doi:10.1093/brain/aww142 BRAIN 2016: 139; 2261–2274 | 2261

BRAIN
A JOURNAL OF NEUROLOGY

Biomarker clusters are differentially associated with longitudinal cognitive decline in late midlife

Annie M. Racine,[1,2,3] Rebecca L. Koscik,[4] Sara E. Berman,[1] Christopher R. Nicholas,[1,5] Lindsay R. Clark,[1,4,5] Ozioma C. Okonkwo,[1,4] Howard A. Rowley,[1,6] Sanjay Asthana,[1,5] Barbara B. Bendlin,[1,4] Kaj Blennow,[7,8] Henrik Zetterberg,[7,8,9] Carey E. Gleason,[1,5] Cynthia M. Carlsson[1,4,5] and Sterling C. Johnson[1,4,5,10]

The ability to detect preclinical Alzheimer's disease is of great importance, as this stage of the Alzheimer's continuum is believed to provide a key window for intervention and prevention. As Alzheimer's disease is characterized by multiple pathological changes, a biomarker panel reflecting co-occurring pathology will likely be most useful for early detection. Towards this end, 175 late middle-aged participants (mean age 55.9 ± 5.7 years at first cognitive assessment, 70% female) were recruited from two longitudinally followed cohorts to undergo magnetic resonance imaging and lumbar puncture. Cluster analysis was used to group individuals based on biomarkers of amyloid pathology (cerebrospinal fluid amyloid-β_{42}/amyloid-β_{40} assay levels), magnetic resonance imaging-derived measures of neurodegeneration/atrophy (cerebrospinal fluid-to-brain volume ratio, and hippocampal volume), neurofibrillary tangles (cerebrospinal fluid phosphorylated tau_{181} assay levels), and a brain-based marker of vascular risk (total white matter hyperintensity lesion volume). Four biomarker clusters emerged consistent with preclinical features of (i) Alzheimer's disease; (ii) mixed Alzheimer's disease and vascular aetiology; (iii) suspected non-Alzheimer's disease aetiology; and (iv) healthy ageing. Cognitive decline was then analysed between clusters using longitudinal assessments of episodic memory, semantic memory, executive function, and global cognitive function with linear mixed effects modelling. Cluster 1 exhibited a higher intercept and greater rates of decline on tests of episodic memory. Cluster 2 had a lower intercept on a test of semantic memory and both Cluster 2 and Cluster 3 had steeper rates of decline on a test of global cognition. Additional analyses on Cluster 3, which had the smallest hippocampal volume, suggest that its biomarker profile is more likely due to hippocampal vulnerability and not to detectable specific volume loss exceeding the rate of normal ageing. Our results demonstrate that pathology, as indicated by biomarkers, in a preclinical timeframe is related to patterns of longitudinal cognitive decline. Such biomarker patterns may be useful for identifying at-risk populations to recruit for clinical trials.

1 Alzheimer's Disease Research Center, University of Wisconsin School of Medicine and Public Health, USA, Madison, WI 53705, USA
2 Institute on Aging, University of Wisconsin-Madison, USA, Madison, WI 53706, USA
3 Neuroscience and Public Policy Program, University of Wisconsin-Madison, USA, Madison, WI 53705, USA
4 Wisconsin Alzheimer's Institute, University of Wisconsin School of Medicine and Public Health, USA, Madison, WI 53705, USA
5 Geriatric Research Education and Clinical Center, Wm. S. Middleton Veterans Hospital, USA, Madison WI 53705, USA
6 Department of Radiology, University of Wisconsin School of Medicine and Public Health, USA, Madison, WI 53705, USA
7 Department of Psychiatry and Neurochemistry, Institute of Neuroscience and Physiology, the Sahlgrenska Academy at the University of Gothenburg, Mölndal, Sweden
8 Clinical Neurochemistry Laboratory, Sahlgrenska University Hospital, Mölndal, Sweden
9 Institute of Neurology, University College London, London, UK
10 Waisman Laboratory for Brain Imaging and Behavior, University of Wisconsin-Madison, USA, Madison, WI 53705, USA

Received December 18, 2015. Revised April 29, 2016. Accepted May 5, 2016. Advance Access publication June 20, 2016

Correspondence to: Sterling C. Johnson, Ph.D.,
William S. Middleton Memorial VA Hospital,
2500 Overlook Terrace (11G),
GRECC, Madison,
WI, 53705, USA
E-mail: scj@medicine.wisc.edu

Keywords: preclinical Alzheimer's disease; cluster analysis; neuroimaging; cerebrospinal fluid; longitudinal

Abbreviations: ASCVD = Atherosclerotic Cardiovascular Disease; MMSE = Mini-Mental State Examination; WADRC = Wisconsin Alzheimer's Disease Research Center; WRAP = Wisconsin Registry for Alzheimer's Prevention

Introduction

Alzheimer's disease pathology begins decades before clinical symptoms emerge (Morris, 2005; Sperling *et al.*, 2011; Jansen *et al.*, 2015; Dubois *et al.*, 2016) and cognitive impairment is a continuous process that begins several years before a diagnosis of mild cognitive impairment or dementia due to Alzheimer's disease (Howieson *et al.*, 2008; Villemagne *et al.*, 2013; Rajan *et al.*, 2015). This long preclinical stage provides a critical window for intervention with disease-modifying pharmaceutical or behavioural therapies. However, a reliable way of detecting preclinical Alzheimer's disease in late midlife has yet to be determined. Cohorts in this age range recruited by risk factors such as the apolipoprotein E (*APOE*) gene and parental family history may be enriched with people, yet unidentified, who have or will have preclinical Alzheimer's disease. As Alzheimer's disease is characterized by multiple but distinct pathological changes, biomarker panels that reflect disease-specific and co-occurring pathology will likely be most useful for early detection. Although considerable research has been dedicated to identifying and validating neuroimaging, fluid, and other biomarkers for various stages of Alzheimer's disease, the individual and combined power of these biomarkers to detect preclinical Alzheimer's disease is, as of now, still not clearly established.

Cluster analysis is a potentially powerful statistical tool to apply to Alzheimer's disease biomarkers that can be detected during the preclinical stage. Cluster analysis can group individuals based on heterogeneity within a single biomarker, or an array of biomarkers; therefore, multiple co-occurring pathological features can simultaneously be captured in a single clustering, making it a particularly promising approach for detecting groups of individuals who may be on different trajectories relevant to Alzheimer's disease. Biomarker-based cluster analysis has been used in cognitively normal elderly subjects and patients with mild cognitive impairment or dementia due to Alzheimer's disease, but it has yet to be thoroughly investigated for use in identifying preclinical Alzheimer's disease in midlife (Iqbal *et al.*, 2005; van der Vlies *et al.*, 2009; Nettiksimmons *et al.*, 2010, 2013, 2014; Wallin *et al.*, 2010; Escudero *et al.*, 2011; Pike *et al.*, 2011; Vemuri *et al.*, 2011; Skillback *et al.*, 2013; Noh *et al.*, 2014). As noted by Jack *et al.* (2013) and others, there is a paucity of empirical data regarding the important preclinical timeframe. The primary goal of this study, therefore, was to use an integrated multidimensional approach to detect preclinical heterogeneity in a late middle-aged cohort. In addition to markers of amyloid pathology (amyloid-β_{42}/amyloid-β_{40} levels in CSF) and MRI-derived measures of neurodegeneration/atrophy (CSF-to-brain volume ratio and hippocampal volume), we additionally include a CSF marker of neurofibrillary tangles (phosphorylated tau$_{181}$ levels) and a MRI-based marker of cumulative vascular risk, total white matter hyperintensity lesion volume. Combining an unsupervised learning approach, multimodal biomarkers, and longitudinally measured cognitive data provides a unique opportunity to investigate preclinical conditions as they emerge from normal ageing. Our first hypothesis was that in addition to a healthy ageing group (biomarker negative), a preclinical Alzheimer's disease-like group would be identified based on an integrative biomarker profile (amyloidosis + neurofibrillary tangle pathology + neurodegeneration). Additionally, several studies have found a high prevalence of mixed dementia due to Alzheimer's disease and vascular dementia, with perhaps as many as half of persons clinically diagnosed with dementia due to Alzheimer's disease showing signs of vascular pathology at autopsy (Jellinger, 2007; Schneider *et al.*, 2007). Therefore, we further hypothesized that a group of participants would present with evidence suggestive of simultaneous Alzheimer's disease and vascular pathology (amyloidosis + neurodegeneration + white matter hyperintensities). Our second hypothesis was that these groups would show evidence of early cognitive decline consistent with their biomarker profiles, with greater decline observed in the Alzheimer's disease-like and mixed dementia-like clusters compared to the healthy ageing cluster.

Materials and methods

Participants

Participants were selected for this study on the basis of participation in one of two large observational and longitudinally followed cohorts at the University of Wisconsin-Madison: the Wisconsin Registry for Alzheimer's Prevention (WRAP) and

the Wisconsin Alzheimer's Disease Research Center (WADRC), having available biomarkers from MRI and CSF, and being cognitively normal at baseline. Of the 181 participants initially selected by these criteria, six were excluded based on an incomplete MRI sequence ($n = 1$), incomplete CSF assays ($n = 1$), having an interval from the MRI to lumbar puncture greater than 1 year ($n = 2$, 374 and 780 days, interval range without these participants is −112 to 117 days), or white matter hyperintensity outlier status [$n = 2$, >5 standard deviations (SD) from mean] resulting in a sample of $n = 175$ ($n = 92$ WRAP, $n = 83$ WADRC).

Both WRAP and WADRC are designed to identify biological and lifestyle risk factors associated with development of subsequent clinical Alzheimer's disease in cohorts enriched for Alzheimer's disease risk factors due to parental family history of Alzheimer's disease (Sager *et al.*, 2005; Jonaitis *et al.*, 2013; Koscik *et al.*, 2014). The WRAP study consists of 1545 participants (mean age = 53.6 ± 6.6 years at first cognitive assessment), of which 72.4% have a parental family history of dementia due to Alzheimer's disease. Recruitment for the WADRC cohort is ongoing. For the 300+ WADRC participants currently being followed, mean age at enrolment was ∼57 years, and approximately two-thirds of the cohort have a parental family history of dementia due to Alzheimer's disease.

