Subtypes based on cerebrospinal fluid and magnetic resonance imaging markers in normal elderly predict cognitive decline

J. Nettiksimmons[a], D. Harvey[a], J. Brewer[b], O. Carmichael[c], C. DeCarli[c], C.R. Jack Jr[d], R. Petersen[d], L.M. Shaw[e], J.Q. Trojanowski[e], M.W. Weiner[e], L. Beckett[a,*], The Alzheimer's Disease Neuroimaging Initiative[†]

[a] *Division of Biostatistics, Department of Public Health Sciences, University of California, Davis, Davis, CA 95616, USA*
[b] *Departments of Radiology and Neurosciences, University of California, San Diego, La Jolla, CA 92093-0949, USA*
[c] *Department of Neurology, University of California, Davis, Davis, CA 95616, USA*
[d] *Department of Neurology, Mayo Clinic College of Medicine, Rochester, MN, USA*
[e] *Institute on Aging, Alzheimer's Disease Core Center, Center for Neurodegenerative Disease Research, Department of Pathology and Laboratory Medicine, University of Pennsylvania School of Medicine, Philadelphia, PA, USA*
[e] *Center for Imaging of Neurodegenerative Diseases, University of California, San Francisco, San Francisco, CA, USA*

Abstract

Cerebrospinal fluid (CSF) and structural magnetic resonance imaging (MRI) show patterns of change in Alzheimer's disease (AD) that precede dementia. The Alzheimer's Disease Neuroimaging Initiative (ADNI) studied normal controls (NC), subjects with mild cognitive impairment (MCI), and subjects with AD to identify patterns of biomarkers to aid in early diagnosis and effective treatment of AD. Two hundred twenty-two NC underwent baseline MRI and clinical examination at baseline and at least one follow-up. One hundred twelve also provided CSF at baseline. Unsupervised clustering based on initial CSF and MRI measures was used to identify clusters of participants with similar profiles. Repeated measures regression modeling assessed the relationship of individual measures, and of cluster membership, to cognitive change over 3 years. Most individuals showed little cognitive change. Individual biomarkers had limited predictive value for cognitive decline, but membership in the cluster with the most extreme profile was associated with more rapid decline in ADAS-cog. Subtypes among NC based on multiple biomarkers may represent the earliest stages of subclinical cognitive decline and AD.
© 2010 Elsevier Inc. All rights reserved.

Keywords: Alzheimer's disease; Dementia; Early diagnosis; Cerebrospinal fluid; Tau protein; Amyloid beta-protein; Structural magnetic resonance imaging; Hippocampal volume; Cognition; Clustering; Normal controls

Alzheimer's disease (AD) is a neurocognitive disorder currently estimated to affect some five million people in the USA and more than 25 million worldwide (Brookmeyer et al., 2007; Evans et al., 1990; Harvey et al., 2003; Hebert et al., 2003). Mild cognitive impairment (MCI) has gained recognition as an intermediate clinical category between normal cognitive function and AD, with a greatly increased risk of onset of AD (Bennett et al., 2002; Petersen et al., 2009). AD is characterized not only by cognitive decline, but also by underlying neurobiological changes that likely precede the diagnosis of AD by a considerable period, during MCI and possibly even earlier, before measurable clinical impairment. The process is hypothesized to begin with amyloid deposition, followed by cortical atrophy and decreased metabolism, with effects only gradually becom-

ing apparent in decreased cognitive performance and function (Jack et al., 2010). Identification of early markers of disease would be of great interest to facilitate early diagnosis, improved clinical trials for prevention by targeting individuals at greatest risk, and, ultimately, effective treatment before widespread irreversible neurodegeneration (Clark et al., 2008).

A large number of potential markers have been proposed, including volumetric measures based on magnetic resonance imaging (MRI), cerebrospinal fluid (CSF) biomarkers, FDG PET and others (Hampel et al., 2007; Shaw et al., 2007). Cortical atrophy, for example, is evident on structural MRI not only in AD but also to some extent in people with MCI (Morra et al., 2009; Nestor et al., 2008) and in normal elderly before the onset of MCI (Carlson et al., 2008). Between-person differences in CSF protein levels have also been reported to be associated with AD and MCI (Clark et al., 2008; Maddalena et al., 2003; Shaw et al., 2009). White matter hyperintensity (WMH) has been reported to be increased in patients with AD, suggesting that vascular lesions may also play a role in the neurodegenerative process (Barber et al., 1999). Alternatively, WMH may be a vascular pathology contributing to cognitive impairment in an additive or even multiplicative manner. Homocysteine, a risk factor for vascular damage, has also been hypothesized as a possible risk factor for dementia (Smith, 2008). Most studies of markers have focused on AD and MCI, as the clinical decline is most evident in these groups and association of candidate markers with clinical benchmarks is more readily established. An earlier biomarker horizon, however, would be of great scientific interest and have substantial clinical relevance.

The Alzheimer's Disease Neuroimaging Initiative (ADNI), jointly funded by NIH, pharmaceutical partners, and the Alzheimer's Association, is a multisite research initiative whose aim is to identify biomarkers that would allow the pathological changes of AD to be diagnosed earlier, well before the clinical criteria for dementia are met, and to be tracked more precisely. The goal is to provide earlier diagnosis and better assessment of disease progression and response to therapy. The ADNI participants included normal controls (NC) with detailed standardized assessment of many potential candidate biomarkers and longitudinal follow-up of cognitive outcomes for up to 3 years. We examined a set of imaging and cerebrospinal fluid measures previously proposed in the literature as candidate markers for early diagnosis, and assessed their distribution in normal controls and their relationship to cognitive outcomes over the follow-up period. Our hypothesis was that despite cognitive homogeneity at baseline in the NC subjects, there would be underlying biological heterogeneity in candidate markers, reflecting the earliest detectable changes in the brain. These biological differences would be correlated with each other in a structured way, leading to the ability to construct subgroups based on markers alone, and such sub-

groups would subsequently have different cognitive trajectories.

1. Methods

1.1. The Alzheimer's Disease Neuroimaging Initiative

Data used in the preparation of this article were obtained from the ADNI database (www.loni.ucla.edu/ADNI). The ADNI was launched in 2003 by the National Institute on Aging (NIA), the National Institute of Biomedical Imaging and Bioengineering (NIBIB), the Food and Drug Administration (FDA), private pharmaceutical companies, and non-profit organizations, as a US$60 million, 5-year public-private partnership. The research plan called for recruiting 800 adults, ages 55 to 90: approximately 200 healthy elderly controls, 400 subjects with MCI, and 200 subjects with mild but probable AD. Subjects are followed longitudinally for up to 3 years, with MRI scans, complete cognitive testing, and blood/urine samples collected at 6-month intervals, depending on baseline diagnosis. In addition, subsets of the subjects undergo FDG-PET scans and CSF collection and testing (Mueller, 2005a; Mueller, 2005b).

