D-16-00057

Clinical Studies

Structural neuroimaging correlates of cognitive status in older adults: A person-oriented approach

Charles B. Malpas for the Alzheimer's Disease Neuroimaging Initiative

Department of Medicine, Royal Melbourne Hospital, The University of Melbourne,
Grattan Street, Parkville, VIC 3010, Australia

Developmental Imaging, Clinical Sciences Theme, Murdoch Children's Research
Institute, Melbourne, VIC, Australia

*Corresponding author. Tel.: +61 451 116 434.

E-mail address: charles.malpas@unimelb.edu.au (C.B. Malpas).

Abstract

Person-oriented approaches to clinical research aim to uncover subgroups of patients with different patterns of clinically relevant variables. Such approaches, however, are not yet widely employed in clinical neuroimaging research. This paper demonstrates an accessible approach to person-oriented research using model-based clustering in high-dimensional structural neuroimaging data. Cortical thickness measurements for 369 older adults (182 women, 187 men) were obtained from the Alzheimer's Disease Neuroimaging Initiative. Model-based cluster analysis was performed on these imaging variables and then validated using variables that were not used in the clustering process. Variable selection identified two specific regions that contributed to cluster formation: the left and right entorhinal cortices. Two subgroups were uncovered: a "typical" cluster with higher entorhinal thickness ($M = 3.59$ mm, 95% confidence interval = 3.57, 3.62), and an "atypical" cluster with relatively lower thickness ($M = 2.84$ mm, 95% confidence interval = 2.75, 2.92). Members of the atypical cluster also had lower hippocampal volumes, memory scores, and executive function scores, and were also more likely to be clinically classified as cognitively impaired. These findings demonstrate the utility of model-based clustering of structural neuroimaging data in studies of ageing. The role of the entorhinal cortices in cluster formation is consistent with the known pathological substrate of Alzheimer's disease. The entorhinal cortices are implicated in the early genesis of the disease and atrophy of these regions is strongly associated with the cognitive phenotype. Overall, this approach can be readily applied to future neuroimaging investigations.

Keywords: Alzheimer's disease; Cortical thickness; Model-based cluster analysis;
Person-oriented methods; Structural imaging

1. Introduction

Analytical approaches to human inquiry can be classified as either variable or person oriented. The aim of variable-oriented research is to identify relationships between variables, such as the relationship between a clinical variable (for example, diagnostic status) and a neuroimaging marker (for example, grey matter volume). In this approach, the individual is viewed as a “data carrier”, and the interest is in the inter-relations between the variables at the aggregate level [1]. This approach is typical for the analysis of neuroimaging data, where techniques such as group comparisons, correlations, and the general linear model predominate. In contrast, the aim of person-oriented research is to identify patterns or groupings of individuals under the assumption that the study sample might be drawn from multiple populations. Statistical techniques such as cluster analysis are typical for this approach to research.

Person-oriented approaches are yet to be employed extensively in neuroimaging research. One reason for this involves the high-dimensional nature of neuroimaging data. Cortical thickness data, for example, typically contains >100 datapoints for each participant [2]. While there are now techniques readily available to analyse such data in a person-oriented framework, they have yet to permeate into the neuroimaging community. The aim of this paper is to introduce the concept of person-oriented research in neuroimaging, and demonstrate its application to high-dimensional structural imaging data. Readily available software packages are used [3, 4], which make the uptake of this approach relatively straightforward.

Von Eye and Bogat [5] describe three criteria for person-oriented research. First, an underlying assumption of the analysis is that the sample was drawn from more than one population. Statistical techniques are used to identify different groups within the data that are characterised by different patterns of variables. The second criterion involves the validation of the different groupings. To avoid circularity, it is essential to use variables that were not used to form the groupings. In the case of neuroimaging data, phenotypic data are often suitable candidates. Finally, the groupings must be interpreted in the context of a theory. In the case of neuroimaging in neurodegenerative disease, the known pathological substrate provides this context.

