Cognitive subtypes of probable Alzheimer's disease robustly identified in four cohorts

N.M.E. Scheltens[1], B.M. Tijms[1], T. Koene[2], F. Barkhof[3,4], C.E. Teunissen[5], S. Wolfsgruber[6], M. Wagner[6], J. Kornhuber[7], O. Peters[8], B.I. Cohn-Sheehy[9], G.D. Rabinovici[9], B.L. Miller[9], J.H. Kramer[9], Ph. Scheltens[1], W.M. van der Flier[1,10];
Alzheimer's Disease Neuroimaging Initiative*, German Dementia Competence Network,
University of California San Francisco Memory and Aging Center,
and Amsterdam Dementia Cohort.

[1] Department of Neurology, Alzheimer Center, Amsterdam Neuroscience, VU University Medical Center, Amsterdam, The Netherlands
[2] Department of Medical Psychology, VU University Medical Center, Amsterdam, The Netherlands
[3] Department of Radiology and Nuclear Medicine, Amsterdam Neuroscience, VU University Medical Center, Amsterdam, The Netherlands
[4] Institutes of Neurology and Healthcare Engineering, UCL, London, UK
[5] Neurochemistry Laboratory and Biobank, Department of Clinical Chemistry, Amsterdam Neuroscience, VU University Medical Centre, Amsterdam, The Netherlands
[6] Department of Psychiatry, University of Bonn, Bonn, Germany, and German Center for Neurodegenerative Diseases, Bonn, Germany
[7] Department of Psychiatry, Friedrich-Alexander-University Erlangen, Erlangen, Germany
[8] Department of Psychiatry, Charité Berlin, Campus Benjamin Franklin, Berlin, Germany
[9] Memory and Aging Center, Department of Neurology, University of California San Francisco, San Francisco, USA
[10] Department of Epidemiology and Biostatistics, VU University Medical Center, Amsterdam, The Netherlands

Title characters incl. spaces:	86	
Abstract word count:	150	150 allowed
Research in context	141	150 allowed
Text word count:	3651	3.500 allowed (excluding the abstract, references, figures, and tables)
Tables:	2	+ 4 supplementary
Figures:	3	+ 4 supplementary
References:	47	50 allowed

ABSTRACT

Introduction Patients with Alzheimer's disease (AD) show heterogeneity in profile of cognitive impairment. We aimed to identify cognitive subtypes in four large AD cohorts using a data-driven clustering approach.

Methods We included probable AD dementia patients from the Amsterdam Dementia Cohort (ADC, n=496), Alzheimer's Disease Neuroimaging Initiative (ADNI, n=376), German Dementia Competence Network (DCN, n=521), and University of California, San Francisco (UCSF, n=589). Neuropsychological data were clustered using Nonnegative Matrix Factorization. We explored clinical and neurobiological characteristics of identified clusters.

Results In each cohort, a two-clusters solution best fitted the data (cophenetic correlation >.9): One cluster was memory-impaired, and the other relatively memory-spared. Pooled analyses showed that the memory-spared clusters (29-52% of patients) were younger, more often APOE e4 negative, and had more severe posterior atrophy compared to the memory-impaired clusters (all *p*<.05).

Conclusions We could identify two robust cognitive clusters in four, independent large cohorts with distinct clinical characteristics.

ABBREVIATIONS ADC = Amsterdam Dementia Cohort, ADNI = Alzheimer's Disease Neuroimaging Initiative, APOE = Apolipoprotein E, CSF = cerebrospinal fluid, DCN = German Dementia Competence Network, MMSE = Mini-Mental State Examination, MRI = magnetic resonance imaging, MTA = medial temporal lobe atrophy, NMF = nonnegative matrix factorization, PA = atrophy of the posterior cortex, UCSF = University of California, San Francisco.

HIGHLIGHTS (max 85 characters including spaces per point)
- Alzheimer's disease (AD) is a heterogeneous disorder (54)
- We identified two cognitive AD subtypes in four cohorts with a data-driven approach (82)
- Non-amnestic AD is associated with distinct neurobiological characteristics (77)

RESEARCH IN CONTEXT

1.Systematic review: Alzheimer's disease (AD) is characterized by cognitive heterogeneity. We searched PubMed for clinical and neurobiological heterogeneity in AD profiles, and for data-driven approaches used to identify AD subtypes based on neuropsychological test scores. Several studies demonstrated the potential of clustering methods to identify cognitive AD subtypes. Identified subtypes showed distinct clinical characteristics. However none of the previous studies tested the generalizability of the cluster solutions, since those results were based on single-cohort studies.