The University of Wisconsin Institutional Review Board approved all study procedures, each subject provided signed informed consent before participation, and all research was completed in accordance with the Declaration of Helsinki.

MRI collection and calculation of neuroimaging variables

All participants were scanned on a GE 3.0 T MR750 using an 8-channel head coil. T_1-weighted, T_2-weighted, and fluid-attenuated inversion recovery (FLAIR) anatomical scans were acquired as described previously (Johnson *et al.*, 2014; Racine *et al.*, 2014; Berman *et al.*, 2015). T_2-weighted and FLAIR anatomical scans were reviewed by a neuroradiologist (H.A.R.) for exclusionary abnormalities. The T_1-weighted volume was segmented into tissue classes (CSF, grey matter, and white matter) using the segmentation tool in SPM12 (www.fil.ion.ucl.ac.uk/spm). CSF-to-brain volume ratio was calculated as the tissue volume ratio of CSF/(grey matter + white matter). Hippocampal volume was calculated using FSL-FIRST, a model-based segmentation/registration tool (Patenaude *et al.*, 2011) and corrected for intracranial volume calculated in SPM12. Total white matter hyperintensity lesion volume was measured using the SPM Lesion Segmentation Tool (Schmidt *et al.*, 2012).

CSF collection and quantification

CSF was collected as described previously (Almeida *et al.*, 2015; Starks *et al.*, 2015). CSF collection and processing methods were identical across the WRAP and WADRC studies. Phosphorylated tau_{181} was quantified with a sandwich ELISA [Phospho-Tau(181P), Fujirebio Europe]. For the amyloid-β_{42}/amyloid-β_{40} ratio, CSF levels of amyloid-β_{42} and amyloid-β_{40} (a less amyloidogenic amyloid-β fragment as compared to amyloid-β_{42}) were quantified by electrochemiluminescence using an amyloid-β triplex assay (MSD Human Aβ peptide Ultra-Sensitive Kit, Meso Scale Discovery).

Cognitive data collection

Longitudinal cognitive data are collected for both the WRAP and WADRC studies. WRAP study participants undergo follow-up cognitive testing ∼4 years post their initial visit and every 2 years thereafter. WADRC participants undergo annual or biannual cognitive testing. At each session of testing, participants complete a comprehensive neuropsychological battery consisting of measures that span traditional cognitive domains of memory, attention, executive function, language, and visuospatial ability.

We selected tests that were consistent across both studies that assessed domains of episodic memory (total trials 1–5 and delayed recall from the Rey Auditory Verbal Learning Test; immediate and delayed recall from the Wechsler Memory Scale-Revised Logical Memory), semantic memory (Boston Naming Test, Animal Naming), aspects of executive function (Trail Making Test Part B; Wechsler Adult Intelligence Scale-Revised Digit Symbol), and global cognitive function (Mini-Mental State Examination, MMSE). Table 1 summarizes the longitudinal cognitive data for each test across the five sessions. Rey Auditory Verbal Learning Test, Boston Naming, and Trail Making Test Part B started at Session 1 for both WRAP and WADRC; Logical Memory, Digit Symbol, and MMSE started at Session 2 for WRAP and session 1 for WADRC; and Animal Naming started at Session 3 for WRAP and Session 1 for WADRC. Therefore, for analyses and in Table 1 for WRAP participants only, neuropsychological testing 2 and 3 are called neuropsychological testing 1 for tests initiated at Sessions 2 and 3, respectively. WADRC participants received only story A for Logical Memory and a 30-item test for Boston Naming; their scores were doubled to maintain continuity between cohorts. Additionally, the protocol for the neuropsychological battery underwent revisions during the course of data collection for WADRC subjects. As a result, some WADRC subjects were not administered Logical Memory, Animal Naming, Digit Symbol, and MMSE at the second neuropsychological visit. The structurally missing data do not affect the validity of the statistical tests performed, as linear mixed effects models are robust to unbalanced designs, missing data, and unequally spaced data points.

Clustering

Our first hypothesis was that we could use biomarker data to identify groups of participants whose biomarker profiles were consistent with profiles associated with healthy ageing, Alzheimer's pathology, and mixed Alzheimer's and vascular pathology. To examine this hypothesis, we first used cluster analysis to identify groups as described in this section and then we compared participant characteristics across the identified clusters as described below. Clustering broadly refers to the process of grouping a set of individuals into clusters based on how similar they are on certain criteria. Cluster analysis should classify individuals so that individuals within a cluster are as similar to each other as possible and clusters are as different from each other as possible.

Table 1 Summary of longitudinal cognitive data

NP test	NP testing 1	NP testing 2	NP testing 3	NP testing 4	NP testing 5
Episodic memory					
RAVLT total trials 1–5	*n* = 175 50.98 (8.15) 32–71	*n* = 170 52.76 (8.99) 28–71	*n* = 166 52.40 (8.91) 24–72	*n* = 113 53.32 (8.46) 37–71	*n* = 39 53.00 (8.89) 37–71
RAVLT delay	*n* = 174 10.30 (2.86) 2–15	*n* = 170 10.94 (2.84) 0–15	*n* = 166 10.77 (2.90) 0–15	*n* = 113 11.00 (2.49) 4–15	*n* = 39 10.49 (3.03) 4–15
Logical Memory immediate[a,b]	*n* = 174 29.17 (6.47) 14–46	*n* = 110 28.77 (5.76) 13–42	*n* = 146 29.52 (6.7) 10–42	*n* = 67 31.76 (5.64) 18–44	*n* = 14 31.86 (7.82) 18–46
Logical Memory delay[a,b]	*n* = 174 26.48 (7.27) 1–46	*n* = 110 26.00 (6.93) 0–43	*n* = 146 27.77 (6.77) 4–42	*n* = 67 30.03 (6.56) 18–48	*n* = 14 31.29 (7.87) 18–46
Semantic memory					
Animal Naming[a]	*n* = 151 23.19 (5.60) 9–41	*n* = 89 23.06 (5.39) 12–34	*n* = 100 23.36 (4.99) 13–37	*n* = 42 25.33 (4.49) 16–34	*n* = 14 24.21 (4.54) 15–32
Boston Naming[b]	*n* = 173 57.20 (2.69) 46–60	*n* = 170 57.30 (2.60) 47–60	*n* = 164 57.70 (2.48) 30–60	*n* = 113 57.91 (2.33) 44–60	*n* = 39 57.54 (2.32) 52–60
Executive function					
Trail Making Test Part B	*n* = 174 59.44 (19.62) 23–138	*n* = 170 56.56 (18.46) 14–144	*n* = 166 55.79 (18.86) 26–125	*n* = 113 60.66 (24.43) 23–182	*n* = 39 62.15 (23.27) 38–152
Digit Symbol[a]	*n* = 174 58.57 (10.46) 5–82	*n* = 109 58.24 (10.17) 33–83	*n* = 145 58.34 (11.14) 31–89	*n* = 67 57.79 (11.23) 33–89	*n* = 14 62.71 (9.37) 49–86
Global function					
MMSE[a]	*n* = 174 29.44 (0.77) 27–30	*n* = 109 29.28 (1.13) 24–30	*n* = 146 29.44 (0.86) 27–30	*n* = 67 29.25 (1.11) 25–30	*n* = 14 29.21 (0.98) 27–30

Values are given as *n*, mean (SD), range. One participant had a sixth neuropsychological visit (data not shown).

[a]For WRAP participants only, neuropsychological (NP) testing 1 is Session 2 for Logical Memory, Digit Symbol, and MMSE, and NP testing 1 is Session 3 for Animal Naming; same neuropsychological tests were not administered to some of the WADRC participants at NP testing 2.

[b]Test scores were scaled as appropriate to maintain consistency between WRAP and WADRC.

RAVLT = Rey Auditory Verbal Learning Test. Mean baseline age for RAVLT, Trail Making Test Part B, and Boston Naming is 55.9 ± 5.7 years; 58.2 ± 5.6 years for Logical Memory, Digit Symbol, and MMSE; and 59.3 ± 5.8 years for Animal Naming.

Variables included in the cluster analysis were CSF amyloid-β_{42}/amyloid-β_{40}, CSF phosphorylated tau, MRI-derived CSF-to-brain volume ratio, intracranial volume-corrected hippocampal volume, and white matter hyperintensity total lesion volume. MRI data were also collected longitudinally in WRAP and WADRC, therefore, the three neuroimaging variables were selected from the MRI scan collected closest to the lumbar puncture (mean = 2.62 ± 20.56 days, range = −112 to 117). Because these variables are known to change with age, all variables were corrected for age by saving the unstandardized residual from a linear regression for each clustering variable with only age at lumbar puncture in the model. Hippocampal volume was additionally corrected for intracranial volume. The unstandardized residuals were then transformed into z-scores before being entered into the clustering algorithm. To ensure that there was not a high degree of collinearity among the clustering variables, which could lead to specific aspects being overrepresented in the clustering solution, bivariate correlations were conducted on all clustering variables using both Pearson's correlation coefficient and Spearman's rho to test for parametric and non-parametric relationships. A threshold of 0.7 was chosen to signify high collinearity (Dormann *et al.*, 2013).