This study used data from a large Alzheimer's disease (AD) study, containing participants with varying degrees of cognitive function. The advantage of using this population is that there are likely to be structural changes in some participants, driving cluster formation [6]. The known pathological substrate of the disease provides a framework for interpreting the findings, and the associated measures of cognitive impairment provide a means of validating the cluster solution. As described in detail below, this study used a stepwise variable selection approach to reduce the number of variables used for clustering. Cluster analysis was then performed using a validated model-based approach [7]. These approaches overcome the challenges of high-dimensional data and facilitate the objective selection of the number of clusters to extract [8]. This approach has been successfully applied in the analysis of high-dimensional genetic data [9]. Given the known distribution of pathology in AD, it is predicted that clustering information will be encoded in the mesial temporal, parietal, and posterior cingulate structures [10-12]. Groups with greater atrophy were expected

to have lower scores on cognitive measures, and would be more likely to carry a clinical diagnosis of cognitive impairment.

2. Material and methods

2.1. Participants

The data used in this study were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database. The ADNI was launched in 2003 by the USA National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, the Food and Drug Administration, private pharmaceutical companies and non-profit organisations, as a US\$60 million, 5 year public-private partnership. The primary goal of ADNI has been to test whether serial MRI, positron emission tomography, other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of mild cognitive impairment (MCI) and early AD. Determination of sensitive and specific markers of early AD progression is intended to aid researchers and clinicians to develop new treatments and monitor their effectiveness, as well as lessen the time and cost of clinical trials. ADNI is the result of efforts of many co-investigators from a broad range of academic institutions and private corporations, with subjects recruited from over 50 sites across the USA and Canada. The initial target for ADNI was to recruit 800 adults, age 55 to 90 years, to participate in the research, approximately 200 cognitively normal older individuals to be followed for 3 years, 400 people with MCI to be followed for 3 years and 200 people with early AD to be followed for 2 years (for more information, see www.adni-info.org).

The ADNI database was queried for all participants who underwent structural imaging at the 12 month timepoint. This timepoint was chosen essentially arbitrarily, except with the aim of maximising the number of patients included in the analysis. Only patients with complete demographics, cognitive, and neuroimaging data were included. Cases that failed strict quality control procedures were excluded. The study sample comprised 369 participants (182 women, 187 men). One hundred and nineteen (32.2%) were classified as having no dementia, 223 had questionable dementia (60.4%), 23 had mild dementia (6.2%), and four had moderate dementia (1.1%). Full subject characteristics are shown in Table 1.

2.2. *Neuroimaging data*

Structural neuroimaging data were downloaded from the ADNI database. Cortical thickness estimates were performed by investigators at The University of California, San Francisco, USA (full details of this analysis are available on the ADNI website: <http://adni.loni.usc.edu/methods/mri-analysis>). Non-accelerated T1 structural images (either multiplanar reconstruction or inversion recovery spoiled gradient echo) were analysed following pre-processing that included gradient warping, B1 correction, and N3 inhomogeneity correction. Cortical thickness estimates were obtained using the open source FreeSurfer software (version 5.1) [13]. Quality control procedures were followed, and only metrics that passed were included in the analysis. Mean thickness measurements for 34 cortical regions were extracted for each hemisphere [14]. Hippocampal volume measurements computed by the FreeSurfer subcortical pipeline were also downloaded in order to validate the clustering solution. Estimates of intracranial volume were also obtained from the FreeSurfer output.

2.3. *Clinical variables*

A number of clinical variables were extracted in order to validate the clustering solution and to characterise the resulting subgroups. These include the Mini Mental Status Examination [15], a cognitive screening measure where higher scores indicate better cognitive function. Another screening measure, the Alzheimer's Disease Assessment Scale [16] was also used. On this scale, higher scores indicate greater impairment. The Functional Assessment Questionnaire [17] provided a measurement of impairments in activities of daily living, where higher scores indicate greater impairment. For diagnostic characterisation, the Clinical Dementia Rating scale [18] was used. This clinician-rated scale classifies patients as either having no dementia, questionable dementia, mild dementia, or moderate dementia. Validated composite scores of memory and executive function were used, and are described in detail elsewhere [19, 20].