2. Interpretation: In four large AD cohorts, we consistently identified two cognitive clusters using Nonnegative Matrix Factorization of neuropsychological test scores. In each cohort one cluster most

prominently showed memory-impairment, and the other cluster was relatively memory-spared. These clusters were associated with distinct clinical characteristics.

3. Future directions: Future research should aim to further study the underlying biological disease mechanisms that cause a non-memory AD phenotype.

1. Introduction

Alzheimer's disease (AD) dementia is characterized by progressive cognitive impairment in multiple cognitive domains, e.g. memory, language, visuospatial and executive functioning, and attention. Typically, AD is characterized by early and prominent memory loss.[1] A minority of patients has a prominent and relatively focal cognitive presentation, such as logopenic-variant primary progressive aphasia (lvPPA), posterior cortical atrophy (PCA), or a behavioural/dysexecutive subtype.[2-5] Atypical variants have been associated with specific demographic, genetic, and neuroimaging/biomarker findings that are distinct from those of typical amnestic patients (e.g. age at onset, apolipoprotein E [APOE] genotype, distribution of cortical atrophy, hypometabolism, tau deposition, cerebrospinal fluid biomarker concentrations, and pathological findings).[6-10] However, even patients who do not display a defined subtype also show considerable variation in patterns of cognitive impairment. Earlier studies demonstrated the potential to capture cognitive heterogeneity in AD using a data-driven clustering approach.[11-14] Studies differed in sample size, clinical diagnosis of included patients, available neuropsychological test results, available neurobiological characteristics to compare clusters with, and clustering technique. This has resulted in different numbers of clusters, with different cognitive and neurobiological characteristics. Although those studies were clearly suggestive of variability in underlying pathological mechanisms, it is difficult to generalise the findings, since they result from single studies that show considerable variability in patient population and methodological approaches.

In the present study, we aimed to identify cognitive subtypes and to study whether these subtypes could be replicated in three independent AD dementia cohorts. For the identification of cognitive AD subtypes, we used Nonnegative Matrix Factorization (NMF).[15-18] Based on earlier descriptions of cognitive heterogeneity, we expected NMF to identify at least a cluster including patients with typical amnestic AD, and one or more other clusters including patients with non-amnestic features.[15-18]

4

2. Methods

2.1. Patients

We selected AD patients from four large cohorts: The Amsterdam Dementia Cohort (ADC), the Alzheimer's Disease Neuroimaging Initiative (ADNI), the German Dementia Competence Network (DCN), and the University of California, San Francisco Memory and Aging Center research cohort (UCSF). Patients were selected based on 1) clinical diagnosis of probable AD dementia, 2) availability of neuropsychological test results, and 3) Mini-Mental State Examination (MMSE) score >16/30.[19] In the ADC and UCSF cohort, patients with focal presentations lvPPA, PCA, and the behavioural/dysexecutive subtype of probable AD dementia were included, while such subjects were explicitly excluded from participation in the ADNI and DCN studies.

From the ADC we selected 496 patients with probable AD.[20] Patients visited the outpatient memory clinic of the VU University Alzheimer Center between 2008 and 2013. Standard dementia screening included for most patients medical history and medication use, physical and neurological examination, extensive neuropsychological evaluation, screening laboratory tests, APOE genotyping, magnetic resonance imaging (MRI), and lumbar puncture (LP). In the ADC, level of education was defined according to a rating scale ranging from 1 (low, primary school not finished) to 7 (high, university degree).[21] All participants provided written informed consent to use their clinical data for research purposes. The local ethical committee approved the study.

From the ADNI database (adni.loni.usc.edu) we selected 376 probable AD patients. Patients were recruited in over 50 sites across the U.S. and Canada (www.adni-info.org). Standard workup included medical history, physical and neurological examination, extensive neuropsychological evaluation, screening laboratory tests, APOE genotyping, neuroimaging including MRI, and LP. For the present study, we used data of screening and baseline visits, acquired for ADNI-1 or ADNI-2 between 2005 and 2013. All patients gave written informed consent at screening.