1.2. Subjects

The individuals studied were recruited between 17 August 2005 and 4 September 2007 as ADNI participants and were identified at baseline clinical evaluation as cognitively normal. NC, MCI and AD participants were frequency-matched by age-group to a common target age profile. NC participants underwent cognitive testing and clinical examination by a physician at baseline and every 6 months for the first year and then annually for the next 2 years. MRI scans (1.5 Tesla) were performed in each subject (www.loni.ucla.edu/ADNI/Research/Cores/index.shtml) at baseline, repeated at 6, 12, 24, and 36 months. Approximately half the participants also provided CSF at the baseline and m12 visits. Additional details are given in Petersen et al. (Petersen et al., 2010). This study was approved by the Institutional Review Boards of all the participating institutions. Informed written consent was obtained from all participants at each site. A detailed description of the study design and inclusion criteria is available at clinicaltrials.gov/show/NCT00106899. Data used in this analysis were downloaded from the ADNI database (www.loni.ucla.edu/ADNI) on 27 September 2009. Analysis focused on NC but the MCI and AD group were described at baseline for comparison purposes.

1.3. Measures

Biomarker summary measures for cluster analysis were selected using a list initially specified by researchers from the ADNI Imaging and Biomarker Cores at the time of grant submission, for core hypothesis tests. MRI summary measures were calculated by the Anders Dale Laboratory at UC San Diego and normalized by their measure of intracranial

volume (ICV) (Fennema-Notestine et al., 2009). Summary measures used were cerebral volume, hippocampal volume, entorhinal cortex thickness, and ventricle volume. WMH were detected by Charles DeCarli and the Imaging of Dementia and Aging laboratory at UC Davis based on coregistered T1-, T2-, and PD-weighted images using an automated protocol described previously (Schwarz et al., 2009). Total WMH volume was used as the primary summary; it was also standardized by ICV. CSF samples were obtained by the individual centers, then banked and batch-processed using a standardized protocol, under the direction of Drs Leslie Shaw and John Trojanowski of the ADNI Biomarker Core at the University of Pennsylvania School of Medicine (Shaw, 2008). CSF measures at baseline included $A\beta_{1-42}$, total tau protein (t-tau), tau protein phosphorylated at the 181 threonine position (P-tau$_{181}$), and the ratios t-tau/$A\beta_{1-42}$ and P-tau$_{181}$/$A\beta_{1-42}$ as previously described (Shaw et al., 2009). Apolipoprotein E genotype was determined and serum homocysteine assayed using a validated enzyme immunoassay methodology by the ADNI Biomarker Core. Two cognitive performance tests were considered for measurement of longitudinal change: the Rey Auditory-Verbal Learning Test (RAVLT) total of five trials, and the Alzheimer's disease Assessment Scale-cognitive subsection (ADAS-cog), selected because some NC showed modest changes in these scales. Higher values of RAVLT and lower values of ADAS-cog correspond to better performance.

1.4. Statistical analysis

All demographic, clinical and marker data were first summarized descriptively (means, standard deviations, graphical summaries). Quantitative marker data (Imaging, CSF and serum) were standardized by subtracting the NC mean and dividing by the standard deviation (SD) before inclusion in regression models and clustering, to facilitate comparisons across markers and analyses. As a comparison for the utility of clusters as predictors, univariate and multivariate mixed models for longitudinal data were used to model changes in RAVLT and ADAS-cog over time using candidate biomarkers as predictors (Laird and Ware, 1982). The annualized rate of change was the model-estimated slope, in units of cognitive test score change per year. Main effects of each predictor estimated the difference in cognitive score at baseline, and interaction terms between predictor and time estimated the effect of a 1 SD increase in predictor on the annualized rate of change in cognitive test score. Random effects were included for slope and intercept, and an unstructured correlation matrix was specified for errors. All models also controlled for education effects on cognitive test score and change. Variables in multivariate models were removed step-wise based on the *p*-value of the variable's slope component. Univariate *p* values were reported without adjustment for multiple comparisons, because the goal was not variable selection, but a reference

comparison for the utility of any specific, individual biomarker compared with cluster membership.

Risk groups (clusters) based on baseline levels of CSF and serum biomarkers and MRI summaries were created by unsupervised hierarchical clustering, without reference to baseline cognitive test scores or to longitudinal trajectories using the function hclust in R. In this agglomerative approach, each individual begins as a cluster of one person. Clusters are then iteratively combined based on dissimilarity. The calculation of dissimilarity relies on maximizing or minimizing an objective function and a metric for distance between individuals. Many options exist for objective functions and distance metrics; we explored seven different objective function methods (Ward's method, single linkage, complete linkage, average, Mcquitty's method, median and centroid) and six different distance metrics (Euclidean, supremum norm, Manhattan, Canberra, binary, and Minkowski). Some combinations of these have a tendency to produce clusters containing a very small number of individuals, essentially isolating individual outliers. These methods are not useful in this context because the resulting sample sizes are too small to allow for further exploration. We restricted consideration to methods that had at least five individuals per cluster. Our primary approach used Ward's method, which is defined as minimal increase in the error sum of squares after combining clusters (Ward, 1963). Error sum of squares was calculated as the sum of the squared Euclidean distances of each cluster member to the cluster center. Similarity across clustering methods was assessed by Rand's statistic, a measure of agreement between two partitions, with 1 representing perfect agreement and 0 corresponding to equal amounts of concordance and discordance (Rand, 1971). We also examined descriptively the degree of agreement between methods on individuals belonging to the group with "unusual" profiles compared with the typical NC. The choice of the number of clusters (k) was based on a combination of observed visual separation, plots of the log within-cluster dissimilarity against k, and restrictions due to sample size. We chose $k = 3$ based on these considerations. Although we are forming subgroups, we hypothesize that these subgroups represent a categorization of an unobserved continuum. Thus, for a larger sample size, a reasonable choice for k may be well beyond 3. Secondary analyses examined whether the effect of cluster membership could be accounted for by ApoE genotype or by age. Preliminary analyses examined conversions from NC to MCI but found too few conversions to allow formal statistical analysis. Both regression models and clustering algorithms used a complete-case approach for biomarkers, excluding any individual missing one or more of the biomarkers in that analysis, and regression models treated missing cognitive scores at scheduled visits as missing at random. Statistical analyses were carried out using the R package (R 2009) and SAS/STAT® software (SAS Institute, 2004).