2.4. *Variable selection and model-based clustering*

Variable selection was performed using the *clustvarsel* package in the R environment [3]. This package implements the variable selection procedure first described by Raftery and Dean [21] and subsequently extended by Maugis and colleagues [22, 23]. Briefly, the algorithm begins with a set of variables and then assesses a new variable for exclusion or removal based on the Bayesian information criteria (BIC). This process is repeated until a final set of variables is selected. The subset of variables was then entered into model-based clustering using the *MClust* package in the R environment [24]. This package performs normal mixture modelling via expectation-maximisation in combination with the BIC to allow for automated detection and extraction of clusters of cases. The optimal number of clusters is

compared to the case of one cluster (that is, no clusters), and cluster membership is saved for follow-up analysis.

2.5. Additional statistical analyses

Secondary variables were analysed to validate the cluster solution. For each variable, a standardised effect size was calculated in the form of Bonett's delta [25]. As with Cohen's *d*, 0.2 is considered a small effect, 0.5 is medium, and 0.8 is large [26]. 95% confidence intervals were also computed. Confidence intervals that do not contain zero can be interpreted as statistically significant at the 5% level. Cross-tabulated data is presented with adjusted standardised residuals representing the difference between the actual and expected cell counts. Adjusted standardised residuals with an absolute value >2 are considered statistically significant at the 5% level. Odds ratios with 95% confidence intervals are presented for binary cross-tabulations.

3. Results

3.1. Cluster solution

The variable selection procedure identified two variables that contributed significantly to cluster membership. These were the left and right entorhinal cortices. Model-based clustering was then performed on these two variables. As shown in Figure 1, the best fitting clustering solution contain two groups (BIC = -745.18).

The cluster solution is shown in Figure 2 with reference to left and right entorhinal cortex thickness. Participants in the first cluster, labelled the “typical” subgroup, had entorhinal cortex thicknesses in the upper range of the data set. This group was

relatively homogeneous, forming a tight cluster. In contrast, participants in the second “atypical” cluster had lower entorhinal thicknesses. This cluster was more heterogeneous.

3.2. Cluster characteristics

Cluster differences on key metrics are shown in Table 2, while Figure 3 shows these differences in the form of standardised effect sizes. Participants in the atypical cluster had much smaller entorhinal cortex thicknesses. This difference was more pronounced on the right than the left. Hippocampal volumes were also lower in the atypical cluster, however there was no difference between left and right volumes. Participants in the atypical cluster were slightly older. They also had lower scores on all cognitive measures. Large differences were observed on measures of global cognitive function, including the Mini Mental Status Examination and Alzheimer’s Disease Assessment Scale. Scores on the memory and executive function composites were also lower, with memory showing the greatest impairment. There was no evidence for a relationship between cluster and sex (odds ratio: 1.30, 95% confidence interval 0.84, 2.02).

Table 3 shows the relationship between cluster membership and clinical classification. There were relatively fewer participants classified as having no dementia in the atypical cluster (17%) compared to the typical cluster (39%). There were relatively more participants with mild dementia in the atypical cluster (16%) compared to the typical cluster (2%). All four participants with moderate dementia were assigned to the atypical cluster.

4. Discussion

This study demonstrated the application of model-based clustering with variable selection in a high-dimensional structural neuroimaging dataset. In a large group of aged participants, two regions were found to discriminate between subgroups, namely the left and right entorhinal cortices. Model-based clustering performed on these variables revealed two subgroups. The first, named the “typical” group, was relatively homogeneous with entorhinal cortical volumes in the upper-range of sample. This cluster had higher scores on all cognitive measures, as well as larger hippocampal volumes. In contrast, the “atypical” group had lower, and more heterogeneous, entorhinal cortex thicknesses. This group had lower scores on all cognitive measures, as well as lower hippocampal volumes. The differences observed on these secondary measures strongly support the validity of the clustering solution.

The contribution of entorhinal cortex thickness to cluster membership is consistent with the known pathological staging of AD. The formation of tau-related neurofibrillary tangles first occurs in transentorhinal cortex, and this region remains the most heavily affected by this pathology throughout the disease [27]. Levels of phosphorylated tau in cerebrospinal fluid has been shown to correlate with reduced functional connectivity in the entorhinal regions in patients with dementia of the Alzheimer’s type (DAT) [28]. The significance of this region in AD is supported by evidence that *in vivo* measurement of entorhinal atrophy allows accurate identification of patients with DAT, and also accurately predicts the evolution of the clinical syndrome from MCI to DAT [29]. The significant differences in hippocampal volume

were also consistent with the pathological substrate of AD. Hippocampal atrophy is observed early in the disease, and has been confirmed by a number of neuroimaging studies [30].

