image and Function, Skåne University Hospital, 211 84 Lund, Sweden
29 Department of Neurology, Skåne University Hospital, Lund University, 211 84 Lund, Sweden
30 Wallenberg Center for Molecular Medicine, Lund University, 22184 Lund, Sweden
31 Memory Clinic, Skåne University Hospital, 205 02 Malmö, Sweden
32 Department of Neurology, University of Bonn, 53127 Bonn, Germany
33 Section for Dementia Research, Hertie Institute for Clinical Brain Research, 72076 Tübingen, Germany
34 Munich Cluster for Systems Neurology (SyNergy), 81377 Munich, Germany
35 Ageing Epidemiology Research Unit (AGE), School of Public Health, Imperial College London, London SW7 2AZ, UK
36 Department of Psychiatry and Psychotherapy, Technical University of Munich, 81675 Munich, Germany
37 Centre for Clinical Brain Sciences, University of Edinburgh and UK DRI, Edinburgh EH16 4SB, UK
38 Excellence Cluster on Cellular Stress Responses in Aging-Associated Diseases (CECAD), University of Cologne, 50931 Cologne, Germany
39 Division of Neurogenetics and Molecular Psychiatry, Department of Psychiatry and Psychotherapy, University of Cologne, 50931 Cologne, Germany

40 Department of Psychiatry & Glenn Biggs Institute for Alzheimer's and Neurodegenerative Diseases, The University of Texas at San Antonio, San Antonio, TX 78229, USA
41 Institute of Biomedicine (iBiMED), Department of Medical Sciences, University of Aveiro, 3810-193 Aveiro, Portugal
42 Luxembourg Centre for Systems Biomedicine (LCSB), University of Luxembourg, 4362, Belvaux, Luxembourg
43 Center for Behavioral Brain Sciences (CBBS), Otto-von-Guericke University Magdeburg, 39106 Magdeburg, Germany

Correspondence to: Hannah Baumeister
Deutsches Zentrum für Neurodegenerative Erkrankungen e.V. (DZNE)
Leipziger Strasse 44
39120 Magdeburg, Germany
E-mail: hannah.baumeister@dzne.de

Correspondence may also be addressed to: David Berron
E-mail: david.berron@dzne.de

Keywords: disease heterogeneity; Alzheimer's disease; structural MRI; executive function; episodic memory

Introduction

Sporadic Alzheimer's disease (AD) is characterized by two hallmark proteinopathies—accumulations of amyloid-β (Aβ) and hyperphosphorylated tau—in addition to progressive neurodegeneration.[1] Traditionally, these pathological entities have been assumed to progress along a stereotypical trajectory in which brain regions of the medial temporal lobe (MTL) are among the earliest affected by tau accumulation and related atrophy.[2-4] The late-life amnestic syndrome commonly observed in AD has been linked to this severe pathological load on the MTL.[3,5] This traditional view of AD progressing in one stereotypical pattern has been challenged by findings of AD pathology occurring as the primary pathology in clinical syndromes that deviate from amnestic late-onset AD, such as posterior cortical atrophy,[6] behavioural variant of AD[7] or logopenic variant of primary progressive aphasia.[8] Interestingly, these syndromes exhibit distinct spatial distribution patterns of AD pathological hallmarks, in particular of tau accumulation and neurodegeneration.[9]

Independent of clinical phenotypes, data-driven methods have been used to capture the biological heterogeneity of AD, with most studies applying statistical clustering approaches to identify diverging spatial distribution patterns of AD pathology. Studies aiming to identify distinct atrophy subtypes have commonly relied on cross-sectional structural MRI.[10-16] A recent review of these efforts suggests that spatial variability in AD-related atrophy may occur from two sources: disease severity and typicality.[17] In fact, this model points to one of the major challenges faced when applying clustering approaches to cross-sectional biomarker data. Variance in these data may originate from interindividual differences in disease progression and disease subtype, assuming that multiple exist. Thus, the identification of atrophy subtypes requires that variance in disease progression is accounted for appropriately. The Subtype and Stage Inference (SuStaIn) algorithm was designed to address this shortcoming of conventional clustering techniques.[18,19] It exploits principles from clustering and event-based modelling to recover distinct pseudo-longitudinal progression sequences, i.e. disease subtypes, from cross-sectional biomarker information. SuStaIn has been applied to structural MRI volumetric data, suggesting AD atrophy subtypes that were termed 'typical', 'cortical' and 'subcortical',[19,20] or 'typical', 'hippocampal-sparing' and 'limbic-predominant'.[21]

Despite a surging academic interest in the biological heterogeneity of AD, gaps in knowledge remain regarding the implications of recent findings in clinical settings. For instance, it is not known to what extent atrophy heterogeneity occurs in patients presenting at memory clinics. Not only does the co-occurrence of various risk factors and pathologies in many memory clinic patients render their population generally more heterogeneous than participants from many observational studies and clinical trials for AD, but also these patients are typically characterized in less detail.

Moreover, the mixed results of recent phase III trials of potential disease-modifying agents for AD suggest that the optimal window for effective treatment might lie in the preclinical to early prodromal stages of the disease.[22,23] One of the challenges faced when targeting these groups of patients is that their trajectories on end point measures are heterogeneous and deviate only marginally from those of healthy controls. In turn, the small effects expected from successful treatment at this disease stage pose a significant challenge to those aiming to design powerful yet cost-efficient pharmacological trials.[24,25] Understanding and accounting for disease heterogeneity thus appears crucial in this context. However, previous studies linking atrophy heterogeneity to distinct clinical profiles have focused on AD patients diagnosed with mild cognitive impairment (MCI) or dementia of the Alzheimer's type (DAT[17]; although see Shand et al.[26] for a non-peer-reviewed report on atrophy heterogeneity in preclinical AD). To assess the potential utility of considering atrophy heterogeneity for pharmacological trial design, investigations of the prevalence and clinical implications of atrophy heterogeneity in the absence of significant cognitive impairment are warranted.

In addition, a common limitation of disease progression models is their lacking or unknown generalizability to external datasets (i.e. testing outside of the training dataset).[27-29] Recent work by Chekroud et al.[27] demonstrated in various examples of disease progression models that, even when performance in internal validation procedures such as *k*-fold cross-validation was satisfactory, models performed poorly, at around chance level, in external testing data. However, the reliable classification of previously unseen data is a key requirement for models to make their way into applied settings.

In the present study, we investigate atrophy subtypes in the DZNE Longitudinal Cognitive Impairment and Dementia Study (DELCODE[30]), which is a multicentre, memory clinic-based cohort targeting earliest at-risk states for DAT. We examine whether atrophy subtypes are related to distinct cross-sectional and longitudinal clinical profiles, focusing on individuals who are cognitively

unimpaired (CU) or report subjective cognitive decline (SCD). Core analyses are repeated in an external sample of the Swedish BioFINDER-2 study to ensure group- and subject-level generalizability to inform potential model application.

Materials and methods

Participants

Out of 1078 individuals enrolled in DELCODE, 813 had available baseline structural MRI segmentation data that passed visual quality control and were thus included in the present study. Participants were classified as either CU ($n = 285$) or fulfilled diagnostic criteria for SCD ($n = 342$), MCI ($n = 118$) or DAT ($n = 68$; see Supplementary material, Methods section for inclusion criteria).[30] Participants in the SCD, MCI and DAT groups were referrals, including self-referrals, from 10 German university-based memory clinics. CU participants were recruited via public advertisement. All participants provided their written informed consent to participate in the study according to the Declaration of Helsinki. The study protocol was approved by the local institutional review boards of all participating institutions. DELCODE has been registered with the German Clinical Trials Registry (DRKS; DRKS00007966) prior to inclusion of first participants.

Fluid biomarkers

Lumbar CSF samples were available in a subsample of 385 (47.36%) participants. Levels of phosphorylated tau (p-tau) 181 (Innotest essay; Fujirebio), total tau (t-tau), $A\beta_{40}$ and $A\beta_{42}$ (all Mesoscale platform; Meso Scale Diagnostics) were determined. A threshold for Aβ-positivity of ≤0.08 was obtained through two-component Gaussian mixture modelling of $A\beta_{42}/A\beta_{40}$ ratios. Plasma samples were analysed for neurofilament light chain, $A\beta_{x\text{-}42}$ and $A\beta_{x\text{-}40}$ levels. In participants with missing CSF data, Aβ-positivity was determined from EDTA plasma $A\beta_{x\text{-}42}/A\beta_{x\text{-}40}$ ratios (Supplementary material, Methods section).

Neuropsychological assessment

Cognitive functioning was measured in two domains: learning and memory (MEM) and executive functions and mental processing speed (EXEC). These scores were previously obtained from the DELCODE neuropsychological test battery using confirmatory factor analysis and z-standardized to a reference group of CU and SCD participants.[29] A difference score (MEM–EXEC) was calculated to estimate the relative impairment of participants in these domains. Negative values reflected a relative impairment in the MEM domain, whereas positive values reflected a relative impairment in the EXEC domain. In addition, the Preclinical Alzheimer Cognitive Composite (PACC-5) was used to measure a set of cognitive functions that have been shown to exhibit early AD-related decline.[31] Cognitive trajectories were estimated from longitudinal PACC-5 scores. To reflect a realistic time frame of a hypothetical clinical trial, we used available data collected ≤240 weeks after baseline, as done in previous trials of monoclonal antibodies targeting Aβ (e.g. solanezumab[32] or lecanemab[33]). These data were available in 676 participants [83.14%; mean ± standard deviation (SD) number of follow-up visits = 2.84 ± 1.09, mean ± SD follow-up interval = 1.08 ± 0.27 years].

Among participants diagnosed as CU or SCD at baseline, progression to MCI was determined based on the available neuropsychological data (see Stark *et al.*[34]). Again, only incident MCI diagnoses given ≤240 weeks after baseline were considered.