Participants were grouped using agglomerative hierarchical clustering in Statistical Package for the Social Sciences (SPSS) 22 with Ward's method of minimum variance and the squared Euclidean distance metric (Ward, 1963; Clatworthy *et al.*, 2005; Murtagh and Legendre, 2014). Ward's method joins two clusters to produce the smallest increases in the pooled within-cluster variation. The Euclidean distance is one of the most common and straightforward ways of computing distance between objects and refers to the geometric distance in multidimensional space. Squaring the Euclidean distance is a common approach that places greater emphasis on objects that are further apart (Clatworthy *et al.*, 2005) and is the preferred metric when using Ward's method in SPSS (Murtagh and Legendre, 2014). The number of resulting clusters was determined by observing where an inconsistent increase occurred in the dissimilarity measure (squared Euclidean distance) on the dendrogram (Supplementary Fig. 1A) and stopping the

clustering process one stage prior (Clatworthy *et al.*, 2005). This was further quantified using the stopping rule as described by Hair *et al.* (2013). With Ward's method the coefficients in the agglomeration schedule refer to the within-cluster sum of squares, whereby an increase in the coefficient indicates an increase in heterogeneity. Because the goal of clustering is to identify relatively homogenous groups within a heterogeneous sample, an increase in heterogeneity within a cluster is considered undesirable. We calculated the percentage change in heterogeneity between subsequent clustering steps as (agglomeration coefficient at stage j − agglomeration coefficient at stage j − 1)/agglomeration coefficient at stage j − 1. While comparable and smaller increases in heterogeneity are observed up until the last few stages (<8%), a disproportionately larger increase is observed moving from four clusters to three clusters (13.35%). By the stopping rule, the agglomeration should stop at the stage prior to the disproportionate increase, which, here, would favour a four-cluster solution. This is visualized on the scree plot of the agglomeration coefficients (Supplementary Fig. 1B). However, both a three-cluster and four-cluster solution would satisfy our criteria for adequate sample sizes and inconsistent increases in dissimilarity, as visualized on the dendrogram as well as by the increase in the agglomeration coefficient; the four-cluster solution was selected because it allowed for a slightly more comprehensive examination of biomarker groups than simply a three-cluster solution, and because moving from the four- to three-cluster solution indicated the first substantial increase in heterogeneity according to the agglomeration schedule.

Statistical analyses

ANOVA: cluster characteristics

Characteristics of clusters were compared by ANOVA (for continuous variables) or chi-square (for categorical variables) with SPSS on the means of the five clustering variables, demographics, and clinical characteristics that may be related to dementia due to Alzheimer's disease or vascular-mediated cognitive decline. ANOVA on clustering variables was performed with raw values and no covariates (with the exception of hippocampal volume for which intracranial volume was included) as validation that the clustering solution worked and to facilitate *post hoc* characterization of the relative means of the cluster variables. Demographic variables were assessed next and included age at lumbar puncture, sex, years of education, parental family history, whether a participant carries at least one copy of the *APOE* ε4 allele (*APOE4*), cognitive follow-up interval for tests initiated at Session 1 for both WRAP and WADRC (years), baseline neuropsychological testing age, and cohort composition (whether the participant was selected from WRAP or WADRC). Clinical characteristics were also assessed including blood pressure (systolic and diastolic), pulse pressure (systolic minus diastolic), total cholesterol, non-high density lipoprotein (HDL) cholesterol (total cholesterol minus HDL), fasting glucose (mg/dl), body mass index (kg/m^2), waist-to-hip ratio, and atherosclerotic cardiovascular disease (ASCVD) risk score, a formula developed by the American College of Cardiology and the American Heart Association to estimate risk of an ASCVD event in the next 10 years (Goff *et al.*, 2014). ASCVD risk score was highly skewed, so it was log-transformed to improve normality. Because clinical characteristics are known to vary by age and sex, these were included as covariates for analyses of clinical data, in addition to the interval (months) between the lumbar puncture and the medical examination/lab draw date. One participant had missing data for ASCVD, non-HDL and HDL cholesterol, fasting glucose, and waist-to-hip ratio. Due to outliers compared to the entire sample (>3 SD from the mean) for some clinical variables, ANCOVA results are reported with and without outliers removed. Results from analyses comparing cluster characteristics are evaluated against a threshold of $P < 0.05$. In cases where the omnibus test for group differences was significant, *post hoc* pairwise tests using Fisher's least significant difference were performed for each cluster pair. When the chi-square was significant, *post hoc* analyses tested whether each cluster was significantly different compared to the other three clusters combined.

Linear mixed effects: cognitive decline

To test our second hypothesis that the biomarker-based Alzheimer's-like and mixed aetiology groups would exhibit greater cognitive decline than the healthy ageing group, we used linear mixed effects models in SPSS for each neuropsychological test of interest. Cognitive decline was measured by slope from the linear mixed effects models with cluster groups as the independent variable and longitudinally measured neuropsychological scores as the dependent variables. For each neuropsychological test, we first ran unconditional growth models adjusting for random effects of intercept and time-associated slope to determine significant random effects. We explored multiple covariance structures (e.g. unstructured and variance components), and selected unstructured because it allows covariances to vary and had a similar model fit measured by Akaike's Information Criteria. Next, conditional models were run with cluster as the grouping variable, which included significant random effects plus fixed effects of sex, years of education, *APOE4*, interval from the closest MRI to the lumbar puncture (days), interval from the lumbar puncture to the first neuropsychological assessment (months), cohort (binary: WRAP or WADRC), mean-centred baseline age (mean age at entry into the cohort 55.92 years), time (interval in years from the first neuropsychological testing visit), cluster, and the interaction of time × cluster. The interaction term tested the hypothesis that rates of cognitive decline differed between clusters.

Results are evaluated at a threshold of $P < 0.05$. In cases where the omnibus test for group differences was significant, *post hoc* tests were performed using Cluster 4, which was in the average range on all biomarkers (biomarker-negative; Fig. 1), as the comparison group. In cases where one of the three biomarker-positive clusters was significantly different from Cluster 4, additional *post hoc* comparisons were performed comparing that cluster to the other two clusters. The statistics from linear mixed effects models do not have exact F distributions so degrees of freedom are estimated by a Satterthwaite approximation and are reported rounded to the nearest integer.

Results

Clustering variables

There were no relationships between the variables included in the cluster analysis that exceeded a Pearson's correlation coefficient or Spearman's rho > 0.3. Four clusters emerged (Fig. 1 and Supplementary Fig. 2). The first cluster ($n = 22$) demonstrated elevated CSF phosphorylated tau and low CSF amyloid-β_{42}/amyloid-β_{40}, consistent with an Alzheimer's disease-like profile. The second cluster ($n = 32$) showed evidence of elevated CSF-to-brain volume ratio and white matter hyperintensities and low CSF amyloid-β_{42}/amyloid-β_{40}, suggestive of vascular/mixed pre-dementia and/or amyloid angiopathy. The third cluster ($n = 45$) had elevated CSF-to-brain volume ratio and the smallest hippocampal volume, suggestive of an atrophic pattern. The fourth and largest cluster ($n = 76$) was in the average range for all five biomarkers, suggestive of healthy ageing.

ANOVA supported the visual characterization of the clusters. Cluster 1 had significantly elevated phosphorylated tau compared to all other clusters ($P < 0.001$) and Cluster 2 had significantly lower phosphorylated tau compared to all other clusters ($P < 0.05$). Cluster 1 also had the lowest CSF amyloid-β_{42}/amyloid-β_{40} compared to all other clusters ($P < 0.001$), and Cluster 2 had significantly lower amyloid-β_{42}/amyloid-β_{40} compared to Clusters 3 and 4 ($P < 0.001$). Both Clusters 2 and 3 had significantly greater CSF-to-brain volume ratio compared to Cluster 4 ($P < 0.001$). Cluster 3 had significantly smaller hippocampal volume compared to all other clusters ($P < 0.05$), and Cluster 4 had significantly greater hippocampal volume compared to all other clusters ($P < 0.05$). Cluster 2 had significantly greater white matter hyperintensity total lesion volume compared to all other clusters ($P < 0.001$).

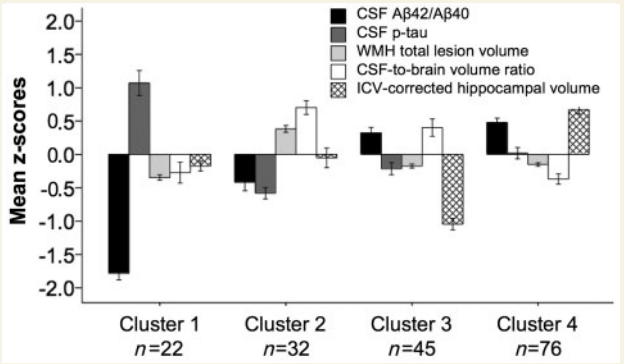

Figure 1 Results of hierarchical cluster analysis with Alzheimer's disease biomarkers. *Y*-axis: Mean z-scores of the five age-corrected clustering variables: CSF amyloid-β_{42}/amyloid-β_{40} (black), CSF phosphorylated tau (dark grey), white matter hyperintensity total lesion volume (light grey), CSF-to-brain volume ratio (white), and intracranial volume-corrected hippocampal volume (cross-hatched). *X*-axis: Cluster. Error bars represent 95% confidence intervals. P-tau = phosphorylated tau_{181}; WMH = white matter hyperintensity; ICV = intracranial volume.

Demographics and clinical characteristics

Demographic and clinical characteristics by cluster are summarized in Table 2. Cluster 1 was significantly older than Clusters 2 ($P = 0.004$) and 4 ($P = 0.027$) and had significantly more *APOE4* carriers than all other clusters combined ($\chi^2 = 7.60$, $P = 0.006$). Clusters differed on the composition of WRAP and WADRC participants such that Cluster 1 had significantly more participants from WRAP than WADRC compared to all other clusters combined ($\chi^2 = 6.16$, $P = 0.013$) and Cluster 4 had more participants from WADRC than WRAP ($\chi^2 = 7.48$, $P = 0.006$). Clusters did not differ on sex, years of education, parental family history, neuropsychological follow-up, or baseline neuropsychological testing age.