Table 1

Demographic, clinical, imaging and biomarker characteristics of ADNI participants at baseline, comparing NC, MCI, and AD participants. All participants with baseline data were included, but not all participants had data on each measure

Variables	NC	MCI	AD	*p* value
Demographic	mean (SD)	mean (SD)	mean (SD)	
N[a]	222	382	181	
Age	76.0 (5.0)	74.8 (7.4)	75.1 (7.6)	0.12
Percent male	52%	64%	53%	0.005
Education (years)	16.1 (2.9)	15.6 (3.1)	14.8 (3.1)	< 0.001
Percent ApoE4+	27%	54%	67%	< 0.001
Clinical				
RAVLT score	43.5 (8.9)	30.8 (9.1)	23.3 (7.5)	< 0.001
ADAS-cog score	6.2 (2.9)	11.6 (4.5)	18.5 (6.4)	< 0.001
Imaging				
Cerebral volume[b]	0.685 (0.025)	0.671 (0.027)	0.659 (0.026)	< 0.001
Ventricle volume[b]	0.026 (0.012)	0.031 (0.014)	0.036 (0.016)	< 0.001
Hippocampal volume[b]	0.0050 (0.0005)	0.0044 (0.0007)	0.0040 (0.0007)	< 0.001
Entorhinal cortex thickness	6.50 (0.60)	5.85 (0.93)	5.14 (0.88)	< 0.001
White matter hyperintensity	2.65 (2.48)	2.63 (2.47)	3.92 (7.32)	< 0.001
Serum biomarker				
Homocysteine[c]	9.93 (2.88)	10.6 (2.90)	10.6 (3.16)	0.02
CSF biomarkers				
CSF sample size	112	189	98	
$A\beta_{1\text{-}42}$[d]	206 (55)	164 (55)	143 (41)	< 0.001
t-tau[d]	69 (28)	102 (54)	119 (56)	< 0.001
P-tau$_{181}$[d]	25 (14)	35 (17)	41 (19)	< 0.001
t-tau/$A\beta_{1\text{-}42}$	0.4 (0.25)	0.7 (0.56)	0.9 (0.48)	< 0.001
P-tau$_{181}$/$A\beta_{1\text{-}42}$	0.1 (0.12)	0.3 (0.17)	0.3 (0.18)	< 0.001

[a] Sample sizes are based on subjects with at least one clinical follow-up assessment. Means and standard deviations are based on participants with data at baseline, so calculations for imaging and serum biomarker measures may be based on fewer subjects.

[b] Presented as fraction of ICV.

[c] Presented as μmol/L.

[d] Presented in pg/mL.

2. Results

2.1. Normal control participant characteristics

Baseline MRI were available for 222 NC and CSF biomarker data for 112 NC. However, not all image summary measures were available for all images (range: 192–222). In comparison with ADNI MCI and AD participants, the NC were similar in age but had higher education, fewer males, and fewer participants with one or more ApoE4 alleles (Table 1). As expected, given ADNIs design to yield three clearly differentiated groups, the NC had substantially higher RAVLT scores and lower (better) ADAS-cog scores at baseline; only about 5% of the NC group had scores as bad as the average MCI participant at baseline. Imaging and CSF biomarkers, however, showed substantial overlap between groups; the 25% of the NC group with the worst baseline measurement for almost all biomarkers was at or beyond the baseline median of the MCI group (Petersen et al., 2010). Means on summaries of MRI showed smaller regional and global volumes, thinner entorhinal cortex, and larger ventricles for AD than MCI, and for MCI than NC. However, the MCI mean was typically just one standard deviation worse than the NC mean. CSF biomarker mean levels for participants who had baseline lumbar puncture were consistent with previously reported patterns with higher tau protein and lower amyloid beta associated with

greater cognitive impairment as captured by clinical diagnosis. Considerable variability within the diagnostic groups led to substantial overlap across the groups.

The NC showed little change in clinical status during the first 36 months of follow-up. Only 10 converted to MCI during this period, and 1 of these 10 subsequently to AD. These NC had a mean of 4.4 (SD = 0.7) observations per person of ADAS-cog and a mean of 4.3 (0.8) observations of RAVLT. For most participants, there was little change either in ADAS-cog score (Figure 1) or in RAVLT performance (Figure 2), as would be expected in normal controls. Random effects models for RAVLT and ADAS-cog showed significant between-person variance in initial cognitive test scores (random intercept) and rate of change (random slope), indicating unexplained heterogeneity in trajectories and supporting exploration of baseline imaging summaries and fluid biomarkers as possible predictors of early cognitive decline.

2.2. Imaging and cerebrospinal fluid measures as predictors of cognitive decline

We fitted mixed models for possible predictors of cognitive decline, separately with RAVLT and ADAS-cog trajectories as the outcome. Each imaging or fluid biomarker predictor, standardized as described above was first considered individually in a model adjusted for education, and the
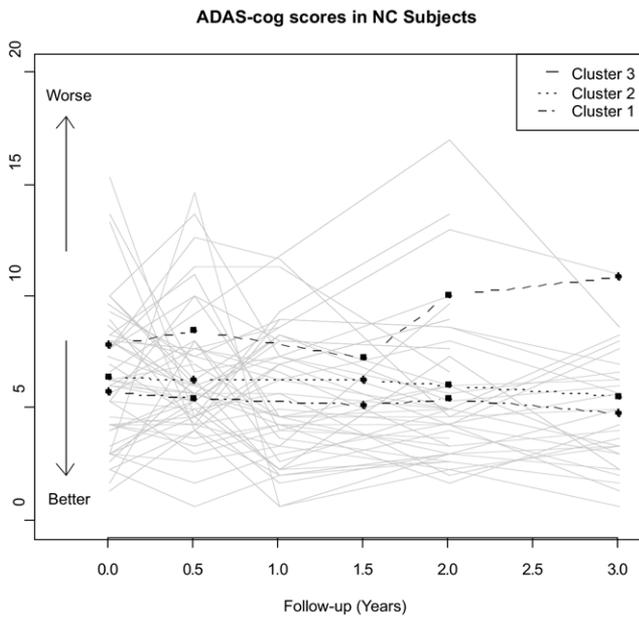
Fig. 1. Longitudinal trajectories of ADAS-cog scores for a random sample of 50 ADNI normal controls. The average longitudinal trajectory for each of the three clusters is also presented.