Structural MRI acquisition and processing

All structural MRI data were collected using 3 T Siemens MRI systems. T1-weighted images were acquired using a 3D whole-brain magnetization prepared rapid gradient echo sequence [MPRAGE; echo time/repetition time (TE/TR) = 437/2500 ms, inversion time = 1100 ms, 7° flip angle, 1 mm isotropic resolution]. Coronal T2-weighted turbo spin-echo (TSE) images were acquired orthogonally to the longitudinal axis of the hippocampus on a slab that covered the MTL (TE/TR = 354/3500 ms, 120° flip angle, 0.5 mm × 0.5 mm × 1.5 mm resolution).

FreeSurfer (v.7; http://surfer.nmr.mgh.harvard.edu/) and the automated segmentation of hippocampal subfields algorithm (ASHS,[35] using the Penn ABC-3T atlas[36]) were used to obtain region of interest-based grey matter volume and average cortical thickness data. ASHS was implemented using an in-house longitudinal segmentation pipeline if usable follow-up structural MRI data were available (Supplementary material, Methods section). For participants without longitudinal data, ASHS was implemented conventionally. We summed up volumes of the amygdala label from FreeSurfer's aseg atlas and of the hippocampal and entorhinal ASHS labels to obtain bilateral MTL volumes. Grey matter volumes were adjusted for their relationship with total intracranial volume in Aβ-negative CU participants ($n = 187$).[37] Average cortical thickness was calculated for temporal, parietal, frontal and occipital regions of interest based on FreeSurfer's Desikan–Killiany parcellation (Supplementary material, Methods section).[38]

SuStaIn modelling

To identify atrophy subtypes from structural MRI data, we implemented SuStaIn using python (v.3.9).[19] All structural MRI markers were z-standardized to a reference group comprising Aβ-negative CU participants and corrected for their relationships with age, sex and years of education in this reference group. Across markers, $z = -1$ and -2 were chosen as atrophy event thresholds in line with earlier work.[19] In total, the model identified sequences of 10 atrophy events, here referred to as SuStaIn stages, with each stage reflecting the surpassing of one out of two z-score thresholds in one out of five structural MRI markers. Note that SuStaIn stages do not represent continuous, time-equivalent intervals. As in previous studies, linear modelling across SuStaIn stages was applied.[19,39] Model selection (i.e. selection of number of subtypes) and sequence stability assessment were performed as previously reported.[19] Participants who did not exhibit significant atrophy were not assigned a subtype but were referred to as a separate atrophy-negative group. Note that for these participants, as for others, probabilities of belonging to each subtype were calculated.

The SuStaIn model was applied to follow-up structural MRI data ($n = 546$, mean ± SD follow-up time = 1.03 ± 0.08 years). This allowed us to test whether participants progressed in line with the theoretical assumptions of subtype robustness and stage monotonicity.

To ensure that atrophy subtypes could be identified reliably in subsamples of particular interest, SuStaIn modelling was repeated in participants representing the biologically defined AD spectrum (i.e. only the reference group and Aβ-positive individuals; $n = 457$) and in participants without manifest cognitive impairment (i.e. CU and SCD participants).

Further details on SuStaIn modelling parameters, internal 10-fold cross-validation, model selection, longitudinal model validation and subsample replication are provided in the Supplementary material, Methods section.

Replication cohort

Replication analyses and assessment of model generalizability were performed on data from 779 participants enrolled in the Swedish BioFINDER-2 cohort. To ensure comparability with DELCODE, only BioFINDER-2 participants aged ≥60 years and who belonged to the CU ($n = 283$), SCD ($n = 152$), MCI ($n = 212$) or DAT ($n = 132$) diagnostic groups were included (see Palmqvist *et al.*[40,41] and Supplementary material, Methods for inclusion criteria).

Latent MEM and EXEC domain scores were unavailable for BioFINDER-2. Instead, the Symbol Digit Modalities Test (SDMT;[42] one point for every correct answer within the response time of 90 s) was chosen as a measure of executive functions and attention. The delayed 10-word list recall test from the Alzheimer's Disease Assessment Scale-Cognitive subscale (ADAS-Cog[43]; number of errors) was used as a measure of episodic memory. ADAS-Cog delayed recall scores were inverted such that lower scores on all cognitive measures reflected worse performance. Both ADAS-Cog and SDMT scores were z-standardized to CU participants. In BioFINDER-2, a modified version of the PACC (mPACC) was used.[41,44] Longitudinal mPACC scores obtained ≤240 weeks after baseline were available for 438 participants (56.23%; mean ± SD number of follow-up visits = 1.85 ± 0.92, mean ± SD follow-up interval = 1.41 ± 0.58 years).

Further methodological details for the BioFINDER-2 sample are provided in the Supplementary material, Methods section.

Statistical analysis

Statistical analyses were carried out using R (v.4.2.2).

We performed analyses of variance and Tukey's *post hoc* comparisons to test for differences in mean SuStaIn stage in each patient group against CU participants. We assessed the cross-sectional relationships between atrophy subtype and demographic, as well as clinical variables. Using ordinary least squares regression models for continuous variables and binomial logistic regression models for binary dependent variables, subtypes were first contrasted against the atrophy-negative group and, second, against each other.

Longitudinal PACC-5 slopes were examined using linear mixed effects models with random intercepts and slopes using restricted maximum likelihood (REML) and Bound Optimization BY Quadratic Approximation (BOBYQA) implemented in the lme4 package.[45] First, we calculated PACC-5 slopes for each baseline subtype and for the atrophy-negative group in separate models. Second, we tested for differences in slopes by including pairwise interaction terms of Time × Baseline subtype. Third, we tested for interaction effects of Time × Baseline SuStaIn stage in each subtype.

Cox proportional hazard regression models were used to predict the effect of baseline atrophy subtype and SuStaIn stage on risk of progression to MCI among CU and SCD participants.

We ran binomial logistic regression models in the atrophy-negative group to test whether baseline subtype probability could predict conversion to each subtype at follow-up. To investigate whether the inclusion of additional baseline variables would improve predictive power, forward step-wise regression was implemented using the stepAIC function from the MASS package.[46] The maximum model included subtype probability, age, sex, years of education, diagnostic group, plasma neurofilament light chain level, PACC-5 score and *APOE* ϵ4 status. Finally, a model including all predictors of the maximum model except subtype probability was calculated. We did not include baseline CSF biomarkers and Aβ status owing to limited sample sizes, or cognitive domain scores owing to their low practical applicability.

The pseudo-longitudinal development of structural MRI markers across SuStaIn stages was modelled using monotone penalized cubic regression splines. Models of cognitive scores across SuStaIn stages were fitted using natural cubic regression splines.

To test generalizability, we correlated SuStaIn stages (using Spearman rank correlation) and subtype probabilities (using Pearson correlation) obtained from internally and externally trained SuStaIn models in both the DELCODE and BioFINDER-2 samples.

The threshold for statistical significance was $P < 0.05$. Where appropriate, false discovery rate (FDR) correction was applied. Models always controlled for diagnostic group where CU and SCD groups were collapsed, because SCD does not occur in all AD patients[47] and because we assumed that SCD might be related to atrophy subtype. When fluid biomarkers were assessed, age and sex were additionally controlled for. Analyses of cognitive measures and incident MCI diagnoses also controlled for years of education.

Table 1 Baseline characteristics of the DELCODE sample

Variable	Missing	Overall N = 813	Diagnostic group			
			CU n = 285	SCD n = 342	MCI n = 118	DAT n = 68
Age, years	0 (0.00%)	70.67 (6.07)	68.61 (5.35)	70.93 (6.05)	72.99 (5.87)	73.97 (6.31)
Females	0 (0.00%)	423 (52.03%)	165 (57.89%)	159 (46.49%)	57 (48.31%)	42 (61.76%)
Education, years	0 (0.00%)	14.54 (2.98)	14.65 (2.76)	14.88 (2.99)	14.07 (3.20)	13.19 (3.02)
Atrophy subtype						
Atrophy-negative	0 (0.00%)	459 (56.46%)	199 (69.82%)	214 (62.57%)	40 (33.90%)	6 (8.82%)
Limbic-predominant	0 (0.00%)	188 (23.12%)	28 (9.82%)	56 (16.37%)	48 (40.68%)	56 (82.35%)
Hippocampal-sparing	0 (0.00%)	166 (20.42%)	58 (20.35%)	72 (21.05%)	30 (25.42%)	6 (8.82%)
APOE ϵ4 carrier	5 (0.62%)	274 (33.91%)	70 (24.91%)	104 (30.50%)	59 (50.00%)	41 (60.29%)
Aβ-positive	110 (13.53%)	270 (38.41%)	56 (23.05%)	106 (34.87%)	61 (60.40%)	47 (85.45%)
PACC-5, score	41 (5.04%)	−0.34 (1.16)	0.17 (0.59)	−0.08 (0.67)	−1.31 (0.95)	−3.70 (1.33)

For continuous variables, the mean (standard deviation) are reported. For categorical variables, *n* (%) are reported. All percentages refer to subsamples with available data for the respective variable. CU = cognitively unimpaired; DAT = dementia of the Alzheimer's type; MCI = mild cognitive impairment; PACC-5 = Preclinical Alzheimer Cognitive Composite; SCD = subjective cognitive decline.

Results

Two atrophy subtypes identified from cross-sectional MRI

Table 1 provides an overview of DELCODE participants. The majority did not exhibit abnormal atrophy and were thus classified as atrophy-negative ($n = 459$, 56.46%; 46.67% of Aβ-positive participants). In the remaining 354 participants, two distinct atrophy progression sequences were identified (Fig. 1A and C, Supplementary Fig. 1 and Supplementary material, Results section). A limbic-predominant subtype ($n = 188$, 53.11% of participants assigned a subtype; 70.14% of Aβ-positive participants assigned a subtype) showed initial abnormal volume loss in the MTL, followed by cortical thinning in the temporal, parietal, frontal and, eventually, occipital lobes. A hippocampal-sparing subtype ($n = 166$, 46.89% of participants assigned a subtype; 29.86% of Aβ-positive participants assigned a subtype) initially exhibited reduced cortical thickness in the frontal, occipital and parietal lobes, before the temporal lobe and, ultimately, the MTL were affected by atrophy.