ANOVA or ANCOVA omnibus tests were not significant for any clinical characteristics examined when the entire sample was included (Table 2). However, using a threshold of >3 SD from the overall mean, nine participants met criteria for outlier status ($n = 5$ for fasting glucose; $n = 1$ for body mass index; $n = 1$ for waist-to-hip ratio, and $n = 2$ for HDL cholesterol). When these outliers were removed from the ANCOVA for their respective clinical variable, only one result changed. The outlier for body mass index was from Cluster 3 and when this participant was removed, the omnibus ANCOVA was significant [$P = 0.035$, $F(3,167) = 2.9$] (Supplementary Table 1). *Post hoc* tests showed both clusters 2 and 4 had greater body mass index compared to clusters 1 (*P*-values < 0.05) and 3 (*P*-values < 0.05) but Clusters 2 and 4 and Clusters 1 and 3 did not significantly differ from one another.

Cognitive decline

Results from the linear mixed effects models of cognitive change over time by cluster are displayed in Fig. 2.

Female sex was associated with better performance on Rey Auditory Verbal Learning Test total trials 1–5 [$P < 0.001$, $\beta = 7.62$, $t(169) = 7.384$] and delayed [$P < 0.001$, $\beta = 2.08$, $t(168) = 5.672$], Logical Memory immediate [$P = 0.014$, $\beta = 2.14$, $t(168) = 2.478$] and delayed [$P = 0.016$, $\beta = 2.39$, $t(167) = 2.433$], and MMSE [$P = 0.043$, $\beta = 0.22$, $t(164) = 2.035$]. More years of education were associated with better performance on Rey Auditory Verbal Learning Test total trials 1–5 [$P = 0.034$, $\beta = 0.41$, $t(168) = 2.141$], Logical Memory immediate [$P = 0.001$, $\beta = 0.55$, $t(168) = 3.471$] and delayed [$P = 0.003$, $\beta = 0.55$, $t(167) = 3.034$], Boston Naming [$P = 0.002$, $\beta = 0.20$, $t(175) = 3.119$], Digit Symbol [$P = 0.006$, $\beta = 0.74$, $t(169) = 2.779$], and Trail Making Test B [$P = 0.024$, $\beta = -1.07$, $t(175) = -2.282$]. Older baseline age was associated with worse performance on Rey

Table 2 Cluster comparisons on biomarker, demographic, and clinical variables

Comparison variable	Cluster 1 *n* = 22	Cluster 2 *n* = 32	Cluster 3 *n* = 45	Cluster 4 *n* = 76	*F*(df)	*P*-value
Clustering variable						
CSF amyloid-β_{42}/amyloid-β_{40}	0.065 (0.02)	0.094 (0.02)	0.107 (0.01)	0.110 (0.01)	64.3 (3, 171)	<0.001
CSF phosphorylated tau	58.23 (15.3)	32.94 (9.8)	39.13 (10.7)	42.01 (13.1)	19.9 (3, 171)	<0.001
CSF-to-brain volume ratio	0.291 (0.06)	0.319 (0.05)	0.314 (0.08)	0.269 (0.05)	8.2 (3, 171)	<0.001
Hippocampal volume[a]	7646 (611)	7959 (873)	7134 (634)	8349 (732)	45.3 (3, 171)	<0.001
WMH total lesion volume	13.58 (5.8)	25.76 (11.0)	14.61 (6.4)	15.35 (6.7)	18.8 (3, 171)	<0.001
Demographic variable						
Age at lumbar puncture (years)	61.77 (4.9)	57.22 (7.3)	59.46 (5.4)	58.68 (5.3)	2.9 (3, 170)	0.037
Sex (% female)[b]	72.7	68.8	62.2	73.7	1.9 (3)	0.599
Education (years)	17.05 (1.9)	16.09 (2.7)	16.18 (2.8)	16.24 (2.4)	0.8 (3, 170)	0.499
APOE4 carrier (%)[b]	68.2	50.0	35.6	32.9	10.4 (3)	0.015
Parental family history (%)[b]	68.2	81.3	77.3	84.2	3.0 (3)	0.394
Neuropsychological follow-up (years)[d]	7.58 (3.1)	6.36 (3.8)	5.84 (3.2)	5.43 (3.6)	2.3 (3, 170)	0.080
Baseline neuropsychological age (years)[d]	57.2 (4.6)	53.8 (7.0)	56.5 (5.5)	56.1 (5.4)	2.11 (3, 171)	0.101
Cohort (WRAP/WADRC)[b]	17/5	20/12	24/21	31/45	10.9 (3)	0.012
Clinical variable[c]						
ASCVD risk[e]	6.39 (5.1)	5.70 (6.1)	6.47 (6.2)	4.93 (4.3)	0.6 (3, 167)	0.608
Systolic blood pressure (mmHg)	121.73 (16.3)	126.88 (11.3)	123.71 (13.7)	124.16 (15.8)	1.3 (3, 168)	0.274
Diastolic blood pressure (mmHg)	71.05 (8.5)	76.31 (8.5)	73.98 (8.9)	72.93 (8.8)	1.9 (3, 168)	0.138
Pulse pressure	50.68 (13.6)	50.56 (10.1)	49.73 (13.0)	51.22 (11.9)	0.3 (3, 168)	0.806
Total cholesterol (mg/dl)	204.09 (39.0)	195.47 (34.4)	196.73 (39.0)	201.05 (32.9)	0.3 (3, 168)	0.822
Non-HDL cholesterol[e] (mg/dl)	143.05 (29.4)	138.32 (32.2)	135.44 (35.9)	139.09 (32.1)	0.2 (3, 167)	0.881
HDL cholesterol[e]	61.05 (21.7)	57.87 (15.5)	61.29 (18.6)	61.96 (15.8)	0.7 (3, 167)	0.557
Fasting glucose (mg/dl)[e]	95.05 (8.6)	98.03 (22.8)	97.18 (14.9)	93.01 (12.0)	1.3 (3, 167)	0.287
Body mass index (kg/m^2)	27.03 (4.6)	29.06 (5.5)	27.52 (5.3)	28.88 (5.2)	2.1 (3, 168)	0.108
Waist-to-hip ratio[e]	0.87 (0.14)	0.87 (0.11)	0.88 (0.09)	0.85 (0.09)	0.7 (3, 167)	0.539

Values are mean (SD) unless otherwise indicated. *P*-value is for omnibus ANOVA unless otherwise indicated.
[a]*P*-value is for ANCOVA controlled for intracranial volume.
[b]Effect size and *P*-value is for chi-square test.
[c]*P*-value is for ANCOVA controlled for age at lab visit, sex, and interval between lumbar puncture and lab visit.
[d]Neuropsychological follow-up interval and baseline age are provided for tests administered at session one for both WRAP and ADRC.
[e]One participant in Cluster 2 is missing data for indicated clinical variables.
WMH = white matter hyperintensity. ASCVD risk = estimate risk (%) of an atherosclerotic cardiovascular disease event in the next 10 years; ANCOVA for ASCVD risk was conducted on log transformed data because raw percentages were not normally distributed; means reported for each cluster are raw risk values. Please refer to 'ANOVA: cluster characteristics' section for details of these models.

Auditory Verbal Learning Test total trials 1–5 [$P < 0.001$, $\beta = -0.36$, $t(170) = -4.227$] and delayed [$P = 0.004$, $\beta = -0.09$, $t(169) = -2.888$], Animal Naming [$P = 0.015$, $\beta = -0.15$, $t(174) = -2.456$], Digit Symbol [$P < 0.001$, $\beta = -0.81$, $t(171) = -6.843$], Trail Making Test B [$P < 0.001$, $\beta = 1.34$, $t(178) = 6.37$], and MMSE [$P = 0.007$, $\beta = -0.02$, $t(166) = -2.728$].

Omnibus tests revealed differences in intercepts between clusters on Rey Auditory Verbal Learning Test total trials 1–5 [$P = 0.030$, $F(254) = 3.023$], Logical Memory delayed [$P = 0.009$, $F(249) = 3.962$], and Boston Naming [$P = 0.046$, $F(251) = 2.708$]. Cluster 1 had a larger intercept on Rey Auditory Verbal Learning Test total trials 1–5 compared to Cluster 4 [$P = 0.003$, $\beta = 5.24$, $t(253) = 2.969$], Cluster 2 [$P = 0.037$, $\beta = 4.09$, $t(261) = 2.098$], and Cluster 3 [$P = 0.012$, $\beta = 4.71$, $t(259) = 2.54$]. Cluster 1 also had a larger intercept on Logical Memory delayed compared to Cluster 4 [$P = 0.010$, $\beta = 4.28$, $t(243) = 2.597$], Cluster 2 [$P = 0.001$, $\beta = 6.21$, $t(249) = 3.405$], and Cluster 3 [$P = 0.011$, $\beta = 4.45$, $t(247) = 2.573$]. Cluster 2 had a smaller intercept on Boston Naming compared to Cluster 4 [$P = 0.031$, $\beta = -0.105$, $t(248) = -2.171$] and Cluster 1 [$P = 0.008$, $\beta = -1.71$, $t(256) = -2.676$] but not Cluster 3 [$P = 0.067$, $\beta = -0.96$, $t(257) = -1.829$].