association of a one SD higher level of the biomarker at baseline with cognitive score at baseline and with annualized rate of change in cognitive score was estimated using models with random effects for intercept and slope (Table 2). Baseline test scores were significantly correlated in univariate models with baseline levels of many MRI and CSF
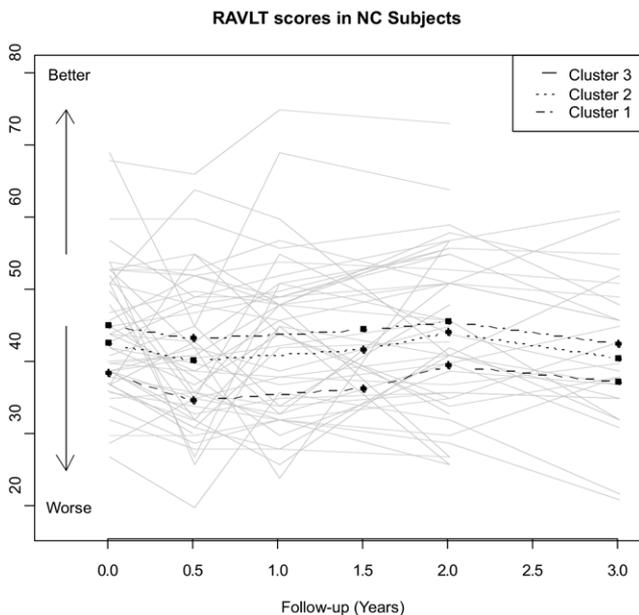

Fig. 2. Longitudinal trajectories of RAVLT scores for a random sample of 50 ADNI normal controls. The average longitudinal trajectory for each of the three clusters is also presented.

Table 2
Univariate estimates from mixed models (random slope and intercept) of association of one-standard deviation higher level in baseline imaging, serum, or CSF marker with cognitive test outcome at baseline and with annualized rate of change in cognitive test score. All imaging summaries were normalized by total intracranial volume before standardizing by baseline mean and standard deviation of normal controls. Predictors significant at 0.05 level are indicated with bold face

Predictor	Outcome: ADAS-cog score			Outcome: RAVLT score		
	Estimate	SE	*p*-value	Estimate	SE	*p*-value
Effects on baseline score						
Male	**−1.01**	**0.35**	**0.004**	**5.92**	**1.15**	**< 0.001**
ApoE4+	0.60	0.38	0.11	1.01	1.32	0.44
Homocysteine	0.28	0.17	0.10	**−1.85**	**0.58**	**0.001**
Cerebral vol	−0.31	0.17	0.07	**1.34**	**0.60**	**0.026**
Hipp vol	**−0.41**	**0.17**	**0.02**	**1.62**	**0.60**	**0.007**
Vent vol	**0.52**	**0.17**	**0.003**	**−1.96**	**0.59**	**0.001**
Entorhin thick	**−0.71**	**0.17**	**< 0.001**	**2.20**	**0.60**	**< 0.001**
White matter	0.09	0.18	0.63	−0.06	0.65	0.93
t-tau	0.20	0.22	0.36	−0.27	−0.77	0.73
P-tau$_{181}$	0.30	0.22	0.18	−1.21	0.77	0.11
A$\beta_{1\text{-}42}$	**−0.52**	**0.22**	**0.017**	0.60	0.76	0.43
t-tau/A$\beta_{1\text{-}42}$	0.36	0.22	0.10	−0.73	0.76	0.33
P-tau$_{181}$/A$\beta_{1\text{-}42}$	**0.49**	**0.22**	**0.027**	**−1.57**	**0.76**	**0.041**
Effects on annual change						
Male	0.05	0.16	0.76	0.47	0.39	0.23
ApoE4+	0.29	0.18	0.10	0.21	0.44	0.63
Homocysteine	**0.16**	**0.08**	**0.047**	0.10	0.20	0.64
Cerebral vol	−0.09	0.08	0.28	0.14	0.20	0.50
Hipp vol	**−0.16**	**0.08**	**0.046**	0.17	0.20	0.38
Vent vol	0.08	0.08	0.31	0.01	0.20	0.97
Entorhin thick	−0.05	0.08	0.55	−0.06	0.20	0.76
White matter	−0.01	0.08	0.90	−0.20	0.19	0.29
t-tau	0.02	0.11	0.90	0.05	0.25	0.83
P-tau$_{181}$	**0.21**	**0.11**	**0.048**	0.08	0.24	0.73
A$\beta_{1\text{-}42}$	−0.09	0.11	0.42	−0.27	0.25	0.29
t-tau/A$\beta_{1\text{-}42}$	0.08	0.12	0.50	0.26	0.27	0.33
P-tau$_{181}$/A$\beta_{1\text{-}42}$	**0.23**	**0.11**	**0.036**	0.16	0.24	0.52

markers. Higher (worse) initial ADAS-cog scores were associated with lower cortical volume at baseline as measured by hippocampal volume (0.41 points better for each SD larger, $p = 0.02$), entorhinal cortex (0.71 points better for every SD thicker, $p < 0.001$), and ventricular volume (0.52 points worse for every SD larger, $p = 0.003$). CSF biomarker differences at baseline were also associated with worse initial performance on the ADAS-cog; a one SD lower CSF A$\beta_{1\text{-}42}$ ($p = 0.017$) predicted about a half-point worse ADAS-cog, as did a one SD higher P-tau$_{181}$/A$\beta_{1\text{-}42}$ ratio ($p = 0.027$). Higher (better) initial RAVLT scores were associated with larger cortical volume, whether measured by a one SD increase in cerebral volume (1.34 point higher score, $p = 0.026$), hippocampal volume (1.62 point better, $p = 0.007$), or entorhinal thickness (2.20 points better, $p < 0.001$). A one SD increase in ventricular volume was associated with initial RAVLT scores almost two points worse ($p = 0.001$). Higher homocysteine levels at baseline were also associated with significantly worse initial RAVLT score (1.85 points worse for a one SD higher level, $p = 0.001$.) A one SD higher P-tau$_{181}$/A$\beta_{1\text{-}42}$ ratio was also

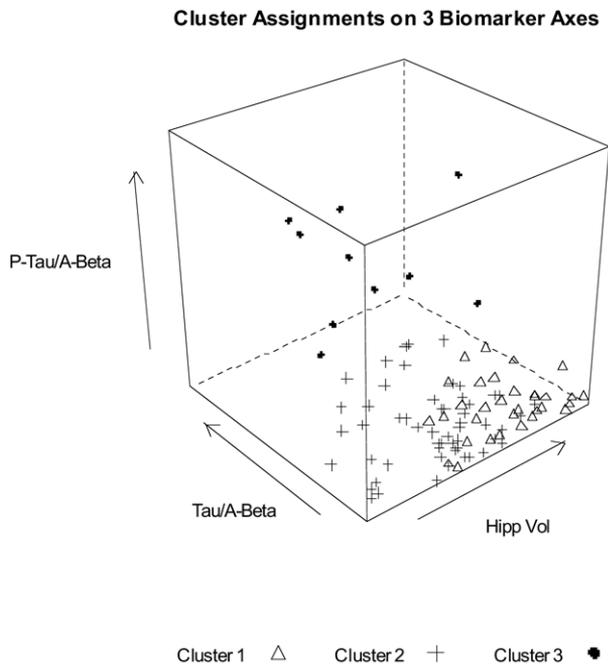**Cluster Assignments on 3 Biomarker Axes**

Fig. 3. Three-dimensional representation of cluster assignments for 96 ADNI normal controls, based on unsupervised clustering using 11 baseline MRI and CSF markers and serum homocysteine, without use of clinical data to define clusters. This representation is shown on three axes of the original 11-dimensional data space, chosen to maximize separation of cluster centers for clearest visualization on paper of the cluster locations.

associated with a 1.57 point lower baseline RAVLT score ($p = 0.041$).