Analyses of follow-up MRI scans revealed that most observed atrophy trajectories (92.52% of participants with limbic-predominant and 86.24% of participants with baseline hippocampal-sparing atrophy) were in line with the modelled atrophy sequences (Supplementary material, Results section). Evidence for highly similar atrophy subtypes was found in the Aβ-positive and CU/SCD subsamples (Supplementary Fig. 2 and Supplementary material, Results section).

The two atrophy subtypes were represented across diagnostic groups. Atrophy-negative cases dominated the CU and SCD groups, whereas their proportion was smaller in the MCI and DAT groups. Among participants who were assigned a subtype, the hippocampal-sparing subtype was more common among CU and SCD participants, whereas the limbic-predominant subtype was more prevalent in the MCI and DAT groups (Fig. 1B).

Mean SuStaIn stage differed among diagnostic groups in both atrophy subtypes [limbic-predominant, $F(3,184) = 16.44$; hippocampal-sparing, $F(3,162) = 13.80$; both $P < 0.001$; Fig. 1D]. Mean SuStaIn stage in CU participants (limbic-predominant, 1.46 ± 0.88; hippocampal-sparing, 2.64 ± 1.58) was not significantly different from that in SCD patients (limbic-predominant, 2.14 ± 1.73, $P_{FDR} = 0.167$; hippocampal-sparing, 2.69 ± 1.83, $P_{FDR} = 0.865$) but was significantly lower than in MCI (limbic-predominant, 2.85 ± 2.13, $P_{FDR} = 0.009$; hippocampal-sparing, 3.80 ± 2.33, $P_{FDR} = 0.010$) and DAT patients (limbic-predominant, 4.43 ± 2.78; hippocampal-sparing, 7.33 ± 2.66, both $P_{FDR} < 0.001$). This indicates that atrophy stage increases with prodromal cognitive impairment across both subtypes.

Atrophy subtypes demonstrate distinct clinical profiles at baseline

Baseline characteristics of individuals in both atrophy subtypes were compared with the atrophy-negative group (Fig. 2A–E and Supplementary Tables 1 and 2). Hippocampal-sparing atrophy was related to higher educational levels ($b = 0.76$, $\eta^2_{partial} = 0.01$, $P_{FDR} = 0.015$). Meanwhile, limbic-predominant atrophy was related to older age ($b = 2.82$, $\eta^2_{partial} = 0.09$, $P_{FDR} < 0.001$), higher educational levels ($b = 0.77$, $\eta^2_{partial} < 0.01$, $P_{FDR} = 0.011$), higher odds of *APOE* ε4 carriership [$b = 0.63$, odds ratio (OR) = 1.87, $P_{FDR} = 0.006$] and being Aβ-positive ($b = 0.60$, OR = 1.82, $P_{FDR} = 0.015$), in addition to more pathological levels across fluid biomarkers, with the exception of plasma neurofilament light chain (CSF $A\beta_{42}/A\beta_{40}$ ratio, $b = -0.01$, $\eta^2_{partial} = 0.22$; CSF p-tau 181, $b = 16.18$, $\eta^2_{partial} = 0.18$; CSF t-tau: $b = 119.94$, $\eta^2_{partial} = 0.21$; all $P_{FDR} < 0.001$). Regarding cognitive performance, limbic-predominant atrophy was associated with lower

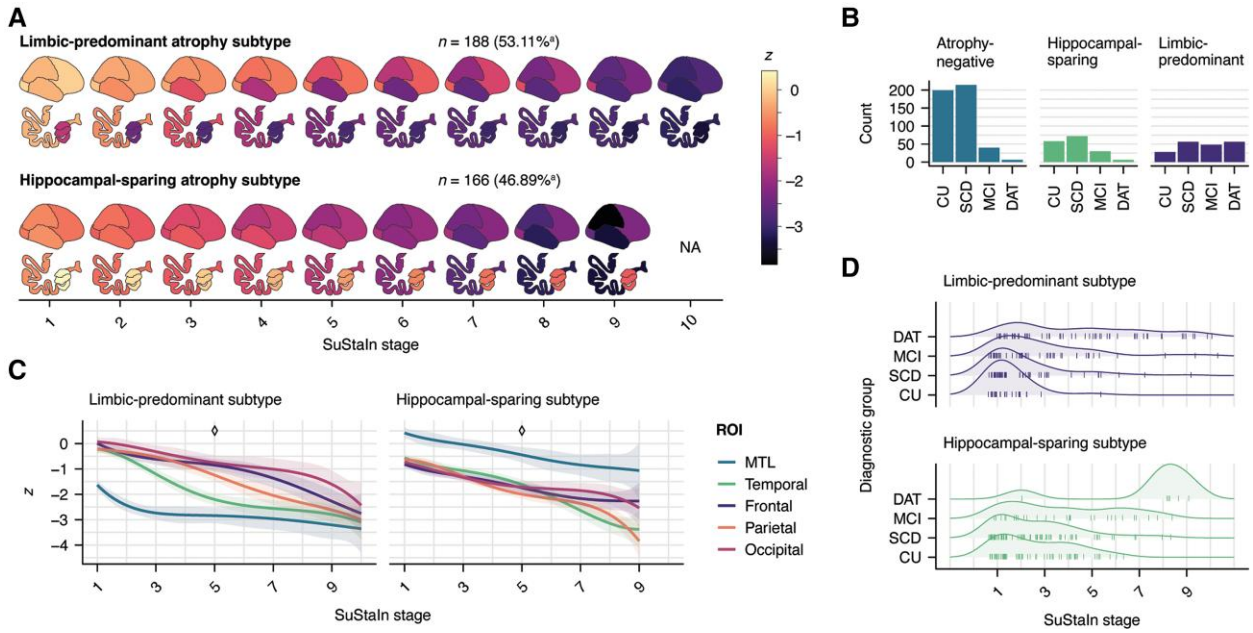

Figure 1 Two atrophy subtypes were identified in the DELCODE sample. (**A**) The progression of atrophy across SuStaIn stages. Atrophy is measured in z-scores that were scaled and centred to Aβ-negative cognitively unimpaired (CU) participants. The displayed values were obtained by (**C**) modelling each region of interest-based atrophy marker across SuStaIn stages using monotone regression splines. The open diamond denotes the knot position at SuStaIn stage = 5. (**B**) Distributions of diagnostic groups across the atrophy-negative group and the two atrophy subtypes. (**D**) Distributions of SuStaIn stages for each subtype and diagnostic group. Data-points are jittered on the x-axis. [a]Percentages refer to the proportion among all participants assigned an atrophy subtype (excluding atrophy-negative individuals). DAT = dementia of the Alzheimer's type; MCI = mild cognitive impairment; MTL = medial temporal lobe; ROI = region of interest; SCD = subjective cognitive decline; SuStaIn = Subtype and Stage Inference.

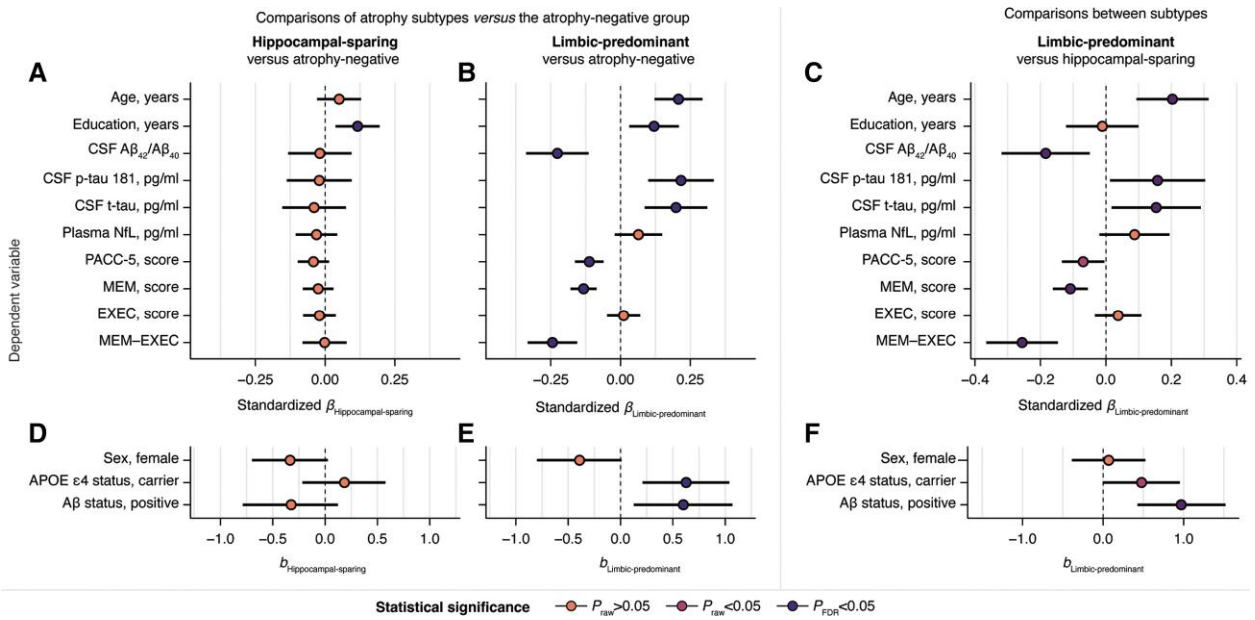

Figure 2 The two atrophy subtypes and the atrophy-negative group exhibit different cross-sectional clinical profiles. (**A**–**C**) The *top row* shows standardized β-coefficients from ordinary least squares linear regression models predicting each continuous dependent variable, with atrophy subtype as the predictor of interest. (**D**–**F**) The bottom row shows unstandardized estimates of the effect of subtype in binomial logistic regression models predicting each binary dependent variable, with subtype as the predictor of interest. Initially, models were fitted to compare participants with (**A** and **D**) hippocampal-sparing atrophy and (**B** and **E**) limbic-predominant atrophy against the atrophy-negative group, before (**C** and **F**) models comparing the two atrophy subtypes against each other were fitted. The displayed effects are controlled for diagnostic group when predicting demographic variables (including *APOE* ε4 status), for age, sex and diagnostic group when predicting fluid biomarkers and for age, sex, education and diagnostic group when predicting cognitive scores. Error bars visualize 95% confidence intervals. EXEC = executive functions and mental processing speed; MEM = learning and memory; NfL = neurofilament light chain; PACC-5 = Preclinical Alzheimer Cognitive Composite.