Damage to the transentorhinal cortex results in associative memory impairment. This has been confirmed by animal research documenting associative learning impairments following rhinal lesions [31]. In accordance with the early accumulation of tau-pathology in mesial temporal regions, memory impairment dominates the early clinical picture and is typically the index complaint [32]. The specific role of associative memory function has been confirmed by evidence showing arbitrary associative learning paradigms are particularly sensitive to the early stages of DAT [33]. Cerebrospinal fluid biomarkers of tau-pathology have been associated with memory impairments in DAT [34]. Entorhinal cortex volumes are reduced in people with subjective memory complaints, even in the absence of objective memory change [35]. In patients with more clinically advanced disease, entorhinal volumes have been shown to negatively correlate with measures of memory function [36].

As discussed above, cluster membership was associated with clinical diagnosis. The value of model-based clustering, however, is not the classification into known categories. In the person-oriented framework, the value of this clustering approach is to identify new, and unknown, clusters of individuals. Taken together, these findings satisfy Von Eye and Bogat's [5] criteria for person-oriented research. Subgroups of individuals were identified using an established clustering approach. The subgroups were validated against external variables that were not included in computing the

clustering solution. The clusters were interpreted in the context of theory, which in this study drew on the known pathological substrate of AD.

While this study used structural neuroimaging data to recover subgroups, future work might investigate other imaging modalities. Functional connectivity analysis, in which brain connectivity is conceptualised as a network with a great number of edges, might be particularly appropriate for this cluster-based approach [37]. The application of this approach in longitudinal settings would also be advantageous. For example, subgroups could be identified in patients at the prodromal stage of AD. These subgroups could be tracked to identify patterns of data that are at higher risk of conversion to clinical dementia.

A potential limitation of this work relates to the interpretation of subgroups derived from cluster analysis. For example, the clusters identified in this study were not completely separated. One question that arises is whether these clusters represented real subgroups, or are instead artefacts of the clustering process. The use of model-based clustering with the BIC helps guard against this problem. By testing the fit of the model against a clustering solution with one single cluster, this process helps ensure that multiple clusters are only identified if they are more appropriate than a solution with no clusters (that is, a homogenous population). The validation of the clustering solution by using external variables also minimises this issue. It is also important to note that the approach taken here was one of unsupervised learning. It is not necessary for person-oriented research to recover known or existing classes of individuals. Rather, the aim is to identify subgroups of individuals with similar

patterns of data. In other applications, such subgroups can form the basis of additional hypothesis generation.

5. Conclusions

This study demonstrates the successful application of person-oriented methods to structural neuroimaging data. The combination of variable selection and model-based clustering identified two clusters in a large sample of participants from an AD study. The two regions that survived variable selection were consistent with the known neuropathology of AD, and the clusters differed on key phenotypic variables that were not used to form the groups. Taken together, model-based clustering of structural neuroimaging data is a promising approach that might complement dominant variable-based approaches.

Acknowledgements

Data collection and sharing for this project was funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI; National Institutes of Health Grant U01 AG024904) and Department of Defense ADNI (Department of Defense award number W81XWH-12-2-0012). ADNI is funded by the USA National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: AbbVie, Alzheimer's Association; Alzheimer's Drug Discovery Foundation; Araclon Biotech; BioClinica, Inc.; Biogen; Bristol-Myers Squibb Company; CereSpir, Inc.; Eisai Inc.; Elan Pharmaceuticals, Inc.; Eli Lilly and Company; EuroImmun; F. Hoffmann-La Roche Ltd and its affiliated company Genentech, Inc.; Fujirebio; GE Healthcare; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research & Development, LLC.; Johnson & Johnson Pharmaceutical

Research & Development LLC.; Lumosity; Lundbeck; Merck & Co., Inc.; Meso Scale Diagnostics, LLC.; NeuroRx Research; Neurotrack Technologies; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal Imaging; Servier; Takeda Pharmaceutical Company; and Transition Therapeutics. The Canadian Institutes of Health Research is providing funds to support ADNI clinical sites in Canada. Private sector contributions are facilitated by the Foundation for the National Institutes of Health (www.fnih.org). The grantee organisation is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer's Disease Cooperative Study at the University of California, San Diego, USA. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of Southern California, USA.