Baseline imaging and biomarker scores, though, had little predictive impact on ADAS-cog or RAVLT change, assuming cognitive test score changes are linear over the period of observation. A one SD higher homocysteine level was associated with a 0.16 point per year faster worsening in ADAS-cog performance ($p = 0.047$), while a one SD smaller hippocampus was associated with about the same magnitude of difference in rate of ADAS-cog decline ($p = 0.046$). A one SD higher P-tau$_{181}$ or P-tau$_{181}$/A$\beta_{1\text{-}42}$ ratio was associated with about a 0.2 point faster annual worsening of ADAS-cog. These effects are modest; it would take about 5 years at this rate to account for a single additional point worse on the ADAS-cog. Adjusting *p*-values in Table 2 for multiple comparisons, as would be done for selection of best predictors, would further reduce the evidence for utility of individual biomarkers as predictors of cognitive baseline or decline.

In multivariate models adjusted for education, no individual baseline measure of cortical volume or CSF remained a significant predictor of cognitive decline, either for the ADAS-cog or the RAVLT.

These analyses found little evidence of associations between either the individual standardized imaging and biomarker measures of this study or weighted sums of the measures and subsequent cognitive decline in NC, over a

period of up to 3 years. These results served as a comparison for analyses which focused on an effort to identify individuals with a distinctive profile of biomarkers that separated them out from typical NC, without reference to their cognitive outcomes, and then to assess the behavior of cognitive trajectories for these subgroups.

2.3. Cluster analysis of normal controls using baseline imaging and biomarker measures

Cluster analysis of the NC using the 11 imaging and biomarker measures (96 complete cases) identified three distinct clusters, of size 32, 54, and 10. The first cluster was relatively compact, while the third cluster was well separated in multidimensional space even when individual measurements might fall in the normal range (Figure 3, projection from 11 dimensions to 3, projection chosen to separate cluster centers for best visualization of cluster locations). The second cluster was intermediate, but closer to the first, "healthier" cluster. We calculated the center (11-dimensional point representing the average across all people in the cluster for each marker) for each of the three clusters and for the MCI and AD groups (Table 3). Note that the MCI and AD data were not used to develop the clusters among the NC, but are included for comparison. Because all the markers were standardized using the NC mean and standard deviation, components of the centers are in units of NC SD. The center for Cluster 3, the atypical group, is 1.5–2.5 SD away from the normal mean for all the baseline CSF biomarkers. The average person in this cluster has a smaller cortical volume than the compact, "healthier" Cluster 1 but not as small as a typical individual in Cluster 2. Overall, Cluster 3 looks very close to the AD group for CSF and closer to MCI than to normal for the MRI variables (Figure 4). We also calculated Euclidean distances for the centers of each of the three clusters to the MCI and AD groups. Cluster 3, the "atypical" cluster, was closer to the MCI and AD than to either of the other two NC clusters (2.6 and 2.8 SD v. 4.8 and 5.6 SD). Cluster 2 was equidistant from Cluster 1, the

Table 3
Marker values at center of NC clusters, compared with center of MCI and AD groups. All values are in SD units away from the overall NC mean

| Variable name	Location of center of group, in SD from NC mean				
	NC cluster 1	NC cluster 2	NC cluster 3	MCI	AD
Cerebral vol	1.08	−0.57	0.00	−0.54	−1.05
Hippo vol	0.84	−0.38	−0.19	−1.11	−1.79
Vent vol	−0.86	0.45	0.00	0.46	0.88
Entor thick	0.62	−0.24	−0.60	−0.83	−1.50
WMH	−0.24	0.12	−0.27	−0.02	0.50
Homocysteine	−0.44	0.23	0.10	0.23	0.23
t-tau	−0.26	−0.09	1.56	1.17	1.80
Ab$_{1\text{-}42}$	0.22	0.03	−1.44	−0.77	−1.15
P-tau$_{181}$	−0.39	−0.12	2.19	0.79	1.22
t-tau/Ab$_{1\text{-}42}$	−0.32	−0.13	2.15	1.43	2.10
P-tau$_{181}$/Ab$_{1\text{-}42}$	−0.37	−0.17	2.45	0.92	1.42

Average Z-scores by Group
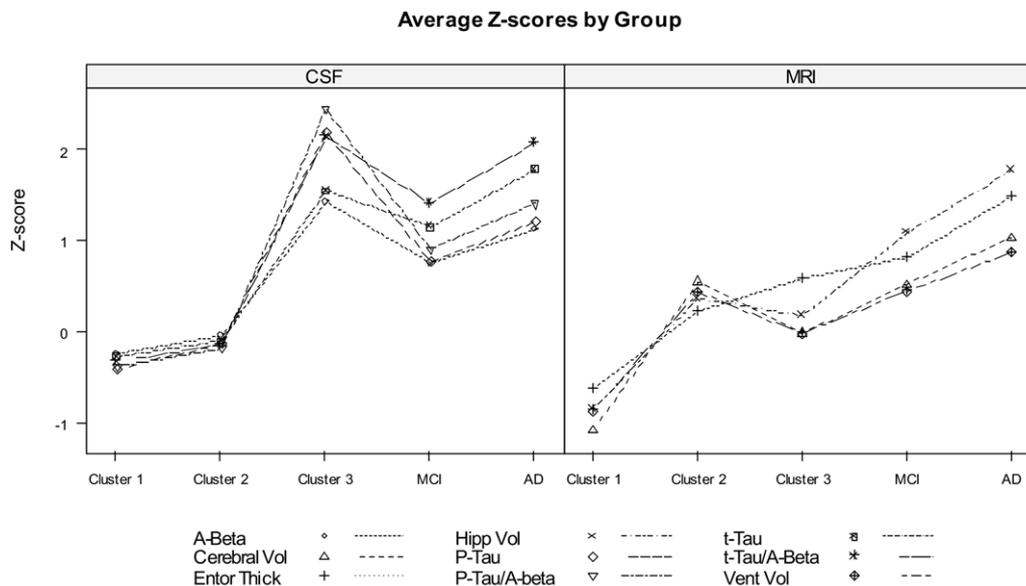
Fig. 4. Difference for each biomarker of cluster means from the mean for all ADNI normal controls, standardized by normal control standard deviation, compared with distances for ADNI MCI and AD groups. Sign of standardized differences was reversed for all variables except $A\beta_{1-42}$ and ventricular volume so that high values (top of figure) represent worse biomarker measurements.