The omnibus tests for group differences in slope (i.e. clusters × time) were significant for Rey Auditory Verbal Learning Test total trials 1–5 [$P = 0.042$, $F(562) = 2.746$], Logical Memory immediate [$P = 0.046$, $F(369) = 2.684$] and delayed [$P = 0.001$, $F(364) = 5.241$], and MMSE [$P = 0.009$, $F(409) = 3.911$]. Cluster 1 had a steeper slope (i.e. faster cognitive decline) on Rey Auditory Verbal Learning Test total trials 1–5 compared to Cluster 4 [$P = 0.005$, $\beta = -0.58$, $t(551) = -2.810$] and Cluster 3 [$P = 0.026$, $\beta = -0.50$, $t(550) = -2.237$] but not Cluster 2 [$P = 0.067$, $\beta = -0.42$ $t(552) = -1.838$]. Cluster 1 had a steeper slope on Logical Memory immediate compared to Cluster 4 [$P = 0.024$, $\beta = -0.67$, $t(367) = -2.273$] and

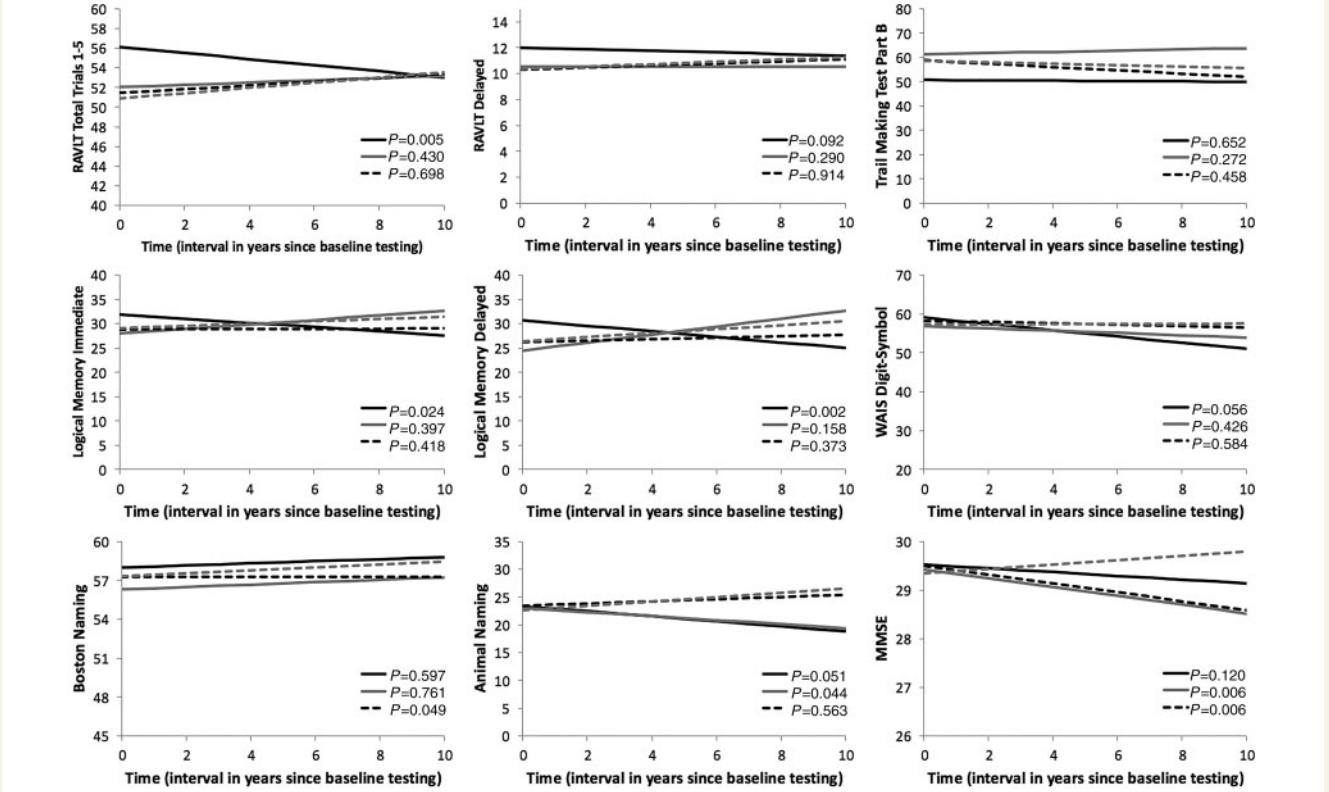

Figure 2 Cognitive decline over time by cluster. Cognitive change over time (interval since first neuropsychological testing visit) corrected for all other variables in the model for Cluster 1 (black solid, preclinical Alzheimer's disease-like), Cluster 2 (grey solid, preclinical mixed aetiology), Cluster 3 (black dashed, suspected non-Alzheimer's aetiology), and Cluster 4 (grey dashed, healthy ageing). The plots were extrapolated based on calculation of group intercepts and slopes using the regression coefficients from the output of estimated fixed effects from the linear mixed effects models. *P*-values are for the comparison of slopes between Cluster 4 and one of the other clusters, indicated by its corresponding colour. *Y*-axis: Raw cognitive test scores. *X*-axis: Time (interval in years since first neuropsychological testing visit). Lower scores indicate worse performance on all tests except Trail Making Test B. RAVLT = Rey Auditory Verbal Learning Test; WAIS = Wechsler Adult Intelligence Scale.

Cluster 2 [$P = 0.008$, $\beta = -0.90$, $t(370) = -2.684$] but not Cluster 3 [$P = 0.175$, $\beta = -0.45$, $t(365) = -1.359$]. Cluster 1 had a steeper slope on Logical Memory delayed compared to Cluster 4 [$P = 0.002$, $\beta = -0.98$, $t(363) = -3.062$], Cluster 2 [$P < 0.001$, $\beta = -1.39$, $t(365) = -3.835$], and Cluster 3 [$P = 0.047$, $\beta = -0.72$, $t(361) = -1.992$]. Cluster 2 [$P = 0.006$, $\beta = -0.14$, $t(419) = -2.772$] and Cluster 3 [$P = 0.006$, $\beta = -0.14$, $t(404) = -2.780$] both had steeper slopes on MMSE compared to Cluster 4, but Clusters 2 and 3 were not different from each other or from Cluster 1 (P-values > 0.3, β-values < 0.1, t-values < 1).

The covariate for interval from the closest MRI to the lumbar puncture was highly and significantly ($P < 0.001$) correlated with time, particularly for tests starting at Session 1 (Pearson correlation coefficient = 0.569) and Session 2 (Pearson correlation coefficient = 0.357) but also Session 3 (Pearson correlation coefficient = 0.258). Because collinearity can inflate the error terms, we reran the models with this covariate removed to ensure our models were stable. We found that cluster-related results did not differ when this covariate was removed.

It's possible that the rate of change could be influenced by baseline age rather than cluster (i.e. baseline age × time might be significant rather than cluster × time). Therefore, we re-examined the cluster × time interaction for each outcome after adding a baseline age × time term to the models. The baseline age × time term was not significant in any of the models and changes to the significance of cluster × time terms were minor: the omnibus tests became marginally non-significant for Rey Auditory Verbal Learning Test total trials 1–5 slope ($P = 0.057$) and Boston Naming intercept ($P = 0.056$) but the *post hoc* tests remained significant ($P < 0.05$). These additional analyses lead us to conclude that significant cluster × time interactions indicate that two or more clusters differed in rate of decline and that baseline age was not driving the observed effects of cluster on cognitive decline.

Discussion

This study provides empirical support for the use of cluster analysis to detect heterogeneity in biomarker profiles and potentially identify a subgroup with preclinical Alzheimer's disease in late midlife. We identified four biomarker groups and then investigated clinical interpretations of the groups by examining differential cognitive decline. In line with our hypotheses, we identified a group that appears to represent healthy ageing as well as a group whose biomarker and cognitive profiles are consistent with preclinical Alzheimer's disease. Specifically, the preclinical Alzheimer's disease cluster presented with elevated CSF phosphorylated tau$_{181}$, low CSF amyloid-β$_{42}$/amyloid-β$_{40}$, and greater decline on tests of episodic memory. Consistent with an Alzheimer's-like profile, this cluster was also older and had greater incidence of *APOE4*. Clusters did not differ on other demographic or clinical variables examined. The interpretation of the other two clusters is less clear. The second cluster had a biomarker profile suggestive of preclinical mixed Alzheimer's/vascular pathology and the third cluster showed evidence of hippocampal vulnerability without Alzheimer's pathology; however, their cognitive deficits at this time appear non-specific, with cognitive decline only detected on a test of global cognitive function, the MMSE.

Preclinical Alzheimer's disease is often simply characterized as 'amyloid-β+' or 'amyloid-β−' or by long-term clinical trajectories—both techniques have their limitations. Artificial cut-offs and dichotomization ignore potentially important variability in a continuous variable. Additionally, while amyloid-β is an important biomarker, it does not capture the multifaceted nature of Alzheimer's disease, which is co-characterized by amyloid-β and neurofibrillary tangle pathology as well as other processes (Querfurth and LaFerla, 2010; Sperling *et al.*, 2011; Storandt *et al.*, 2012; Dubois *et al.*, 2016). Long-term trajectories into late life with clinical endpoints are helpful to researchers trying to understand preclinical Alzheimer's disease, but are not pertinent to the ultimate goal of detecting preclinical Alzheimer's disease so that interventions can be initiated during that stage. In contrast, cluster analysis provides an unbiased way to initially characterize multimodal data that captures multiple pathological characteristics simultaneously and can be used throughout the course of the Alzheimer's disease continuum. Here we provided evidence for the use of cluster analysis to identify individuals who appear to be in a preclinical stage of Alzheimer's disease in late midlife, even preceding mild cognitive impairment.

Using biomarkers to cluster groups along the Alzheimer's disease trajectory has been done previously in cohorts of normal elderly (Nettiksimmons *et al.*, 2010, 2013; Pike *et al.*, 2011; Skillback *et al.*, 2013), patients with mild cognitive impairment (Escudero *et al.*, 2011; Nettiksimmons *et al.*, 2014), and patients with dementia (Iqbal *et al.*, 2005; van der Vlies *et al.*, 2009; Wallin *et al.*, 2010; Vemuri *et al.*, 2011; Noh *et al.*, 2014). For instance, two studies in patients with mild cognitive impairment showed that clusters with biomarker profiles consistent with Alzheimer's disease were more likely to progress to dementia due to Alzheimer's disease, illustrating the potential utility for identifying patterns of variables that predict disease progression (Escudero *et al.*, 2011; Nettiksimmons *et al.*, 2014). While clustering has demonstrated utility in elderly and cognitively impaired cohorts, it had yet to be validated during the preclinical Alzheimer's disease stage in late midlife. The current study fills this gap by using cluster-based analysis with neuroimaging and CSF biomarkers in a late-middle-aged risk-enriched cohort. Because this cohort is still relatively young, cognitive decline rather than clinical endpoints were used to evaluate progression.