PACC-5 ($b = -0.31$, $\eta^2_{partial} = 0.38$, $P_{FDR} < 0.001$) and MEM scores ($b = -0.28$, $\eta^2_{partial} = 0.54$, $P_{FDR} < 0.001$), as well as lower MEM–EXEC difference scores ($b = -0.30$, $\eta^2_{partial} = 0.08$, $P_{FDR} < 0.001$), indicative of pronounced amnestic cognitive impairment.

Next, the two subtypes were contrasted against each other with the hippocampal-sparing subtype as the reference group (Fig. 2C and F and Supplementary Table 3). Participants with limbic-predominant atrophy were older ($b = 2.52$, $\eta^2_{partial} = 0.06$, $P_{FDR} = 0.001$) and more likely to be Aβ-positive ($b = 0.97$, OR = 2.64, $P_{FDR} = 0.002$). CSF $A\beta_{42}/A\beta_{40}$ ratios ($b = -0.01$, $\eta^2_{partial} = 0.19$, $P_{FDR} = 0.019$), p-tau 181 levels ($b = 13.81$, $\eta^2_{partial} = 0.12$, $P_{FDR} = 0.041$) and t-tau levels ($b = 105.60$, $\eta^2_{partial} = 0.16$, $P_{FDR} = 0.041$) were more abnormal in the limbic-predominant subtype. Limbic-predominant atrophy was linked to lower MEM ($b = -0.23$, $\eta^2_{partial} = 0.49$, $P_{FDR} < 0.001$) and MEM–EXEC scores ($b = -0.31$, $\eta^2_{partial} = 0.08$, $P_{FDR} < 0.001$) and tended to be associated with lower PACC-5 scores ($b = -0.20$, $\eta^2_{partial} = 0.29$, $P_{FDR} = 0.054$). Raw data are depicted in Supplementary Fig. 3.

In only Aβ-positive participants, there was no significant association of atrophy subtype with demographics and fluid AD biomarkers, while the relationships of atrophy subtype with cognitive scores were replicated (Supplementary Tables 4–6).

Amnestic cognitive decline already occurs in early stages of limbic-predominant atrophy

Natural cubic regression splines were used to estimate the pseudo-longitudinal trajectories of baseline cognitive scores across SuStaIn stages (Fig. 3A–D). In the limbic-predominant subtype, PACC-5, MEM and EXEC scores declined with increasing SuStaIn stage. MEM scores were lower relative to EXEC scores across SuStaIn stages. Meanwhile, all three cognitive scores remained relatively stable across the first five SuStaIn stages of hippocampal-sparing atrophy before steeply declining. This decline was more pronounced in EXEC scores than in MEM scores, resulting in a larger relative impairment of the EXEC domain with increasing SuStaIn stage.

These pseudo-longitudinal trajectories were in line with the observed slopes of PACC-5 scores over time. PACC-5 scores decreased at trend level in those with hippocampal-sparing atrophy ($b = -0.03$, $\eta^2_{partial} = 0.04$, $P = 0.068$) and significantly at a large effect size in those with limbic-predominant atrophy ($b = -0.16$, $\eta^2_{partial} = 0.34$, $P < 0.001$). The atrophy-negative group exhibited a small practice effect ($b = 0.02$, $\eta^2_{partial} = 0.04$, $P < 0.001$; Fig. 3E). These slopes were significantly different from each other (all $P_{FDR} \leq 0.005$). In both subtypes, PACC-5 slopes were negatively related to SuStaIn stage, with stronger interaction effects of Time × SuStaIn stage in the limbic-predominant subtype ($b = -0.05$, $\eta^2_{partial} = 0.15$, $P = 0.002$) than in the hippocampal-sparing subtype ($b = -0.03$, $\eta^2_{partial} = 0.08$, $P = 0.005$; Fig. 3F and G and Supplementary Table 7).

In only Aβ-positive participants, we did not observe a practice effect in the atrophy-negative group ($b = -0.02$, $\eta^2_{partial} = 0.03$; $P = 0.165$). PACC-5 scores declined significantly in both atrophy subtypes, although the effect size was larger in the limbic-predominant subtype ($b = -0.37$, $\eta^2_{partial} = 0.41$, $P < 0.001$) than in the hippocampal-sparing subtype ($b = -0.20$, $\eta^2_{partial} = 0.24$, $P = 0.005$). Increasing SuStaIn stage predicted an accelerated decline of PACC-5 scores in the hippocampal-sparing subtype ($b = -0.08$, $\eta^2_{partial} = 0.16$, $P = 0.013$). A similar trend was observed in the limbic-predominant subtype ($b = -0.07$, $\eta^2_{partial} = 0.07$, $P = 0.077$; Supplementary Table 8).

Early limbic-predominant atrophy predicts cognitive decline in clinically normal older adults

To test the predictive properties of atrophy subtype and stage before the onset of manifest cognitive and functional impairment, the

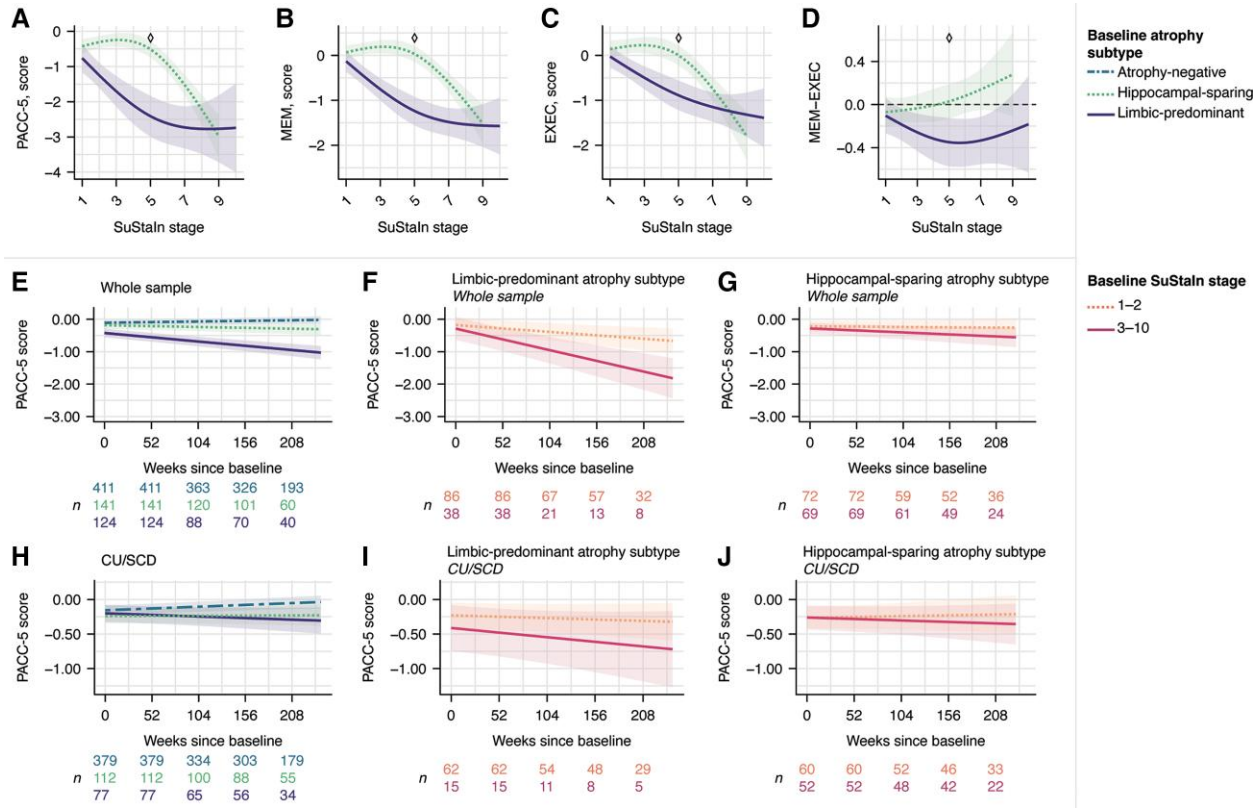

Figure 3 Visualization of cognitive trajectories across pseudo-longitudinal SuStaIn stages and longitudinal annual assessments. (**A**) PACC-5 scores, (**B**) MEM and (**C**) EXEC domain scores, as well as (**D**) MEM–EXEC scores, were fitted across SuStaIn stages using natural cubic regression splines that controlled for age, sex and years of education. The open diamond denotes denotes the knot position at SuStaIn stage = 5. (**D**) Negative values (below dashed line) represent a pronounced impairment of the MEM domain, whereas positive values (above dashed line) represent a pronounced impairment of the EXEC domain. Linear mixed effects models were used to estimate longitudinal PACC-5 slopes for (**E**) each atrophy subtype and (**F** and **G**) across SuStaIn stages (here displayed stratified by median split). (**H–J**) These analyses were repeated in only CU and SCD participants. CU = cognitively unimpaired; EXEC = executive functions and mental processing speed; MEM = learning and memory; PACC-5 = Preclinical Alzheimer Cognitive Composite; SCD = subjective cognitive decline; SuStaIn = Subtype and Stage Inference.