"typical" normal group, and MCI (2.7SD). Cluster 1 was furthest from AD (6.1 SD). The MCI and AD groups were closer together than any other two groups (1.7 SD).

Alternative clustering techniques generally led to similar cluster assignments, particularly with regard to the atypical cluster (Cluster 3). Among methods that did not result in clusters with too few individuals, the results were fairly consistent with the results from Ward's method, as measured by Rand's statistic (average 0.64, range = [0.45, 1]) and by the tendency to locate the same atypical cluster (over half kept at least 9 of the 10 members of the atypical cluster together, most of these locating exactly the same cluster of 10). Using just MRI summary measures and omitting CSF biomarkers approximately doubled the available sample size but led to less clear separation between clusters, consistent with hypotheses that changes in amyloid levels precede volumetric change (Jack et al., 2010; Petersen et al., 2009).

We fitted mixed effects models to assess the predictive value of cluster membership at baseline on subsequent cognitive change (Table 4). Members of Cluster 3, the profile farthest from a typical NC, had baseline scores 2 points worse on average than the typical individuals in Cluster 1 for the ADAS-cog (p = 0.021) and 9 points worse on average for the RAVLT (p = 0.002). The atypical baseline marker profile of Cluster 3 was also accompanied by a faster rate of worsening on the ADAS-cog (1.25 points worse per year, p < 0.001). The rate of decline for RAVLT, while of similar magnitude, was not statistically significant (p = 0.17). A permutation test was performed to determine the likelihood of a different clustering with groups of these sizes resulting in an equal or smaller p-value for the cluster variable indicating decline in ADAS-cog (p = 0.0025). Out

of 5,000 random clusterings, only 21 (less than 1%) resulted in a p-value ≤ 0.0025.

Further characterization of the three clusters found that 50% of the members of Cluster 3 had at least one ApoE4 allele, while 25% and 22% of the members of Clusters 1 and 2 were E4+.However, the association of cluster membership with worse baseline performance and more rapid deterioration of ADAS-cog performance was not accounted for by ApoE genotype or by participant age; results were not materially changed by including in the model the effects of these factors on baseline cognition and rate of change. Cluster size and composition were sensitive at the margins to choice of measures included and algorithms used but were overall quite robust.

Among the 96 people assigned to clusters, only 7 people converted to MCI, 2 in Cluster 3 (20%), 4 in Cluster 2 (8%) and 1 in Cluster 1 (3%). These numbers, while too small for

Table 4
Mixed-model estimates of effects of cluster membership based on baseline imaging and biomarker measures on cognitive test outcome at baseline and on annualized rate of change. Predictors significant at 0.05 level are indicated with bold face. Cluster 3 is the most atypical and Cluster 2 closer to the "typical" Normal Control

Variable	ADAS-cog			RAVLT		
	Estimate	SE	*p*-value	Estimate	SE	*p*-value
Effects at baseline						
Cluster 3	**1.96**	**0.85**	**0.021**	**−9.08**	**2.84**	**0.002**
Cluster 2	0.85	0.52	0.10	**−3.45**	**1.73**	**0.048**
Effects on annual change						
Cluster 3	**1.25**	**0.37**	**<0.001**	1.18	0.87	0.17
Cluster 2	0.03	0.22	0.88	0.32	0.51	0.54

Competing interests. None reported for JN, JB, LB, DH.

statistical analysis, indicate the importance of additional follow-up of the NC.

3. Discussion

Our analyses in a large, well-characterized cohort of normal controls found heterogeneity in imaging summary measures and fluid biomarker summaries, indicating that the biomarker profile could not be encompassed by a single, compact, well-defined set of boundaries. Instead, the patterns were suggestive of several distinct clusters of individuals, even though these individuals were cognitively intact at baseline and showed very little clinical progression over 3 years of follow-up. Regression models suggested that baseline levels of some measures of brain size and CSF biomarkers are associated with cross-sectional differences in cognitive performance, and that increased P-tau$_{181}$ and the P-tau$_{181}$/Aβ_{1-42} ratio may foreshadow slightly more rapid declines in ADAS-cog. The effects, however, were modest, and not improved in multivariate regression analysis. Cluster analysis identified three distinct groups of individuals, based solely on their baseline imaging and biomarker measures without reference to cognitive status or change. One group in particular, comprising about 10% of the NC group, was well separated from the bulk of the NC and lay closer to the centers of the MCI and AD groups, even when some individual marker measurements might be closer to the center for typical NC. This subgroup had strikingly lower baseline scores on the RAVLT, significantly worse scores on the ADAS-cog, and a significantly more rapid deterioration on the ADAS-cog than the typical NC. Membership in this cluster was associated with annual cognitive decline approximately 5 times as rapid as that predicted for a person one standard deviation worse than average on the strongest individual marker. A second group, centered between the atypical group and the more typical NC group, showed structural MRI profiles closer to the MCI group and had somewhat worse baseline performance but showed little difference in cognitive trajectory compared with typical NC. The division into distinct clusters was robust across several clustering algorithms, with the same people consistently identified as a distinct group whose center placed them closer to the profile of MCI and AD. As a complementary approach, unsupervised regression trees were used to generate a proximity matrix which was then plotted on multidimensional scaling axes and colored by cluster membership. The plot showed almost no mixing of clusters, supporting the clusters identified with agglomerative clustering despite the fact that methodologies have almost nothing in common. This suggests that the clusters are not simply an artifact of the clustering process and may instead represent meaningful structure in the data.

Our findings are consistent with previous work suggesting P-tau$_{181}$, Aβ_{1-42}, and their ratio as among the measures most closely reflecting early preclinical neurobiological changes (Shaw et al., 2009; Stomrud et al., 2007). We also found evidence for correlation between brain atrophy measures and cross-sectional measures of cognitive performance, even in this very high-functioning cohort. None of these measures, however, had substantial predictive power for future cognitive decline, either taken individually or as linear combinations. Little previous work has explored the multidimensional structure of imaging and CSF biomarkers in clinically normal older people. A recent study (Fagan et al., 2009) assessed correlations between CSF measures and normalized whole brain volume in cognitively normal elderly subjects and found that Aβ_{1-42}, but not Tau or P-tau$_{181}$, correlated inversely with whole-brain volume in elderly control subjects, while Tau and P-tau$_{181}$ correlated in very mild and mild AD. Such findings have suggested a hypothetical model of the sequence of biomarker changes in the neurobiological process leading to AD, with amyloid changes occurring very early, followed by tau pathology, volumetric and metabolic decline, and with a very gradual onset of measurable cognitive decline, reaching the diagnosis of dementia only very late in the cascade (Jack et al., 2010). Our analyses offer additional support for this conceptual model. We identified a subgroup of normal controls, based solely on biomarker profiles, who were notable primarily for amyloid-related CSF abnormalities. Not only were people in this group much closer to the AD group in their CSF profile, but they also had a striking pattern of ADAS-cog decline over 2–3 years, despite being clinically well within the normal range to start. Figure 4 suggests strong similarity to the conceptual model proposed in Jack et al., 2010. Clustering based just on MRI measures failed to isolate such a distinctive subgroup. Our findings are consistent with Cluster 3 representing a group of people who have already progressed so far in amyloid deposition that they are starting to experience the earliest signs of clinical decline, even though their scores remain, for the most part, within the wider normal range and do not yet verge on dementia.