References

- [1] Bergman L, Trost K (2006) The Person-Oriented Versus the Variable-Oriented Approach: Are They Complementary, Opposites, or Exploring Different Worlds? *Merrill-Palmer Quarterly* **52**, 601-632.
- [2] Fischl B, Dale A (2000) Measuring the thickness of the human cerebral cortex from magnetic resonance images. *Proceedings of the National Academy of Sciences* **97**, 11050-11055.
- [3] Scrucca L, Raftery A (2014).
- [4] Fraley C, Raftery A, Murphy T, Scrucca L (2012).
- [5] von Eye A, Bogat G (2006) Person-Oriented and Variable-Oriented Research: Concepts, Results, and Development. *Merrill-Palmer Quarterly* **52**, 390-420.
- [6] Markesbery W (2010) Neuropathologic alterations in mild cognitive impairment: A review. *Journal of Alzheimer's Disease* **19**, 221-228.
- [7] Banfield J, Raftery A (1993) Model-Based Gaussian and Non-Gaussian Clustering. *Biometrics* **49**, 803-821.
- [8] Fraley C, Raftery A (2002) Model-Based Clustering, Discriminant Analysis, and Density Estimation. *Journal of the American Statistical Association* **97**, 611-631.
- [9] Yeung K, Fraley C, Murua A, Raftery A, Ruzzo W (2001) Model-Based Clustering and Data Transformation for Gene Expression Data. *Bioinformatics* **17**, 977-987.
- [10] Bakkour A, Morris JC, Dickerson BC (2009) The cortical signature of prodromal AD Regional thinning predicts mild AD dementia. *Neurology* **72**, 1048-1055.

- [11] Dickerson B, Bakkour A, Salat D, Feczko E, Pacheco J, Greve DN, Grodstein F, Wright C, Blacker D, Rosas H, others (2009) The cortical signature of Alzheimer's disease: Regionally specific cortical thinning relates to symptom severity in very mild to mild AD dementia and is detectable in asymptomatic amyloid-positive individuals. *Cerebral Cortex* **19**, 497-510.
- [12] Li C, Wang J, Gui L, Zheng J, Liu C, Du H (2011) Alterations of Whole-Brain Cortical Area and Thickness in Mild Cognitive Impairment and Alzheimer's Disease. *Journal of Alzheimer's Disease* **27**, 281-290.
- [13] Fischl B (2012) FreeSurfer. *Neuroimage* **62**, 774-781.
- [14] Desikan RS, Ségonne F, Fischl B, Quinn BT, Dickerson BC, Blacker D, Buckner RL, Dale AM, Maguire RP, Hyman BT (2006) An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *Neuroimage* **31**, 968-980.
- [15] Folstein MF, Folstein SE, McHugh PR (1975) "Mini-mental state": a practical method for grading the cognitive state of patients for the clinician. *Journal of psychiatric research* **12**, 189-198.
- [16] Rosen WG, Mohs RC, Davis KL (1984) A new rating scale for Alzheimer's disease. *The American journal of psychiatry*.
- [17] Fillenbaum GG, Smyer MA (1981) The development, validity, and reliability of the OARS multidimensional functional assessment questionnaire. *Journal of gerontology* **36**, 428-434.
- [18] Morris JC (1993) The Clinical Dementia Rating (CDR): current version and scoring rules. *Neurology* **43**, 2412-2414.
- [19] Gibbons LE, Carle AC, Mackin RS, Harvey D, Mukherjee S, Insel P, Curtis SM, Mungas D, Crane PK, Initiative AsDN (2012) A composite score for