Cluster 1 was most consistent with an Alzheimer's disease-like biomarker profile (decreased CSF amyloid-β$_{42}$/amyloid-β$_{40}$, increased CSF phosphorylated tau). Significant atrophy measured by gross CSF-to-brain volume ratio and hippocampal volume was notably absent, suggesting that this cluster is representative of an early stage of the Alzheimer's disease continuum preceding frank neuronal death and atrophy. This cluster showed the greatest decline in episodic memory performance over time. Our results are largely consistent with a meta-analysis on 47 studies involving 9097 controls and 1207 preclinical Alzheimer's disease cases that found that preclinical Alzheimer's disease (including mild cognitive impairment) is characterized by marked deficits in global cognitive ability, episodic memory, perceptual speed, and executive functioning; smaller deficits in verbal ability, visuospatial skill, and attention; and no preclinical impairment in primary memory (Backman *et al.*, 2005). Dysfunction in episodic memory is consistently found in various stages of Alzheimer's disease (Albert *et al.*, 2001; Backman *et al.*, 2001; Grober *et al.*, 2008) and is postulated to reflect pathological changes in the medial-temporal lobe, which occur early in the disease (Collie and Maruff, 2000). Longitudinal follow-up of cognitive trajectories will be needed to determine at which point individuals with preclinical Alzheimer's disease progress from cognitive decline to cognitive impairment. Curiously, Cluster 1 also initially performed better (indicated by a larger intercept) on Rey Auditory Verbal Learning Test total trials 1–5 and Logical Memory immediate. One potential theory behind this finding is that *APOE4* is an example of antagonistic pleiotropy whereby *APOE4* could have beneficial effects earlier in life despite being associated with cognitive decline and dementia risk later in life (Wright *et al.*, 2003; Alexander *et al.*, 2007; Mondadori *et al.*, 2007). More than two-thirds of Cluster 1 was *APOE4*+ compared to one-third in the healthy cluster, suggesting that Cluster 1 may demonstrate simultaneous benefits of *APOE4* on tests of verbal episodic memory in terms of intercept, but also negative effects in terms of decline over time. However, the main effect of *APOE4* was not significant in any of our models, and the antagonistic pleiotropy theory of *APOE4* is not

definitive. It will be interesting to see if this intercept effect is replicated as more studies probe the earliest stages of preclinical Alzheimer's disease.

According to a theoretical model by Sperling *et al.* (2011), there are three stages of preclinical Alzheimer's disease: (i) asymptomatic amyloidosis; (ii) co-occurring amyloidosis and neurodegeneration; and (iii) co-occurring amyloidosis, neurodegeneration, and subtle cognitive decline. Jack *et al.* (2012) subsequently extended the NIA-AA criteria for preclinical Alzheimer's disease with two additional categories of cognitively normal individuals: (i) Stage 0 indicated by normal Alzheimer's biomarkers and no evidence of subtle cognitive impairment; and (ii) suspected non-Alzheimer's disease pathophysiology (SNAP). While our clustering approach seemed to identify both of the cognitively 'normal' groups (Clusters 3 and 4) and a preclinical Alzheimer's disease group (Cluster 1), it was unable to parse the three subdivisions of preclinical Alzheimer's disease because cognition was not represented among the initial clustering variables. Based on these models, it would be expected that individuals in the preclinical Alzheimer's disease-like cluster would represent individuals spanning the three stages of preclinical disease. Therefore, to further qualitatively assess heterogeneity within this cluster, we made spaghetti plots for individuals in Cluster 1 for two cognitive tests (Rey Auditory Verbal Learning Test total trials 1–5 and Logical Memory delayed) and the neurodegeneration/atrophy clustering variables (CSF-to-brain volume ratio and hippocampal volume) (Supplementary Fig. 3). Of the 22 participants in Cluster 1, $n = 15$ had at least two time points of MRI data; $n = 5$ had three time points, and $n = 5$ had four time points. A general pattern of intraindividual decline over time is observable on the two cognitive tests, but the group is heterogeneous: some participants steadily improve over time and others show an alternating pattern of improvement and worsening over time. CSF-to-brain volume ratio is highly consistent in showing smaller brain volume associated with older age as well as progressive brain atrophy over time. The pattern for hippocampal volume is less clear, likely due to measurement error of this complex structure and disease stage heterogeneity; but in general, individual hippocampal volume appears relatively stable for most participants over this relatively short interval. These plots provide some evidence for the hypothesized three stages of preclinical Alzheimer's disease within this small single cluster, but further studies with much larger sample sizes will be needed to quantitatively parse out this preclinical staging further. Furthermore, we were not able to address heterogeneity potentially related to hippocampal sparing verses limbic predominant subtypes of Alzheimer's disease, which could result in both biomarker and cognitive profile variability (Murray *et al.*, 2011). An ongoing project is to use molecular imaging of both amyloid and tau to gain spatial resolution of pathology because it is likely that distribution in addition to overall burden would contribute to cognitive and clinical phenotypes.

Cluster 2's biomarker profile (elevated white matter hyperintensity total lesion volume and CSF-to-brain volume ratio, decreased CSF amyloid-β_{42}/amyloid-β_{40}) is consistent with preclinical mixed dementia due to cerebrovascular disease and Alzheimer's disease with cerebral atrophy. That Cluster 2 had the highest body mass index potentially further supports a vascular contribution. White matter hyperintensities are visible on T_2-weighted FLAIR MRI and are thought to represent ischaemic injury, increase with age, and are elevated in Alzheimer's disease (Young *et al.*, 2008; Wen *et al.*, 2009; Alosco *et al.*, 2013; Provenzano *et al.*, 2013; Casado Naranjo *et al.*, 2015). An autopsy study found almost half of the brains of clinically diagnosed Alzheimer's disease patients harboured mixed pathology, the most common of which were infarcts, and that infarcts increased the likelihood of cognitive impairment and dementia (Schneider and Bennett, 2010). White matter ischaemia is postulated to be due to hypoperfusion of the white matter and may be accompanied by vascular stenosis and signs of cardiovascular disease (Brun and Englund, 1986). White matter lesions are also seen in non-demented controls (Scheltens *et al.*, 1995), suggesting that these lesions might be more related to cardiovascular conditions than specific Alzheimer's disease pathology. However, white matter hyperintensities are also seen in patients with autosomal dominant Alzheimer's disease who develop the disease at a younger age, making them less likely to have significant vascular risk factors or atherosclerotic small vessel disease (Ryan *et al.*, 2015). White matter hyperintensities in autosomal dominant populations are thought to be due to cerebral amyloid angiopathy (amyloid-β deposits in the walls of the blood vessels of the CNS) rather than a primarily cardiovascular cause. Therefore, co-occurring low CSF amyloid-β_{42}/amyloid-β_{40} and greater white matter hyperintensities in Cluster 2 could also be indicative of another subset of preclinical Alzheimer's disease with advanced amyloid angiopathy. It is unknown whether vascular brain lesions are co-occurring pathologies which contribute to cognitive impairment independent of amyloid and tau pathology, whether vascular lesions incur greater vulnerability which make the brain more susceptible to Alzheimer's disease pathology, or whether vascular and Alzheimer's disease pathology directly interact to accelerate the Alzheimer's disease cascade. Longitudinal evaluation of individuals in this cluster may provide greater understanding of early co-occurring vascular and amyloid pathology.

Previous studies have demonstrated a relationship between vascular pathology and executive dysfunction, particularly for attention/speed tasks (Moser *et al.*, 2001), even in non-demented patients compared to older healthy controls (Kramer *et al.*, 2002) and preclinical populations (Birdsill *et al.*, 2014). Consistent with this theory, we found that Cluster 2 performed worse on Trail Making Test B, indicated by a larger intercept (higher scores indicate worse performance on this test), but did not show evidence of increased decline over time. Differences in

either intercept or slope were not observed on Digit Symbol, the other executive functioning task. A study of ADNI normal controls found that a cluster characterized by substantial brain atrophy and white matter hyperintensities (mean age 76.7 years) did not show decline on six tests of executive function, but did have significantly higher body mass index, Hachinksi score, creatinine levels, triglycerides, and blood glucose (Nettiksimmons *et al.*, 2013). Similarly, when a single outlier was removed, Cluster 2 showed higher body mass index compared to Clusters 1 and 3. Differences between the ADNI findings and the current results could be due to cohort age differences (Cluster 2 is on average almost 20 years younger than ADNI's suspected vascular cluster) and unlike the ADNI suspected vascular cluster, Cluster 2 additionally demonstrated decreased CSF amyloid-β_{42}/amyloid-β_{40}. In addition to modest differences in initial performance on one test of executive function, we also detected global decline over time on the MMSE. While MMSE is relatively non-specific, decline on this test has been observed in mixed dementia, as well as possible and probable Alzheimer's disease (Corey-Bloom *et al.*, 1993). Our results suggest that even if MMSE scores at any single evaluation remain in the normal range, individuals in the preclinical timeframe may still demonstrate meaningful decline suggestive of underlying disease well in advance of a potential dementia diagnosis.