Methods looking at best classification strategies to predict conversion among NC are not yet applicable in ADNI, as there are so few conversions. We are not aware of any work with data mining or other dimension-reduction strategies based on rate of cognitive change as an outcome. Our study represents a first attempt to seek structure based strictly on the imaging and fluid biomarker measures, and to map the observed heterogeneity to distinct subgroups.

One limitation of our approach is that the CSF markers were available only for half the participants, thus reducing the sample size available for clustering. This, in turn, led to a fairly small though stable "atypical" cluster of about 10% of the NC with CSF. The remaining participants split with every clustering algorithm we examined, with some uncertainty in locating the border between a compact, "typical" group and a second group leaning toward the "atypical" group. A larger sample size might have allowed better

definition of the margins, as well as the variables most important for identifying the "atypical" cluster. A second limitation is that length of follow-up and the generally healthy nature of this cohort led to only a few conversions to MCI, precluding our assessment of the predictive value of the cluster for change in clinical status. A third limitation is that we used only a fraction of the many potential biomarkers, especially from the imaging summaries, and we included five participants with some imaging quality control problems whose effect on our findings appeared minimal in the primary analysis. Our selection was based on previous literature, but may have omitted variables with better prognostic power in the ADNI cohort. A fourth limitation is the approach to standardization of the imaging variables. Dividing by ICV may diminish differences between individuals or groups, although the cluster definitions were more dependent on the CSF biomarkers. Finally, we have not used ADNI's FDG PET or PiB imaging data because the number of participants with complete data on the MRI, CSF and PET imaging data would be too small for analysis.

Our approach has several notable strengths. First, the ADNI data offer one of the largest, uniformly ascertained databases available, with standardized protocols not only for data collection but also for image processing and biomarker specimen assays. Second, our approach makes use of 36 months of follow-up with standardized cognitive testing. Third, our clustering strategy was "unsupervised", that is, based on the imaging and biomarker data without reference to cognitive endpoints or diagnostic categories, so that we address directly the question of whether a strikingly unusual profile on these measures may foreshadow clinical decline, without letting the outcome define the profile of interest. Notably, a recent study on ADNI CSF biomarkers used another unsupervised learning approach that identified nearly the same CSF Aβ cut point as Shaw et al. (2009) wherein this cut point was established in subjects with autopsy confirmed diagnoses of AD thereby demonstrating the power of unsupervised analytical approaches (de Meyer et al., 2010). Supervised techniques such as regression trees or discriminant analysis offer a complementary approach, focused on identifying subgroups of clinical interest based on their outcome, but would require cross-validation, potentially problematic in our sample of under 100 NC with little clinical change as yet. In comparison with univariate or multivariate regression models, our approach takes advantage of the correlated relationship among the biomarkers, rather than being limited by this interrelationship. This strength of the clustering approach may account for the improvement in our ability to detect associations with future cognitive change, compared with the more traditional univariate and multivariate regression models.

Our findings indicate that not just individual abnormalities, but a distinctive pattern of imaging and biomarker deviations from typical healthy older adults may be an early warning sign of neurobiological pathology. Additional follow-up would help to establish the prognostic value for longer-term cognitive decline and conversion to MCI and eventually to AD. A larger sample, with CSF on all participants, would help to confirm the cluster pattern for our "atypical" group and better define the boundary between the other two clusters. In addition, the recruitment of early MCI participants, as currently under way in the recently funded ADNI-GO grant, will offer an opportunity to test whether our atypical group have a profile similar to the earliest stages of clinically identified impairment. Longitudinal analysis of the imaging and biomarker measures can also address whether the distance of the atypical cluster relative to the center for typical NC is increasing over time, and whether the location of the intermediate cluster shifts toward the atypical cluster's baseline location over time.

4. Conclusions

A pattern of imaging and biomarker deviations can identify a subgroup of normal controls distinct from typical cognitively normal older participants. This subgroup, defined based strictly on imaging and biomarkers, had substantially worse cognitive performance at baseline and a more rapid rate of deterioration on the ADAS-cog than the typical NC. Additional follow-up and analysis is warranted to investigate the long-term clinical implications and to characterize the imaging and biomarker trajectories of subgroups over time.

Acknowledgements

References

Barber, R., Scheltens, P., Gholkar, A., Ballard, C., McKeith, I., Ince, P., Perry, R., O'Brien, J., 1999. White matter lesions on magnetic resonance imaging in dementia with Lewy bodies, Alzheimer's disease, vascular dementia, and normal aging. BMJ 67, 66–72.

Bennett, D.A., Wilson, R.S., Schneider, J.A., Evans, D.A., Beckett, L.A., Aggarwal, N.T., Barnes, L.L., Fox, J.H., Bach, J., 2002. Natural history of mild cognitive impairment in older persons. Neurology 23, 198–205.

Brookmeyer, R., Johnson, E., Ziegler-Graham, K., Arrighi, H.M., 2007. Forecasting the global burden of Alzheimer's disease. Alzheimers Dement 3, 186–191.

Carlson, N.E., Moore, M.M., Dame, A., Howieson, D., Silbert, L.C., Quinn, J.F., Kaye, J.A., 2008. Trajectories of brain loss in aging and the development of cognitive impairment. Neurology 70, 828–833.

Clark, C.M., Davatzikos, C., Borthakur, A., Newberg, A., Leight, S., Lee, V.M., Trojanowski, J.Q., 2008. Biomarkers for early detection of Alzheimer pathology. Neuro Signals 16, 11–18.

De Meyer, G., Shapiro, F., Vanderstichele, H., Vanmechelen, E., Engleborghs, B., De Deyn, P.-P., Coart, E., Hanson, O., Minthon, L., Zetterberg, H., Blennow, K., Shaw, L.M., Trojanowski, J.Q., The Alzheimer's Disease Neuroimaging Initiative, 2010. Diagnosis-independent Alzheimer's disease biomarker signature in cognitively normal elderly people. Arch Neurol. [in press].