- executive functioning, validated in Alzheimer's Disease Neuroimaging Initiative (ADNI) participants with baseline mild cognitive impairment. *Brain imaging and behavior* **6**, 517-527.
- [20] Crane PK, Carle A, Gibbons LE, Insel P, Mackin RS, Gross A, Jones RN, Mukherjee S, Curtis SM, Harvey D (2012) Development and assessment of a composite score for memory in the Alzheimer's Disease Neuroimaging Initiative (ADNI). *Brain imaging and behavior* **6**, 502-516.
- [21] Raftery AE, Dean N (2006) Variable selection for model-based clustering. *Journal of the American Statistical Association* **101**, 168-178.
- [22] Maugis C, Celeux G, Martin-Magniette ML (2009) Variable selection for clustering with Gaussian mixture models. *Biometrics* **65**, 701-709.
- [23] Maugis C, Celeux G, Martin-Magniette M-L (2009) Variable selection in model-based clustering: A general variable role modeling. *Computational Statistics & Data Analysis* **53**, 3872-3882.
- [24] Fraley C, Raftery AE, Murphy TB, L S.
- [25] Bonett DG (2008) Confidence intervals for standardized linear contrasts of means. *Psychological Methods* **13**, 99.
- [26] Cohen J (1988) *Statistical power analysis for the behavioural sciences*, Lawrence Erlbaum, New Jersey.
- [27] Juottonen K, Laakso M, Insausti R, Lehtovirta M, Pitkänen A, Partanen K, Soininen H (1998) Volumes of the entorhinal and perirhinal cortices in Alzheimer's disease. *Neurobiology of Aging* **19**, 15-22.
- [28] Malpas C, Saling M, Velakoulis D, Desmond P, O'Brien T (2015) Differential functional connectivity correlates of CSF biomarkers in dementia of the Alzheimer's type. *Neurodegenerative diseases* **In Press**.

- [29] Devanand D, Pradhaban G, Liu X, Khandji A, De Santi S, Segal S, Rusinek H, Pelton G, Honig L, Mayeux R (2007) Hippocampal and entorhinal atrophy in mild cognitive impairment Prediction of Alzheimer disease. *Neurology* **68**, 828-836.
- [30] van der Flier W, Scheltens P (2009) Hippocampal volume loss and Alzheimer disease progression. *Nature Reviews Neurology* **5**, 361-362.
- [31] Buckmaster C, Eichenbaum H, Amaral D, Suzuki W, Rapp P (2004) Entorhinal cortex lesions disrupt the relational organization of memory in monkeys. *The Journal of Neuroscience* **24**, 9811-9825.
- [32] Giannakopoulos P, Herrmann F, Bussiere T, Bouras C, Kövari E, Perl D, Morrison J, Gold G, Hof P (2003) Tangle and neuron numbers, but not amyloid load, predict cognitive status in Alzheimer's disease. *Neurology* **60**, 1495-1500.
- [33] Fowler K, Saling M, Conway E, Semple J, Louis W (2002) Paired associate performance in the early detection of DAT. *Journal of the International Neuropsychological Society*, 58-71.
- [34] Malpas C, Saling M, Velakoulis D, Desmond P, O'Brien T (2015) *Journal of Alzheimer's Disease* **47**, 965-975.
- [35] Jessen F, Feyen L, Freymann K, Tepest R, Maier W, Heun R, Schild H-H, Scheef L (2006) Volume reduction of the entorhinal cortex in subjective memory impairment. *Neurobiology of aging* **27**, 1751-1756.
- [36] Sánchez-Benavides G, Gómez-Ansén B, Molinuevo JL, Blesa R, Monte GC, Buschke H, Peña-Casanova J (2009) Medial temporal lobe correlates of memory screening measures in normal ageing, MCI, and AD. *Journal of geriatric psychiatry and neurology*.

- [37] Malpas C, Genc S, Saling M, Velakoulis D, Desmond P, O'Brien T (In press)
MRI correlates of general intelligence in neurotypical adults. *Journal of
Clinical Neuroscience*.

Tables

Table 1 Participant characteristics

Measure	Mean (95% CI)	Min	Max
Age, years	72.59 (71.88, 73.31)	56	90
MMSE	27.73 (27.44, 28.03)	13	30
ADAS-Cog	8.84 (7.85, 9.12)	1	38
FAQ	3.19 (2.63, 3.75)	0	29
Memory composite	0.53 (0.45, 0.61)	-2.58	2.77
Executive function composite	0.49 (0.40, 0.59)	-2.19	2.57

ADAS-Cog = Alzheimer's Disease Assessment Scale, CI = confidence interval, FAQ = Functional Assessment Questionnaire, max = maximum, min = minimum, MMSE = Mini Mental Status Examination