Cluster 3 presented with increased CSF-to-brain volume ratio, the smallest intracranial volume-corrected hippocampal volume, and decline on MMSE performance. The significantly smaller hippocampal volume in this cluster could be indicative of several pathophysiologies (Jack *et al.*, 2012) including hippocampal sclerosis or some other unknown vulnerability. To examine this further, we performed a supplementary linear mixed effects analysis using longitudinal hippocampal volumes as the dependent variable and fixed effects of sex, years of education, *APOE4*, cohort, intracranial volume at each MRI, cluster, mean-centred age at baseline MRI (59.36 years), time (interval from first MRI visit), and time × cluster. Longitudinal MRI data were only available for a subset of participants ($n = 96$ had at least two time points of MRI data; $n = 45$ had three time points, $n = 10$ had four time points, and $n = 2$ had five time points), so these results should be interpreted with some caution. We found a pronounced intercept effect whereby Clusters 1, 2, and 3 all had smaller hippocampal volumes compared to Cluster 4 (P-values < 0.01, Supplementary Fig. 4). The intercepts of Cluster 1 and Cluster 2 did not differ but Cluster 3 had significantly smaller hippocampal volume compared to Clusters 1 and 2 (P-values < 0.01). There was no effect of slope by cluster. These findings suggest the observed biomarker profile for Cluster 3 is more likely due to hippocampal vulnerability rather than a detectable hippocampal volume loss exceeding the rate of normal ageing due to some other process. It is also likely that this group is more heterogeneous and could represent a normally ageing population that is less healthy than Cluster 4. That this cluster showed steeper decline on the MMSE compared to Cluster 4 further supports a dissociation of this group from healthy ageing. However, we refrain from making stronger conclusions as only longitudinal clinical outcomes can confirm these speculations.

There are several limitations of this study. First, CSF analyte levels are only an indirect measure of pathology, however, lumbar puncture is a common clinical and research technique and the CSF measures of interest (amyloid-β_{42}/amyloid-β_{40} and phosphorylated tau) have been validated by neuroimaging and histological studies. Furthermore, it's been suggested that CSF amyloid-β_{42} and Pittsburgh compound B-PET imaging provide partially independent information about a wide range of Alzheimer's measures, and that reduced CSF amyloid-β may be more strongly related to early stage Alzheimer's disease than Pittsburgh compound B-PET (Mattsson *et al.*, 2015). Tau PET imaging is currently underway in this cohort and will be informative in validating CSF phosphorylated tau findings. Second, there are several putative CSF markers of neurodegeneration. While we could have used other CSF markers of neurodegeneration like total-tau, we chose not to because it was highly correlated in our sample with CSF phosphorylated tau (Pearson correlation coefficient = 0.828). Third, agglomerative hierarchical clustering is only one way to classify individuals on biomarkers. It will be important to investigate and compare other techniques, like machine learning, to the methods used in this study. Fourth, while biomarkers and cognitive data provide evidence of an Alzheimer's disease-like trajectory, clinical endpoints are necessary to confirm the diagnostic accuracy of these clusters. Fifth, because our sample was relatively young and healthy, it is possible that some biomarkers—particularly for hippocampal volume and brain atrophy, which are hypothesized to be the last biomarkers to undergo changes during preclinical Alzheimer's disease—were only minimally affected, making it difficult to make stronger interpretations of those variables. It is expected that the biomarker profiles we observed will change with advancing disease. Sixth, not all tests were administered for the same length of time. While we were able to detect differences in some of the neuropsychological tests with less structurally missing data, it is possible that we lacked sufficient power to detect differences on other tests. It may also be that tests with more temporal data provided better estimates. Seventh, it's possible that within-domain cognitive composites may provide more reliable estimates of cognitive functioning. In this multi-cohort sample, there were only two tests per cognitive domain of interest (ex. Rey Auditory Verbal Learning Test delay and Logical Memory delay would comprise an episodic memory composite); therefore, we chose to examine the neuropsychological tests individually. Future studies may find utility in collecting data to facilitate within-domain cognitive composites to assess changes in cognitive functioning in heterogeneous cohorts. Last, our study sample was largely white

and well educated, and so is not generalizable to all populations; future projects should validate these results in more diverse cohorts. We report the findings of this study with these limitations in mind.

Alzheimer's disease is a growing epidemic for which there is currently no treatment. With clinical trials beginning in the preclinical time period (ex. A4 study), there is imperative need for techniques to identify individuals with preclinical Alzheimer's disease for recruitment into clinical trials and eventual treatment (Sperling *et al.*, 2013). Our results suggest that techniques that combine multimodal information about early pathological changes could inform detection of preclinical Alzheimer's disease and that longitudinal slopes of decline could provide further information about disease trajectory during this critical timeframe.

Acknowledgements

The authors gratefully acknowledge Derek Norton, Richard Chappell, Nancy Davenport-Sis, Amy Hawley, Sandra Harding, Jennifer Oh, Chuck Illingworth, and the support of researchers and staff at the Wisconsin Alzheimer's Disease Research Center, the Wisconsin Alzheimer's Institute, and the University of Wisconsin-Madison for their assistance in recruitment, data collection, and data analysis. Above all, we wish to thank our dedicated volunteers for their participation in this research.

Funding

This research was supported by the National Institutes of Health (S.C.J., AG021155, AG027161), (S.A., AG000213, P50 AG033514), and (B.B.B., AG037639); by P30 HD003352; by a Clinical and Translational Science Award (UL1RR025011) to the University of Wisconsin, Madison; by the Neuroscience & Public Policy Program (SES-0849122); by the Neuroscience Training Program (T32GM007507); by the Medical Scientist Training Program (T32GM008692); by the Wisconsin Alzheimer's Institute Holland Research Fund; and by the Swedish Research Council, the Swedish Brain Foundation, the Knut and Alice Wallenberg Foundation, and Torsten Söderberg's Foundation to the University of Gothenburg. Portions of this research were supported by the Veterans Administration including facilities and resources at the Geriatric Research Education and Clinical Center of the William S. Middleton Memorial Veterans Hospital, Madison, WI.

Supplementary material

Supplementary material is available at *Brain* online.

References

Albert MS, Moss MB, Tanzi R, Jones K. Preclinical prediction of AD using neuropsychological tests. J Int Neuropsychol Soc 2001; 7: 631–9.

Alexander DM, Williams LM, Gatt JM, Dobson-Stone C, Kuan SA, Todd EG, et al. The contribution of apolipoprotein E alleles on cognitive performance and dynamic neural activity over six decades. Biol Psychol 2007; 75: 229–38.

Almeida RP, Schultz SA, Austin BP, Boots EA, Dowling NM, Gleason CE, et al. Effect of cognitive reserve on age-related changes in cerebrospnal fluid biomarkers of Alzheimer disease. JAMA Neurol 2015; 72: 699–706.

Alosco ML, Brickman AM, Spitznagel MB, Griffith EY, Narkhede A, Raz N, et al. Independent and interactive effects of blood pressure and cardiac function on brain volume and white matter hyperintensities in heart failure. J Am Soc Hypertens 2013; 7: 336–43.

Backman L, Jones S, Berger AK, Laukka EJ, Small BJ. Cognitive impairment in preclinical Alzheimer's disease: a meta-analysis. Neuropsychology 2005; 19: 520–31.

Backman L, Small BJ, Fratiglioni L. Stability of the preclinical episodic memory deficit in Alzheimer's disease. Brain 2001; 124 (Pt 1): 96–102.

Berman S, Rivera L, Clark L, Racine A, Carlsson C, Bendlin B, et al. Intracranial arterial 4D-Flow is associated with metrics of brain health and Alzheimer's disease. Alzheimer's Dementia 2015; 1: 420–8.

Birdsill AC, Koscik RL, Jonaitis EM, Johnson SC, Okonkwo OC, Hermann BP, et al. Regional white matter hyperintensities: aging, Alzheimer's disease risk, and cognitive function. Neurobiol Aging 2014; 35: 769–76.

Brun A, Englund E. A white matter disorder in dementia of the Alzheimer type: a pathoanatomical study. Ann Neurol 1986; 19: 253–62.

Casado Naranjo I, Portilla Cuenca JC, Duque de San Juan B, Garcia AF, Sevilla RR, Serrano Cabrera A, et al. Association of vascular factors and amnestic mild cognitive impairment: a comprehensive approach. J Alzheimers Dis 2015; 44: 695–704.

Clatworthy J, Buick D, Hankins M, Weinman J, Horne R. The use and reporting of cluster analysis in health psychology: a review. Br J Health Psychol 2005; 10 (Pt 3): 329–58.

Collie A, Maruff P. The neuropsychology of preclinical Alzheimer's disease and mild cognitive impairment. Neurosci Biobehav Rev 2000; 24: 365–74.

Corey-Bloom J, Galasko D, Hofstetter CR, Jackson JE, Thal LJ. Clinical features distinguishing large cohorts with possible AD, probable AD, and mixed dementia. J Am Geriatr Soc 1993; 41: 31–7.

Dormann CF, Elith J, Bacher S, Buchmann C, Carl G, Carre G, et al. Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. Ecography 2013; 36: 27–46.

Dubois B, Hampel H, Feldman HH, Scheltens P, Aisen P, Andrieu S, et al. Preclinical Alzheimer's disease: definition, natural history, and diagnostic criteria. Alzheimers Dement 2016; 12: 292–323.

Escudero J, Zajicek JP, Ifeachor E. Early detection and characterization of Alzheimer's disease in clinical scenarios using Bioprofile concepts and K-means. In: Proceedings of the 2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Boston, MA, 2011: 6470–3.

Goff DC Jr, Lloyd-Jones DM, Bennett G, Coady S, D'Agostino RB, Sr, Gibbons R, et al. 2013 ACC/AHA guideline on the assessment of cardiovascular risk: a report of the American College of Cardiology/American heart association task force on practice guidelines. J Am College of Cardiology 2014; 63 (25 Pt B): 2935–59.