Evans, D.A., 1990. Estimated prevalence of Alzheimer's disease in the United States. Milbank Q. 68, 267–289.

Fagan, A.M., Head, D., Shah, A.R., Marcus, D., Mintun, M., Morris, J.C., Holtzman, D.M., 2009. Decreased cerebrospinal fluid Abeta(42) correlates with brain atrophy in cognitively normal elderly. Ann Neurol 65:176–183.

Fennema-Notestine, C., Hagler, D.J., Jr, McEvoy, L.K., Fleisher, A.S., Wu, E.H., Karow, D.S., Dale, A.M. and the ADNI, 2009. Structural MRI biomarkers for preclinical and mild Alzheimer's Disease. Human Brain Mapping, 30:3238–3253.

Hampel, H., Burger, K.B., Teipel, S.J., Bokde, A.L.W., Zetterberg, H., Blennow, K., 2007. Core candidate neurochemical and imaging biomarkers of Alzheimer's disease. Alzheimers Dement 4, Nos. 1, 38–48. J Alzheimer's Assoc, 4(1), 38–48.

Harvey, R.J., Skelton-Robinson, M., Rossor, M.N., 2003. The prevalence and causes of dementia in people under the age of 65 years. J Neurol Neurosurg Psychiatry 74, 1206–1209.

Hebert, L.E., Scherr, P.A., Bienias, J.L., Bennett, D.A., Evans, D.A., 2003. Alzheimer disease in the US population: prevalence estimates using the 2000 census. Arch Neurol 60, 1119–1122.

Jack, C.R., Jr, Knopman, D.S., Jagust, W.J., Shaw, L.M., Aisen, P., Weiner, M.W., Petersen, R.C., Trojanowski, J.Q., 2010. Hypothetical model of dynamic biomarkers of the Alzheimer's pathological cascade. Lancet Neurol 9, 119–128.

Laird, N.M., Ware, J.H., 1982. Random-effects models for longitudinal data. Biometrics 38, 963–974.

Maddalena, A., Papassotiropoulos, A., Muller-Tillmanns, B., Jung, H.H., Hegi, T., Nitsch, R.M., Hock, C., 2003. Biochemical Diagnosis of Alzheimer Disease by Measuring the Cerebrospinal Fluid Ratio of Phosphorylated tau Protein to beta-Amyloid Peptide42. Arch Neurol 60, 1202.

Morra, J.H., Tu, Z., Apostolova, L.G., Green, A.E., Avedissian, C., Madsen, S.K., Parikshak, N., Toga, A.W., Jack, C.R., Schuff, N., Weiner MW, Thompson PM, Alzheimer's Disease Neuroimaging Initiative, 2009. Automated mapping of hippocampal atrophy in 1-year repeat MRI data from 490 subjects with Alzheimer's disease, mild cognitive impairment, and elderly controls. Neuroimage 45, S3–15.

Mueller, S.G., Weiner, M.W., Thal, L.J., Petersen, R.C., Jack, C., Jagust, W., Janowski, T., J.Q., Toga, A.W., Beckett, L., 2005a. The Alzheimer's disease neuroimaging initiative. Neuroimaging Clin N Am 15, 869–877.

Mueller, S.G., Weiner, M.W., Thal, L.J., Petersen, J.C.R., Jagust, W., Trojanowski, J.Q., Toga, A.W., Beckett, L., 2005b. Ways toward an early diagnosis in Alzheimer's disease: The Alzheimer's Disease Neuroimaging Initiative (ADNI). Alzheimers Dement 1, Nos. 1, 55–66. J Alzheimer's Assoc 1(1), 55–66.

Nestor, S.M., Rupsingh, R., Borrie, M., Smith, M., Accomazzi, V., Wells, J.L., Fogarty, J., Bartha, R., 2008. Ventricular enlargement as a possible measure of Alzheimer's disease progression validated using the Alzheimer's disease neuroimaging initiative database. Brain 131, 2443–2454.

Petersen, R.C., Robers, R.O., Knopman, D.S., Boeve, B.F., Geda, Y.E., Ivnik, R.J., Smith, G.E., Jack, C.R., 2009. Mild Cognitive Impairment: Ten years later. Arch. Neurol. 66, 1477–1455.

Petersen, R.C., Aisen, P.S., Beckett, L.A., Donohue, M.C., Gamst, A.C., Harvey, D.J., Jack, C.R., Jr, Jagust, W.J., Shaw, L.M., Toga, A.W., Trojanowski, J.Q., Weiner, M.W., The Alzheimer's Disease Neuroimaging Initiative, 2010. Alzheimer's Disease Neuroimaging Initiative (ADNI): clinical characterization. Neurology 74:201–209.

R Development Core Team, 2009. R: A Language and Environment for Statistical Computing. In: R Foundation for Statistical Computing, Vienna, Austria.

Rand, W.M., 1971. Objective criteria for the evaluation of clustering methods. J Am Stat Assoc 66, 846–850.

SAS Institute, Inc. 2004. SAS/STATÒ Version 9.0. Cary, NC: SAS Institute, Inc.

Schwarz, C., Fletcher, E., deCarli, C., Carmichael, O., 2009. Fully-automated white matter hyperintensity detection with anatomical prior knowledge and without FLAIR. Inf Proc Med Imaging 21, 239–251.

Shaw, L.M., Korecka, M., Clark, C.M., Lee, V.M.-Y., Trojanowski, J.Q., 2007. Biomarkers of neurodegeneration for diagnosis and monitoring therapeutics. Nat Rev Drug Discov 6, 295–303.

Shaw, L.M., 2008. PENN Biomarker Core of the Alzheimer's Disease Neuroimaging Initiative. Neuro Signals 16, 19–23.

Shaw, L.M., Vanderstichele, H., Knapik-Czajka, M., Clark, C.M., Aisen, P., Petersen, R.C., Blennow, K., Soares, H., Simon, A., Lewczuk, P., Dean, R., Siemers, E., Potter, W., Lee, V.M.-Y., Trojanowski, J.Q., The Alzheimer's Disease Neuroimaging Initiative, 2009. Cerebrospinal fluid biomarker signature in Alzheimer's Disease Neuroimaging Initiative subjects. Ann Neurol 65, 403–413.

Smith, A.D., 2008. The worldwide challenge of the dementias: a role for B vitamins and homocysteine? Food Nutr Bull 29 suppl, S143.

Stomrud, E., Hansson, O., Blennow, K., Minthon, L., Londos, E., 2007. Cerebrospinal fluid biomarkers predict decline in subjective cognitive function over 3 years in healthy elderly. Dement Geriatr Cogn Disord 24, 118–124.

Ward, J., 1963. Hierarchical Grouping to optimize an objective function. J Am Stat Assoc 58, 236–244.