aforementioned analyses were repeated in only CU and SCD participants. In the limbic-predominant ($b = -0.03$, $\eta^2_{partial} = 0.05$, $P = 0.153$) and hippocampal-sparing ($b = 0.00$, $\eta^2_{partial} < 0.01$, $P = 0.844$) atrophy subtypes, PACC-5 scores did not change significantly over time. The atrophy-negative group exhibited a significant practice effect ($b = 0.03$, $\eta^2_{partial} = 0.05$, $P < 0.001$). Analyses of Atrophy subtype × Time interaction effects showed that the PACC-5 slopes associated with limbic-predominant ($b = -0.05$, $\eta^2_{partial} = 0.02$, $P_{FDR} = 0.030$) but not hippocampal-sparing atrophy ($b = -0.02$, $\eta^2_{partial} = 0.01$, $P_{FDR} = 0.202$; Fig. 3H) differed significantly from this practice effect. Decline on the PACC-5 was steeper with increasing SuStaIn stage in the hippocampal-sparing ($b = -0.03$, $\eta^2_{partial} = 0.07$, $P = 0.019$) but not the limbic-predominant ($b = -0.01$, $\eta^2_{partial} = 0.01$, $P = 0.580$) atrophy subtype (Fig. 3I and J and Supplementary Table 9).

Cox proportional hazard regression models showed that limbic-predominant atrophy was associated with an elevated risk of progressing to MCI within 240 weeks in comparison to the atrophy-negative group [hazard ratio (HR) = 2.28, $P_{FDR} = 0.045$], but not in comparison to hippocampal-sparing atrophy (HR = 1.69, $P_{FDR} = 0.312$). There was no difference between participants in the atrophy-negative group and in the hippocampal-sparing subtype (HR = 1.25, $P_{FDR} = 0.509$; Fig. 4A). The risk of progression to MCI increased with higher SuStaIn stage in the limbic-predominant subtype (HR = 1.48, $P = 0.004$) but not in the hippocampal-sparing atrophy subtype (HR = 1.24, $P = 0.147$; Fig. 4B and C).

Future atrophy subtype can be predicted from baseline MRI

Among participants in the atrophy-negative group, baseline subtype probability was predictive of conversion to both hippocampal-sparing atrophy [area under the curve (AUC) = 0.74, Akaike information criterion (AIC) = 131.75; OR = 1.05, $P < 0.001$] and limbic-predominant atrophy (AUC = 0.88, AIC = 70.67; OR = 1.15, $P < 0.001$) at follow-up.

Forward step-wise inclusion of additional baseline characteristics showed that higher baseline plasma neurofilament light chain levels (OR = 1.11, $P < 0.001$) additionally increased the odds of progression to hippocampal-sparing atrophy (AUC = 0.83, AIC = 122.73). In turn, lower baseline PACC-5 scores (OR = 0.28, $P = 0.005$) significantly increased the odds of progression to limbic-predominant atrophy (AUC = 0.92, AIC = 65.38). This model also included years of education as a non-significant term. The effects of baseline subtype probability remained positive and significant in both models. Models relying only on non-imaging baseline characteristics showed lower predictive performance than the forward-fit models (limbic-predominant subtype, AUC = 0.83, AIC = 88.80; hippocampal-sparing subtype, AUC = 0.76, AIC = 147.24). Receiver operating characteristic curves for all models are shown in Supplementary Fig. 4. Model summaries are provided in Supplementary Table 10.

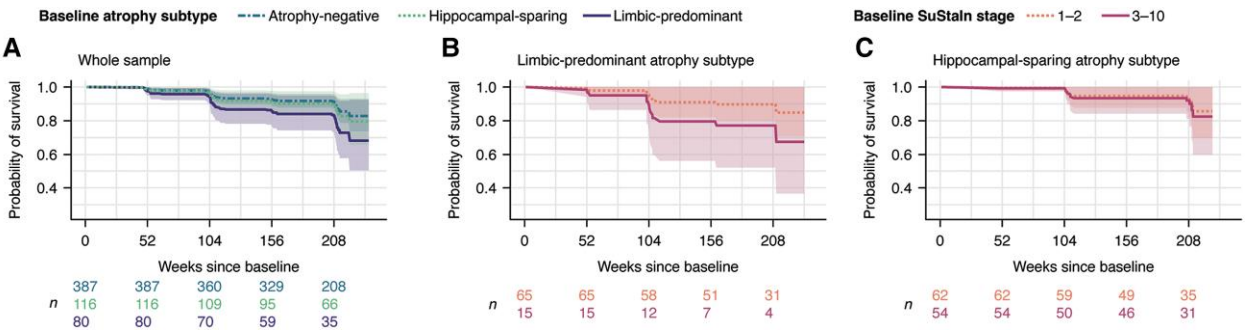

Figure 4 Kaplan–Meier survival curves displaying the estimated probability of remaining without an incident mild cognitive impairment diagnosis over time among cognitively unimpaired subjects and those with subjective cognitive decline. Visualized are: (**A**) the effect of atrophy subtype and the effects of baseline SuStaIn stage (here displayed stratified by median split) in both the (**B**) limbic-predominant and (**C**) hippocampal-sparing atrophy subtypes. SuStaIn = Subtype and Stage Inference.

Analogous atrophy subtypes identified in external cohort

De novo SuStaIn modelling naïve to the main DELCODE model successfully replicated the two atrophy progression sequences in the BioFINDER-2 sample (Fig. 5A and Supplementary material, Results section). The atrophy-negative group comprised 344 participants (44.16%; 34.13% of Aβ-positive participants). Limbic-predominant atrophy was present in 221 participants (50.80% of participants assigned a subtype, 55.74% of Aβ-positive participants assigned a subtype), and 214 (49.20% of participants assigned a subtype, 44.26% of Aβ-positive participants assigned a subtype) participants showed hippocampal-sparing atrophy.

Downstream analyses of SuStaIn results revealed highly similar associations with cognitive performance as shown in DELCODE. Baseline episodic memory performance (ADAS-Cog delayed recall) was lower in limbic-predominant atrophy, whereas attention and executive functioning (SDMT) were more impaired in hippocampal-sparing atrophy. There was no difference in mPACC scores between atrophy subtypes (Supplementary Table 11 and Supplementary Fig. 6).

Longitudinal mPACC scores declined fastest in the limbic-predominant subtype, followed by the hippocampal-sparing subtype and the atrophy-negative group. In both subtypes, increasing SuStaIn stages were linked to faster-declining mPACC scores (Supplementary Table 12 and Supplementary Fig. 7).

High subject-level generalizability of models across cohorts

Out-of-sample testing of both the DELCODE- and BioFINDER-2-based models in the respective other cohort revealed excellent generalizability. In both samples, SuStaIn stages (BioFINDER-2, $\rho = 0.99$, $P < 0.001$; DELCODE, $\rho = 0.97$, $P < 0.001$; Fig. 5B) and probabilities of limbic-predominant atrophy (BioFINDER-2, $r = 0.98$, $P < 0.001$; DELCODE, $r = 0.99$, $P < 0.001$; Fig. 5C) derived from the internally and externally trained models were almost perfectly correlated. Only few participants were assigned different subtypes by the two models (BioFINDER-2, $n = 23$, 2.95%; DELCODE, $n = 29$, 3.57%; Fig. 5D). These results indicate that the proposed model is not only applicable to internal training data but can also be used to determine atrophy subtype and stage reliably in previously unseen individuals.

Discussion

Using data-driven modelling of cross-sectional, MRI-based atrophy markers, the present study uncovered two atrophy subtypes in the memory clinic-based DELCODE cohort. The identified limbic-predominant and hippocampal-sparing atrophy patterns were associated with distinct clinical and cognitive cross-sectional profiles and longitudinal trajectories. Although trained on purely cross-sectional data, the validity of the model was strengthened further by analyses of follow-up MRI scans, where atrophy deviating from the proposed progression sequences was rarely observed. Atrophy subtypes and their cognitive correlates were replicated in the independent BioFINDER-2 cohort. Cross-cohort generalizability was excellent both on the group and the subject level.

Atrophy subtypes resemble known biological subtypes of Alzheimer's disease

Although our sample did not include exclusively individuals with biomarker evidence of AD pathology, the two atrophy subtypes were recovered in a subsample representing the biological AD continuum. This result converges with various previous studies that identified limbic-predominant and hippocampal-sparing AD subtypes in a data-driven manner using *in vivo* imaging methods including structural MRI,[10,11,13,15,16,48-57] tau-PET[39,58] and fluorodeoxyglucose-PET,[59] in addition to *ex vivo* histological investigations[60-63] (for reviews, see Ferreira *et al.*[17] and Habes *et al.*[64]).

Our results are in line with a recently proposed model that describes heterogeneity in AD-related atrophy along two dimensions, namely typicality (limbic-predominant versus hippocampal-sparing atrophy in the present study) and severity (SuStaIn stage in the present study).[17] In addition, our study presents a straightforward implementation of this framework, requiring only a single anatomical MRI scanning session. Given that our study used a heterogeneous yet highly representative memory clinic-based sample, we suggest that applying this framework can be worthwhile in clinical settings where it might be unclear whether AD pathology is present and whether it is the primary aetiology of brain atrophy and cognitive decline.