Grober E, Hall CB, Lipton RB, Zonderman AB, Resnick SM, Kawas C. Memory impairment, executive dysfunction, and intellectual

decline in preclinical Alzheimer's disease. J Int Neuropsychol Soc 2008; 14: 266–78.
Hair JF, Hair JF, Black WC, Babin BJ, Anderson RE. Multivariate data analysis. Pearson Education Limited; Upper Saddle River, NJ, 2013.
Howieson DB, Carlson NE, Moore MM, Wasserman D, Abendroth CD, Payne-Murphy J, et al. Trajectory of mild cognitive impairment onset. J Int Neuropsychol Soc 2008; 14: 192–8.
Iqbal K, Flory M, Khatoon S, Soininen H, Pirttila T, Lehtovirta M, et al. Subgroups of Alzheimer's disease based on cerebrospinal fluid molecular markers. Ann Neurol 2005; 58: 748–57.
Jack CR Jr, Knopman DS, Jagust WJ, Petersen RC, Weiner MW, Aisen PS, et al. Tracking pathophysiological processes in Alzheimer's disease: an updated hypothetical model of dynamic biomarkers. Lancet Neurol 2013; 12: 207–16.
Jack CR Jr, Knopman DS, Weigand SD, Wiste HJ, Vemuri P, Lowe V, et al. An operational approach to National Institute on Aging-Alzheimer's Association criteria for preclinical Alzheimer disease. Ann Neurol 2012; 71: 765–75.
Jansen WJ, Ossenkoppele R, Knol DL, Tijms BM, Scheltens P, Verhey FR, et al. Prevalence of cerebral amyloid pathology in persons without dementia: a meta-analysis. JAMA 2015; 313: 1924–38.
Jellinger KA. The enigma of mixed dementia. Alzheimers Dement 2007; 3: 40–53.
Johnson SC, Christian BT, Okonkwo OC, Oh JM, Harding S, Xu G, et al. Amyloid burden and neural function in people at risk for Alzheimer's disease. Neurobiol Aging 2014; 35: 576–84.
Jonaitis E, La Rue A, Mueller KD, Koscik RL, Hermann B, Sager MA. Cognitive activities and cognitive performance in middle-aged adults at risk for Alzheimer's disease. Psychol Aging 2013; 28: 1004–14.
Koscik RL, La Rue A, Jonaitis EM, Okonkwo OC, Johnson SC, Bendlin BB, et al. Emergence of mild cognitive impairment in late middle-aged adults in the wisconsin registry for Alzheimer's prevention. Dement Geriatr Cogn Disord 2014; 38: 16–30.
Kramer JH, Reed BR, Mungas D, Weiner MW, Chui HC. Executive dysfunction in subcortical ischaemic vascular disease. J Neurol Neurosurg Psychiatry 2002; 72: 217–20.
Mattsson N, Insel PS, Donohue M, Landau S, Jagust WJ, Shaw LM, et al. Independent information from cerebrospinal fluid amyloid-beta and florbetapir imaging in Alzheimer's disease. Brain 2015; 138 (Pt 3): 772–83.
Mondadori CR, de Quervain DJ, Buchmann A, Mustovic H, Wollmer MA, Schmidt CF, et al. Better memory and neural efficiency in young apolipoprotein E epsilon4 carriers. Cereb Cortex 2007; 17: 1934–47.
Morris JC. Early-stage and preclinical Alzheimer disease. Alzheimer Dis Assoc Disord 2005; 19: 163–5.
Moser DJ, Cohen RA, Paul RH, Paulsen JS, Ott BR, Gordon NM, et al. Executive function and magnetic resonance imaging subcortical hyperintensities in vascular dementia. Neuropsychiatry Neuropsychol Behav Neurol 2001; 14: 89–92.
Murray ME, Graff-Radford NR, Ross OA, Petersen RC, Duara R, Dickson DW. Neuropathologically defined subtypes of Alzheimer's disease with distinct clinical characteristics: a retrospective study. Lancet Neurol 2011; 10: 785–96.
Murtagh F, Legendre P. Ward's hierarchical agglomerative clustering method: which algorithms implement ward's criterion? J Classif 2014; 31: 274–95.
Nettiksimmons J, Beckett L, Schwarz C, Carmichael O, Fletcher E, Decarli C. Subgroup of ADNI normal controls characterized by atrophy and cognitive decline associated with vascular damage. Psychol Aging 2013; 28: 191–201.
Nettiksimmons J, DeCarli C, Landau S,: Beckett L, Alzheimer's Disease Neuroimaging I. Biological heterogeneity in ADNI amnestic mild cognitive impairment. Alzheimers Dement 2014; 10: 511–21.e1.
Nettiksimmons J, Harvey D, Brewer J, Carmichael O, DeCarli C, Jack CR Jr, et al. Subtypes based on cerebrospinal fluid and magnetic resonance imaging markers in normal elderly predict cognitive decline. Neurobiol Aging 2010; 31: 1419–28.
Noh Y, Jeon S, Lee JM, Seo SW, Kim GH, Cho H, et al. Anatomical heterogeneity of Alzheimer disease: based on cortical thickness on MRIs. Neurology 2014; 83: 1936–44.
Patenaude B, Smith SM, Kennedy DN, Jenkinson M. A Bayesian model of shape and appearance for subcortical brain segmentation. Neuroimage 2011; 56: 907–22.
Pike KE, Ellis KA, Villemagne VL, Good N, Chetelat G, Ames D, et al. Cognition and beta-amyloid in preclinical Alzheimer's disease: data from the AIBL study. Neuropsychologia 2011; 49: 2384–90.
Provenzano FA, Muraskin J, Tosto G, Narkhede A, Wasserman BT, Griffith EY, et al. White matter hyperintensities and cerebral amyloidosis: necessary and sufficient for clinical expression of Alzheimer disease? JAMA Neurol 2013; 70: 455–61.
Querfurth HW, LaFerla FM. Alzheimer's disease. N Engl J Med 2010; 362: 329–44.
Racine AM, Adluru N, Alexander AL, Christian BT, Okonkwo OC, Oh J, et al. Associations between white matter microstructure and amyloid burden in preclinical Alzheimer's disease: a multimodal imaging investigation. NeuroImage Clin 2014; 4: 604–14
Rajan KB, Wilson RS, Weuve J, Barnes LL, Evans DA. Cognitive impairment 18 years before clinical diagnosis of Alzheimer disease dementia. Neurology 2015; 85: 898–904.
Ryan NS, Biessels GJ, Kim L, Nicholas JM, Barber PA, Walsh P, et al. Genetic determinants of white matter hyperintensities and amyloid angiopathy in familial Alzheimer's disease. Neurobiol Aging 2015; 36: 3140–51.
Sager MA, Hermann B, La Rue A. Middle-aged children of persons with Alzheimer's disease: APOE genotypes and cognitive function in the Wisconsin Registry for Alzheimer's Prevention. J Geriatr Psychiatry Neurol 2005; 18: 245–9.
Scheltens P, Barkhof F, Leys D, Wolters EC, Ravid R, Kamphorst W. Histopathologic correlates of white matter changes on MRI in Alzheimer's disease and normal aging. Neurology 1995; 45: 883–8.
Schmidt P, Gaser C, Arsic M, Buck D, Forschler A, Berthele A, et al. An automated tool for detection of FLAIR-hyperintense white-matter lesions in Multiple Sclerosis. Neuroimage 2012; 59: 3774–83.
Schneider JA, Arvanitakis Z, Bang W, Bennett DA. Mixed brain pathologies account for most dementia cases in community-dwelling older persons. Neurology 2007; 69: 2197–204.
Schneider JA, Bennett DA. Where vascular meets neurodegenerative disease. Stroke 2010; 41 (Suppl 10): S144–6.
Skillback T, Zetterberg H, Blennow K, Mattsson N. Cerebrospinal fluid biomarkers for Alzheimer disease and subcortical axonal damage in 5,542 clinical samples. Alzheimers Res Ther 2013; 5: 47.
Starks EJ, Patrick O'Grady J, Hoscheidt SM, Racine AM, Carlsson CM, Zetterberg H, et al. Insulin resistance is associated with higher cerebrospinal fluid tau levels in asymptomatic APOEvarepsilon4 carriers. J Alzheimers Dis 2015; 46: 525–33.
Sperling RA, Aisen PS, Beckett LA, Bennett DA, Craft S, Fagan AM, et al. Toward defining the preclinical stages of Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. Alzheimers Dement 2011; 7: 280–92.
Sperling RA, Karlawish J, Johnson KA. Preclinical Alzheimer disease-the challenges ahead. Nat Rev Neurol 2013; 9: 54–8.
Storandt M, Head D, Fagan AM, Holtzman DM, Morris JC. Toward a multifactorial model of Alzheimer disease. Neurobiol Aging 2012; 33: 2262–71.
van der Vlies AE, Verwey NA, Bouwman FH, Blankenstein MA, Klein M, Scheltens P, et al. CSF biomarkers in relationship to cognitive profiles in Alzheimer disease. Neurology 2009; 72: 1056–61.
Vemuri P, Simon G, Kantarci K, Whitwell JL, Senjem ML, Przybelski SA, et al. Antemortem differential diagnosis of dementia pathology using structural MRI: differential-STAND. Neuroimage 2011; 55: 522–31.

Villemagne VL, Burnham S, Bourgeat P, Brown B, Ellis KA, Salvado O, et al. Amyloid β deposition, neurodegeneration, and cognitive decline in sporadic Alzheimer's disease: a prospective cohort study. Lancet Neurol 2013; 12: 357–67.

Wallin AK, Blennow K, Zetterberg H, Londos E, Minthon L, Hansson O. CSF biomarkers predict a more malignant outcome in Alzheimer disease. Neurology 2010; 74: 1531–7.

Ward JH. Hierarchical grouping to optimize an objective function. J Am Stat Assoc 1963; 58: 236–44.

Wen W, Sachdev PS, Li JJ, Chen X, Anstey KJ. White matter hyperintensities in the forties: their prevalence and topography in an epidemiological sample aged 44–48. Hum Brain Mapp 2009; 30: 1155–67.

Wright RO, Hu H, Silverman EK, Tsaih SW, Schwartz J, Bellinger D, et al. Apolipoprotein E genotype predicts 24-month bayley scales infant development score. Pediatr Res 2003; 54: 819–25.

Young VG, Halliday GM, Kril JJ. Neuropathologic correlates of white matter hyperintensities. Neurology 2008; 71: 804–11.