Table 2 Differences between the “typical” and “atypical” clusters

Measure	Atypical cluster n=116	Typical cluster n=253	Effect size δ
Age at MRI, years	74.11 (72.85, 75.37)	71.96 (71.09, 72.83)	0.32 (0.09, 0.32)
MMSE	26.18 (25.46, 26.91)	28.43 (28.20, 28.67)	0.74 (0.48, 0.99)
ADAS-Cog	12.73 (11.22, 14.24)	6.56 (6.09, 7.02)	0.97 (0.70, 1.24)
FAQ	6.64 (5.29, 7.99)	1.62 (1.20, 2.04)	0.89 (0.62, 1.15)
Memory composite	0.05 (-0.10, 0.20)	0.74 (0.66, 0.82)	0.95 (0.70, 1.19)
Executive function composite	0.08 (-0.11, 0.27)	0.68 (0.59, 0.78)	0.69 (0.43, 0.91)
Left entorhinal cortex, mm	2.83 (2.73, 2.94)	3.50 (3.47, 3.53)	1.57 (1.21, 1.80)
Right entorhinal cortex, mm	2.83 (2.74, 2.93)	3.69 (3.65, 3.72)	2.11 (1.78, 2.44)
Left entorhinal cortex	1.88 (1.80, 1.95)	2.37 (2.33, 2.40)	1.38 (1.11, 1.65)

(normalised to ICV)			1.64)
Right entorhinal cortex (normalised to ICV)	1.87 (1.80, 1.94)	2.49 (2.46, 2.53)	1.82 (1.55, 2.11)
Left hippocampus volume, mm ³	3094.28 (2978.18 ,3210.39)	3674.25 (3614.12, 3734.38)	1.04 (0.80, 1.30)
Right hippocampus volume, mm ³	3140.13 (3022.76, 3257.49)	3741.70 (3681.37, 3802.03)	1.06 (0.81, 1.31)

Data are presented as mean (95% confidence interval).

ADAS-Cog = Alzheimer's Disease Assessment Scale, FAQ = Functional Assessment Questionnaire, ICV = intracranial volume, MMSE = Mini Mental Status Examination.

Table 3 Clinical Dementia Rating by cluster

		Clinical Dementia Rating			
		No dementia	Questionable	Mild	Moderate
Atypical cluster	n (%)	20 (17%)	74 (64%)	18 (16%)	4 (3%)
	ASR	-4.2 *	0.9	5.0 *	3.0 *
Typical cluster	n (%)	99 (39%)	149 (59%)	5 (2%)	0 (0%)
	ASR	4.2 *	-0.9	-5.0 *	-3.0 *

ASR = adjusted standiardised residual (>2 statistically significant at the 5% level indicated with *).

Figure Captions

Fig. 1. BIC for model-based clustering. The coloured lines represent different parameterisations of the covariance matrix. The cluster solution with maximal BIC contained two clusters with VEV (BIC = -745.18), corresponding to an ellipsoidal distribution, variable volume, equal shape, and variable orientation.

BIC = Bayesian information criteria, EEE = ellipsoidal distribution, equal volume, equal shape, equal orientation, EEI = diagonal distribution, equal volume, equal shape, coordinate axes orientation, EEV = ellipsoidal distribution, equal volume, equal shape, variable orientation, EII = spherical distribution, equal volume, equal shape, EVI = diagonal distribution, equal volume, variable shape, coordinate axes distribution, VEI = diagonal distribution, variable volume, equal shape, coordinate axes orientation, VEV = ellipsoidal distribution, variable volume, equal shape, variable orientation, VII = spherical distribution, variable volume, equal shape, VVI = diagonal distribution, variable volume, variable shape, coordinate axes orientation, VVV = ellipsoidal distribution, variable volume, variable shape, variable orientation.

Fig. 2. Clustering solution. Filled squares (green) show cases in the typical cluster, while unfilled squares (red) show cases in the atypical cluster.

Fig. 3. Cluster differences. Absolute standardised effect sizes are shown for key variables.

ADAS-Cog = Alzheimer's Disease Assessment Scale, CI = confidence interval, FAQ = Functional Assessment Questionnaire, MMSE = Mini Mental Status Examination.