Previous investigations comparing the pathology burden in different biological subtypes of AD presented mixed results, with one study finding more abnormal AD fluid biomarker levels in participants with limbic-predominant tau pathology,[58] whereas other studies showed no differences in Aβ burden among AD subtypes defined from

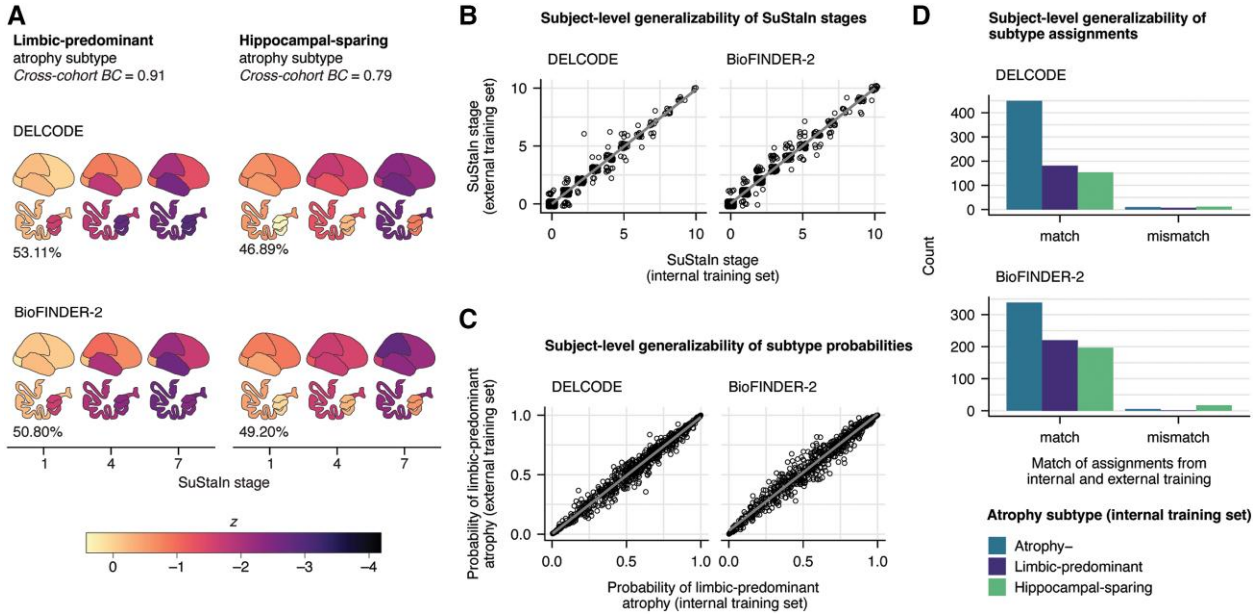

Figure 5 Overview of the BioFINDER-2 model and subject-level generalizability testing across cohorts. (**A**) Two highly comparable atrophy progression patterns were identified in the DELCODE and BioFINDER-2 cohorts. (**B**) SuStaIn stages and (**C**) subtype probabilities generated using the internally and externally trained models were highly correlated. Note that the probability of limbic-predominant atrophy can be interpreted as one minus the probability of hippocampal-sparing atrophy. (**D**) Subtype assignments were highly consistent across internally and externally trained models. BC = Bhattacharyya coefficient; SuStaIn = Subtype and Stage Inference.

histopathology[63] and *in vivo* imaging.[51,59] In the present study, levels of various fluid AD biomarkers were most abnormal in the limbic-predominant subtype. Interestingly, we found no differences in fluid biomarker levels when restricting our analyses to Aβ-positive participants. These results suggest that hippocampal-sparing atrophy might occur owing to various neuropathological processes that include, but are not limited to, hippocampal-sparing AD pathology. Nevertheless, a substantial number of participants who were assigned an atrophy subtype and who had abnormal biomarkers of Aβ were classified as having hippocampal-sparing atrophy (29.86% in DELCODE, 44.26% in BioFINDER-2), suggesting that atrophy in the presence of AD pathology does not necessarily always follow a Braak-like progression pattern. Meanwhile, the limbic-predominant subtype appears to represent a more homogeneous group of patients, presumably with AD as their primary cause of neurodegeneration. It should be noted that the hippocampal-sparing subtype may still include participants with limbic-predominant accumulation of AD pathology along with co-pathology causing a hippocampal-sparing atrophy pattern.

Clinical implications of atrophy heterogeneity in the absence of manifest cognitive impairment

In the present study, limbic-predominant but not hippocampal-sparing atrophy was associated with declining PACC-5 scores over 240 weeks. Meanwhile, those without significant atrophy exhibited practice effects. In both atrophy subtypes, increasing SuStaIn stage predicted steeper decline of PACC-5 scores, highlighting that not only atrophy subtype but also atrophy stage are predictive of the cognitive trajectories of patients.

When including only older adults without manifest cognitive impairment, we did not observe an absolute decline of PACC-5 scores over time in both atrophy subtypes. Instead, limbic-predominant, but not hippocampal-sparing atrophy was related to a diminished practice effect; a phenomenon that has been suggested to be an early cognitive correlate of preclinical AD.[65-67] Our finding of faster cognitive decline with increasing SuStaIn stage in the hippocampal-sparing but not the limbic-predominant atrophy subtype might be attributable to the low variance of atrophy stage in these unimpaired individuals. At the same time, we show that limbic-predominant atrophy, especially with increasing atrophy stage, predicted an increased risk of progression to MCI, highlighting the clinical meaningfulness of atrophy subtype and stage before manifest cognitive impairment.

Assessment of atrophy subtype and stage in pharmacological trials

The recovery of both atrophy progression sequences only in clinically unimpaired individuals suggests that atrophy heterogeneity is prevalent in those potentially qualifying for inclusion in upcoming clinical trials of disease-modifying agents that increasingly target the preclinical and early prodromal stages of AD.[22,23] Our results complement a previous report of SuStaIn-based atrophy subtypes in the A4 study cohort of preclinical AD patients.[26] The identified 'typical', 'subcortical' and 'cortical' subtypes showed different rates of decline on established cognitive measures. It has been demonstrated that cognitive trajectories in these early disease stages vary substantially, impeding the design of clinical trials that aim to demonstrate presumably small-scale treatment effects.[24,25] Given that atrophy subtype and stage explained a significant portion of this variance in our study, the presented model might facilitate the identification of patients with an elevated risk of decline over the typical duration of a clinical trial.

Indeed, sample homogenization has been proposed as a strategy to improve efficiency of trials as, for instance, demonstrated by Edmonds *et al.*[68] In their *post hoc* analysis of the Alzheimer's Disease Cooperative Study trial on the efficacy of donepezil, patients

with false-positive MCI diagnoses were excluded and treatment effects that could not be detected in primary analyses were revealed (see also Oxtoby *et al.*[24] for another approach to sample homogenization in this trial). Our results suggest that sample homogenization based on atrophy subtype and stage could boost statistical power of clinical trials. For instance, trials could target participants with limbic-predominant and/or hippocampal-sparing atrophy. Given that we show that future atrophy subtype can be predicted from a set of accessible cross-sectional clinical markers (e.g. PACC-5 scores, plasma neurofilament light chain levels) along with SuStaIn-based subtype probability, participants with an elevated risk of atrophy progression to a given subtype could also be included when pursuing this strategy. Another way of capitalizing on the proposed model in the context of pharmacological trials might be atrophy subtype- and stage-specific cognitive end point measures, as discussed below. In comparison to sample homogenization, this approach is particularly promising because it allows more potential study participants to be retained.

Atrophy subtype- and stage-specific cognitive assessments may be needed

Complementing previous reports of limbic-predominant and hippocampal-sparing atrophy patterns,[17] our study describes primarily amnestic cognitive impairment in limbic-predominant atrophy (lower latent MEM scores in DELCODE and lower ADAS-Cog delayed recall scores in BioFINDER-2) next to a predominantly executive impairment linked to hippocampal-sparing atrophy (lower relative MEM–EXEC scores in DELCODE and lower SDMT scores in BioFINDER-2). Given that we found no difference between atrophy subtypes in absolute EXEC scores, it should be noted that the executive impairment in the hippocampal-sparing subtype is probably not as pronounced as the amnestic profile associated with limbic-predominant atrophy.

PACC-5 scores declined already in the earliest stages of limbic-predominant atrophy, whereas they were relatively stable in the early- and mid-stages of hippocampal-sparing atrophy. Importantly, we believe that these observations do not necessarily reflect the actual progression of cognitive health in both atrophy subtypes. Rather, they suggest that the PACC-5 is sensitive to cognitive changes occurring in earliest limbic-predominant atrophy. Even more so, as lower PACC-5 scores increased the explanatory power of models predicting conversion to limbic-predominant atrophy, the PACC-5 appears to be sensitive not only to manifest but also to ongoing neurodegenerative change in the MTL. This is expected given that the PACC-5 was designed to detect the earliest Aβ-associated cognitive changes and puts strong emphasis on episodic memory.[31] Meanwhile, more detailed neuropsychological investigations into the cognitive trajectories associated with hippocampal-sparing atrophy are needed. This seems particularly relevant because we observed plummeting cognitive scores once hippocampal-sparing atrophy reached the temporal lobes. This effect was perhaps introduced, in part, by the DELCODE inclusion criteria. Given that amnestic symptoms were required for inclusion of MCI and DAT patients, we cannot exclude the possibility that the early prodromal stages of hippocampal-sparing atrophy (i.e. non-amnestic MCI) were underrepresented in the sample. Accordingly, atrophy subtype- and stage-specific cognitive composites might be needed to ensure the desired performance of cognitive assessments for both diagnostics and disease progression monitoring in the clinic and pharmacological trials.[25]

High model generalizability as a crucial step towards application

By using an external validation set, we not only show that our implementation of SuStaIn can identify two distinct hippocampal-sparing and limbic-predominant atrophy progression patterns in an independent cohort, but we also demonstrate that the models trained in DELCODE and BioFINDER-2 exhibit excellent cross-cohort generalizability, which is an unknown or lacking property of many disease progression models.[27,28] Applying each model to the respective external cohort resulted in very high consistencies of subtype assignments (<4% inconsistencies) and near perfect correlation of SuStaIn stages and atrophy subtype probabilities (correlation coefficients of ≥0.97). A previous study achieved lower rates of successful generalization of SuStaIn-based AD subtypes, especially in SuStaIn stages, while relying on rigorous data harmonization.[20]

Although cross-cohort generalizability is a major requirement for ensuring reliable and meaningful predictions in applied scenarios, it should be noted that both DELCODE and BioFINDER-2 are highly controlled observational studies with access to advanced biomarkers and imaging resources, including a detailed quality control of image segmentations. Further studies are needed to evaluate the full extent of model generalizability and the need for data harmonization by testing model performance in other study cohorts and in real-world settings.

Limitations

The present study has limitations that need to be taken into account when interpreting its results.

First, CSF data were available for only about half of the sample, potentially leading to sampling bias and underpowered analyses. Although likely to be relevant, CSF biomarkers could not be considered in the models predicting atrophy progression in the atrophy-negative group owing to low data availability. Future studies should test the added value of combining disease biomarkers with atrophy subtype and stage information for trial recruitment and sample enrichment.

Second, previous studies suggest that the hippocampal-sparing subtype identified in the present study might comprise more than one AD subtype.[19-21,39] The focus on amnestic cognitive impairment in DELCODE might have impaired the identification of further existing atrophy subtypes. Moreover, our model was trained on anatomically very broadly defined atrophy markers, which might have impeded the detection of small-scale anatomical differences between potential variants of hippocampal-sparing atrophy. Although previous studies have proposed the lateralization of AD pathology in atypical cases, our model cannot detect such lateralization effects because it lacks hemispheric distinctions.[69]

Third, although subtype probabilities among atrophy-negative participants were predictive of conversion to both atrophy subtypes, it should be noted that these predictions were made within the training set due to the small number of converters. To ensure generalizability, future studies should repeat these analyses using independent training and test data.

Finally, our analyses cannot provide insights into the sources of atrophy heterogeneity. These aetiological explanations should be sought to guide an adequate and informed implementation of the proposed model in clinical practice.

Conclusion

In this study, we present a data-driven model that depicts atrophy along limbic-predominant and hippocampal-sparing progression

patterns. Our findings, such as the identified diverging cognitive trajectories associated with atrophy subtype and stage, demonstrate the clinically meaningful explanatory properties of the model even in older adults without manifest cognitive impairment. Although limbic-predominant and hippocampal-sparing atrophy subtypes have been described previously, we demonstrate that implementing this framework is worthwhile in applied settings, such as memory clinics, where access to advanced biomarkers might be limited and cases with ambiguous aetiology of brain atrophy are the rule rather than the exception. Importantly for potential use cases, our model relies only on cross-sectional structural MRI, which is a well-established, cost-effective and non-invasive imaging technique. Important for the transferability of the model to practical settings, such as memory clinics or pharmacological trials, our study showcases excellent generalizability at both group and individual subject levels across cohorts.

Data availability

DELCODE data, study protocol and biomaterials can be shared with partners based on individual data and biomaterial transfer agreements. For BioFINDER, anonymized data will be shared by request from a qualified academic investigator for the sole purpose of replicating procedures and results presented in the article and as long as data transfer is in agreement with EU legislation on the general data protection regulation and decisions by the Ethical Review Board of Sweden and Region Skåne, which should be regulated in a material transfer agreement.

Acknowledgements

The authors thank all DELCODE and BioFINDER-2 participants along with their relatives and caregivers for their valuable contributions.

Funding

The study was funded by the German Center for Neurodegenerative Diseases (Deutsches Zentrum für Neurodegenerative Erkrankungen; DZNE), reference number BN01. Work within the Swedish BioFINDER-2 study was supported by the Alzheimer's Association (SG-23-1061717), Swedish Research Council (2022-00775, 2018-02052, 2021-02219), ERA PerMed (ERAPERMED 2021-184), the Knut and Alice Wallenberg foundation (2017-0383), the Strategic Research Area MultiPark (Multidisciplinary Research in Parkinson's disease) at Lund University, the Swedish Alzheimer Foundation (AF-980907, AF-981132, AF-994229), the Swedish Brain Foundation (FO2021-0293, FO2022-0204 and FO2023-0163), the EU Joint Programme – Neurodegenerative Disease Research (2019-03401), the WASP and DDLS Joint call for research projects (WASP/DDLS22-066), the Rönström Family Foundation (FRS-0003), the Parkinson Foundation of Sweden (1412/22), the Cure Alzheimer's fund, the Konung Gustaf V:s och Drottning Victorias Frimurarestiftelse, the Skåne University Hospital Foundation (2020-O000028), Regionalt Forskningsstöd (2022-1259) and the Swedish Federal Government under the ALF agreement (2022-Projekt0080, 2022-Projekt0107).

Competing interests

S.P. has acquired research support (for the institution) from ki elements/ADDF and Avid. In the past 2 years, he has received consultancy/speaker fees from Bioartic, Biogen, Esai, Lilly and Roche. E.D. reports personal fees from Biogen, Roche, Lilly, Eisai and UCL Consultancy, in addition to non-financial support from Rox Health. D.B. and E.D. are scientific co-founders of neotiv GmbH and own company shares. O.H. has acquired research support (for the institution) from AVID Radiopharmaceuticals, Biogen, C2N Diagnostics, Eli Lilly, Eisai, Fujirebio, GE Healthcare and Roche. In the past 2 years, he has received consultancy/speaker fees from AC Immune, Alzpath, BioArctic, Biogen, Bristol Meyer Squibb, Cerveau, Eisai, Eli Lilly, Fujirebio, Merck, Novartis, Novo Nordisk, Roche, Sanofi and Siemens.

Supplementary material

Supplementary material is available at *Brain* online.

References

1. Jack CR Jr, Bennett DA, Blennow K, et al. NIA-AA research framework: Toward a biological definition of Alzheimer's disease. *Alzheimers Dement*. 2018;14:535-562.
2. Braak H, Braak E. Neuropathological stageing of Alzheimer-related changes. *Acta Neuropathol*. 1991;82:239-259.
3. Berron D, Vogel JW, Insel PS, et al. Early stages of tau pathology and its associations with functional connectivity, atrophy and memory. *Brain*. 2021;144:2771-2783.
4. La Joie R, Visani AV, Baker SL, et al. Prospective longitudinal atrophy in Alzheimer's disease correlates with the intensity and topography of baseline tau-PET. *Sci Transl Med*. 2020;12:eaau5732.
5. Bejanin A, Schonhaut DR, La Joie R, et al. Tau pathology and neurodegeneration contribute to cognitive impairment in Alzheimer's disease. *Brain*. 2017;140:3286-3300.
6. Tang-Wai DF, Graff-Radford NR, Boeve BF, et al. Clinical, genetic, and neuropathologic characteristics of posterior cortical atrophy. *Neurology*. 2004;63:1168-1174.
7. Ossenkoppele R, Singleton EH, Groot C, et al. Research criteria for the behavioral variant of Alzheimer disease: A systematic review and meta-analysis. *JAMA Neurol*. 2022;79:48-60.
8. Gorno-Tempini ML, Dronkers NF, Rankin KP, et al. Cognition and anatomy in three variants of primary progressive aphasia. *Ann Neurol*. 2004;55:335-346.
9. Ossenkoppele R, Schonhaut DR, Schöll M, et al. Tau PET patterns mirror clinical and neuroanatomical variability in Alzheimer's disease. *Brain*. 2016;139(Pt 5):1551-1567.
10. Dong A, Toledo JB, Honnorat N, et al. Heterogeneity of neuroanatomical patterns in prodromal Alzheimer's disease: Links to cognition, progression and biomarkers. *Brain*. 2017;140:735-747.
11. Hwang J, Kim CM, Jeon S, et al. Prediction of Alzheimer's disease pathophysiology based on cortical thickness patterns. *Alzheimers Dement (Amst)*. 2016;2:58-67.
12. Poulakis K, Pereira JB, Mecocci P, et al. Heterogeneous patterns of brain atrophy in Alzheimer's disease. *Neurobiol Aging*. 2018;65:98-108.
13. Park JY, Na HK, Kim S, et al. Robust identification of Alzheimer's disease subtypes based on cortical atrophy patterns. *Sci Rep*. 2017;7:43270.
14. Rauchmann BS, Ersoezlue E, Stoecklein S, et al. Resting-state network alterations differ between Alzheimer's disease atrophy subtypes. *Cereb Cortex*. 2021;31:bhab130.
15. ten Kate M, Dicks E, Visser PJ, et al. Atrophy subtypes in prodromal Alzheimer's disease are associated with cognitive decline. *Brain*. 2018;141:3443-3456.

16. Zhang X, Mormino EC, Sun N, et al. Bayesian model reveals latent atrophy factors with dissociable cognitive trajectories in Alzheimer's disease. *Proc Natl Acad Sci U S A*. 2016;113:E6535-E6544.
17. Ferreira D, Nordberg A, Westman E. Biological subtypes of Alzheimer disease: A systematic review and meta-analysis. *Neurology*. 2020;94:436-448.
18. Aksman LM, Wijeratne PA, Oxtoby NP, et al. Pysustain: A python implementation of the subtype and stage inference algorithm. *Softwarex*. 2021;16:100811.
19. Young AL, Marinescu RV, Oxtoby NP, et al. Uncovering the heterogeneity and temporal complexity of neurodegenerative diseases with subtype and stage inference. *Nat Commun*. 2018;9:4273.
20. Chen H, Young A, Oxtoby NP, et al. Transferability of Alzheimer's disease progression subtypes to an independent population cohort. *NeuroImage*. 2023;271:120005.
21. Archetti D, Young AL, Oxtoby NP, et al. Inter-cohort validation of SuStaIn model for Alzheimer's disease. *Front Big Data*. 2021;4:661110.
22. Cummings J, Lee G, Nahed P, et al. Alzheimer's disease drug development pipeline: 2022. *Alzheimers Dement (N Y)*. 2022;8:e12295.
23. Aisen PS, Jimenez-Maggiora GA, Rafii MS, Walter S, Raman R. Early-stage Alzheimer disease: getting trial-ready. *Nat Rev Neurol*. 2022;18:389-399.
24. Oxtoby NP, Shand C, Cash DM, Alexander DC, Barkhof F. Targeted screening for Alzheimer's disease clinical trials using data-driven disease progression models. *Front Artif Intell*. 2022;5:660581.
25. Jutten RJ, Sikkes SAM, Van der Flier WM, et al. Finding treatment effects in Alzheimer trials in the face of disease progression heterogeneity. *Neurology*. 2021;96:e2673-e2684.
26. Shand C, Markiewicz PJ, Cash DM, et al. Heterogeneity in preclinical Alzheimer's disease trial cohort identified by image-based data-driven disease progression modelling. *medRxiv*. [Preprint] https://doi.org/10.1101/2023.02.07.23285572.
27. Chekroud AM, Hawrilenko M, Loho H, et al. Illusory generalizability of clinical prediction models. *Science*. 2024;383:164-167.
28. Steyerberg EW, Vergouwe Y. Towards better clinical prediction models: Seven steps for development and an ABCD for validation. *Eur Heart J*. 2014;35:1925-1931.
29. Chekroud AM, Zotti RJ, Shehzad Z, et al. Cross-trial prediction of treatment outcome in depression: A machine learning approach. *Lancet Psychiatry*. 2016;3:243-250.
30. Jessen F, Spottke A, Boecker H, et al. Design and first baseline data of the DZNE multicenter observational study on predementia Alzheimer's disease (DELCODE). *Alzheimers Res Ther*. 2018;10:15.
31. Donohue MC, Sperling RA, Salmon DP, et al. The preclinical Alzheimer cognitive composite: Measuring amyloid-related decline. *JAMA Neurol*. 2014;71:961-970.
32. Sperling RA, Donohue MC, Raman R, et al. Trial of solanezumab in preclinical Alzheimer's disease. *N Engl J Med*. 2023;389:1096-1107.
33. Rafii MS, Sperling RA, Donohue MC, et al. The AHEAD 3-45 study: Design of a prevention trial for Alzheimer's disease. *Alzheimer's Dement*. 2023;19:1227-1233.
34. Stark M, Wolfsgruber S, Kleineidam L, et al. Relevance of minor neuropsychological deficits in patients with subjective cognitive decline. *Neurology*. 2023;101:e2185-e2196.
35. Yushkevich PA, Pluta JB, Wang H, et al. Automated volumetry and regional thickness analysis of hippocampal subfields and medial temporal cortical structures in mild cognitive impairment. *Hum Brain Mapp*. 2015;36:258-287.
36. Xie L, Wisse LEM, Wang J, et al. Deep label fusion: A generalizable hybrid multi-atlas and deep convolutional neural network for medical image segmentation. *Med Image Anal*. 2023;83:102683.
37. Voevodskaya O, Simmons A, Nordenskjöld R, et al. The effects of intracranial volume adjustment approaches on multiple regional MRI volumes in healthy aging and Alzheimer's disease. *Front Aging Neurosci*. 2014;6:264.
38. 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*. 2006;31:968-980.
39. Vogel JW, Young AL, Oxtoby NP, et al. Four distinct trajectories of tau deposition identified in Alzheimer's disease. *Nat Med*. 2021;27:871-881.
40. Palmqvist S, Janelidze S, Quiroz YT, et al. Discriminative accuracy of plasma phospho-tau217 for Alzheimer disease vs other neurodegenerative disorders. *JAMA*. 2020;324:772-781.
41. Palmqvist S, Rossi M, Hall S, et al. Cognitive effects of Lewy body pathology in clinically unimpaired individuals. *Nat Med*. 2023;29:1971-1978.
42. Smith A. *Symbol digit modalities test*. Western Psychological Services; 1991.
43. Rosen WG, Mohs RC, Davis KL. A new rating scale for Alzheimer's disease. *Am J Psychiatry*. 1984;141:1356-1364.
44. Pichet Binette A, Franzmeier N, Spotorno N, et al. Amyloid-associated increases in soluble tau relate to tau aggregation rates and cognitive decline in early Alzheimer's disease. *Nat Commun*. 2022;13:6635.
45. Bates D, Mächler M, Bolker B, Walker S. Fitting linear mixed-effects models using lme4. *J Stat Softw*. 2015;67:1-48.
46. Venables WN, Ripley BD. *Modern applied statistics with S*. Springer; 2002. https://www.stats.ox.ac.uk/pub/MASS4/
47. Jessen F, Amariglio RE, van Boxtel M, et al. A conceptual framework for research on subjective cognitive decline in preclinical Alzheimer's disease. *Alzheimers Dement*. 2014;10:844-852.
48. Dong A, Honnorat N, Gaonkar B, Davatzikos C. CHIMERA: Clustering of heterogeneous disease effects via distribution matching of imaging patterns. *IEEE Trans Med Imaging*. 2016;35:612-621.
49. Noh Y, Jeon S, Lee JM, et al. Anatomical heterogeneity of Alzheimer disease. *Neurology*. 2014;83:1936-1944.
50. Oppedal K, Ferreira D, Cavallin L, et al. A signature pattern of cortical atrophy in dementia with Lewy bodies: A study on 333 patients from the European DLB consortium. *Alzheimers Dement*. 2019;15:400-409.
51. Risacher SL, Anderson WH, Charil A, et al. Alzheimer disease brain atrophy subtypes are associated with cognition and rate of decline. *Neurology*. 2017;89:2176-2186.
52. Shiino A, Watanabe T, Maeda K, Kotani E, Akiguchi I, Matsuda M. Four subgroups of Alzheimer's disease based on patterns of atrophy using VBM and a unique pattern for early onset disease. *Neuroimage*. 2006;33:17-26.
53. Shima K, Matsunari I, Samuraki M, et al. Posterior cingulate atrophy and metabolic decline in early stage Alzheimer's disease. *Neurobiol Aging*. 2012;33:2006-2017.
54. Persson K, Eldholm RS, Barca ML, et al. MRI-assessed atrophy subtypes in Alzheimer's disease and the cognitive reserve hypothesis. *PLoS One*. 2017;12:e0186595.
55. 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. *Sci Rep*. 2017;7:46263.
56. Poulakis K, Pereira JB, Muehlboeck JS, et al. Multi-cohort and longitudinal Bayesian clustering study of stage and subtype in Alzheimer's disease. *Nat Commun*. 2022;13:4566.

57. Yang Z, Wen J, Davatzikos C. Surreal-GAN:semi-supervised representation learning via GAN for uncovering heterogeneous disease-related imaging patterns. *arXiv*. https://doi.org/10.48550/arXiv.2205.04523
58. Whitwell JL, Graff-Radford J, Tosakulwong N, et al. [^{18}F]AV-1451 clustering of entorhinal and cortical uptake in Alzheimer's disease. *Ann Neurol*. 2018;83:248-257.
59. Levin F, Ferreira D, Lange C, et al. Data-driven FDG-PET subtypes of Alzheimer's disease-related neurodegeneration. *Alzheimers Res Ther*. 2021;13:49.
60. Yamada M, Itoh Y, Otomo E, Suematsu N, Matsushita M. Dementia of the Alzheimer type and related dementias in the aged: DAT subgroups and senile dementia of the neurofibrillary tangle type. *Neuropathology*. 1996;16:89-98.
61. Murray ME, Graff-Radford NR, Ross OA, Petersen RC, Duara R, Dickson DW. Neuropathologically defined subtypes of Alzheimer's disease with distinct clinical characteristics: A retrospective study. *Lancet Neurology*. 2011;10:785-796.
62. Jellinger KA. Neuropathological subtypes of Alzheimer's disease. *Acta Neuropathol*. 2012;123:153-154.
63. Whitwell JL, Dickson DW, Murray ME, et al. Neuroimaging correlates of pathologically defined subtypes of Alzheimer's disease: A case-control study. *Lancet Neurol*. 2012;11:868-877.
64. Habes M, Grothe MJ, Tunc B, McMillan C, Wolk DA, Davatzikos C. Disentangling heterogeneity in Alzheimer's disease and related dementias using data-driven methods. *Biol Psychiat*. 2020;88:70-82.
65. Samaroo A, Amariglio RE, Burnham S, et al. Diminished learning over repeated exposures (LORE) in preclinical Alzheimer's disease. *Alzheimers Dement (Amst)*. 2020;12:e12132.
66. Jutten RJ, Rentz DM, Fu JF, et al. Monthly at-home computerized cognitive testing to detect diminished practice effects in preclinical Alzheimer's disease. *Front Aging Neurosci*. 2022;13:800126.
67. Hassenstab J, Ruvolo D, Jasielec M, Xiong C, Grant E, Morris JC. Absence of practice effects in preclinical Alzheimer's disease. *Neuropsychology*. 2015;29:940-948.
68. Edmonds EC, Ard MC, Edland SD, Galasko DR, Salmon DP, Bondi MW. Unmasking the benefits of donepezil via psychometrically precise identification of mild cognitive impairment: A secondary analysis of the ADCS vitamin E and donepezil in MCI study. *Alzheimers Dement (N Y)*. 2018;4:11-18.
69. Graff-Radford J, Yong KXX, Apostolova LG, et al. New insights into atypical Alzheimer's disease in the era of biomarkers. *Lancet Neurol*. 2021;20:222-234.