From the DCN cohort database (http://www.kompetenznetz-demenzen.de) we selected 521 probable AD patients.[22] The DCN is a collaboration of fourteen specialized German memory clinics from university hospitals. All patients were offered a uniform dementia screening at first visit between 2003 and 2007, including medical history, physical and neurological examination, extensive neuropsychological evaluation, screening laboratory tests, MRI scan, and LP. The DCN study protocol was approved by the Institutional Review Boards of all participating study centres.[22] All patients, or their legal guardians, provided written informed consent.

From the UCSF research cohort we selected 589 probable AD patients.[23] Patients were either seen in the outpatient memory clinic, or for a research assessment in the UCSF Alzheimer's Disease Research Center. All patients were assessed at first visit between 1998 and 2013. Standardized dementia screening included medical history, physical and neurological examination, neuropsychological evaluation, screening laboratory tests, APOE genotyping, and neuroimaging including MRI. A core screening neuropsychological battery was performed in both the clinical and research settings. All patient and informants provided written informed consent. Surrogate consent was accepted when patients lacked capacity to provide consent themselves. The local medical ethical committee approved the study.

2.2 Neuropsychological tests

Neuropsychological data included tests covering the major cognitive domains in each cohort, but the exact composition of neuropsychological test batteries differed across cohorts. Neuropsychological tests included for analysis in this study are shown in supplementary table 1. The number of missing neuropsychological values differed across cohorts and within neuropsychological test batteries (on average 20% in ADC, 27% in ADNI, 1% in DCN, 12% in UCSF). Main reasons for missingness are practical reasons unrelated to the data (random). In part of the cases however, tests could not be finished due to cognitive impairment, whereas scoring differed across cohorts and between tests (i.e. assignment of missing value or minimum score). The clustering technique NMF does not allow for missing data or negative values. In order to reduce selection bias, we did not select patients based on completeness of datasets, but we completed the datasets using a multiple imputation approach that is commonly used as a reliable method to estimate missing data.

We imputed missing neuropsychological data using R package Multivariate Imputation by Chained Equations (MICE, version 2.25).[24,25] MICE estimates missing neuropsychological values by predicting these values from the relationships with other neuropsychological variables. We also included predictors age, gender, MMSE, and when available education, duration of complaints, Alzheimer's Disease Assessment Scale – Cognitive subscale (ADAS-cog) or Cambridge Cognitive Examination (CAMCOG) or Cognitive Dementia Rating (CDR) sum of boxes in the imputation model. We ran MICE 50 times per cohort, resulting in 50 imputed datasets for each cohort. For further analyses, we included pooled measures over the 50 derived imputed datasets per cohort.[26,27] We inverted values when appropriate, so that for all tests lower scores reflect worse cognitive impairment. Next, the imputed neuropsychological data were normalized and scaled to include only positive values (0-1).

2.3 MRI characteristics

MRI characteristics were available for patients from the ADC, ADNI, and DCN cohorts.

For the ADC cohort, imaging data was obtained on a 1.5T or 3T scan. Visual ratings of medial temporal lobe atrophy (MTA; range 0-4[28]), posterior atrophy (PA; range 0-3[29]), and white matter hyperintensities (WMH; range 0-3[30]) were performed by an experienced neuroradiologist. For ADNI, a structural MRI 1.5T scan was performed on screening or baseline visit.[31] Image processing has been done with cortical reconstruction and volumetric segmentations using FreeSurfer 4.3 (surfer.nmr.mgh.harvard.edu/).[32] Hippocampal and WMH volumes were downloaded from the ADNI website (ida.loni.usc.edu/).[33] For the DCN cohort, MRI data was obtained from multiple 1.5T scanners with standardized MRI acquisition guidelines across centres.[22,34] PA and WMH were scored by experienced neuroradiologists using a visual rating scale in which higher scores reflect more severe abnormalities (range 0-3). Furthermore, hippocampal volumes were measured using FMRIB's Integrated Registration and Segmentation Tool (FIRST) from the FMRIB Software Library (FSL) package of tools.[35,36] For all atrophy measurements, we included the mean of left and right hemisphere.

2.4 APOE e4 genotype

APOE e4 genotype was available in the ADC (n=448; 90%), ADNI (n=196; 52%), DCN (n=397; 76%), and UCSF (n=175; 30%) cohorts. We dichotomised APOE e4 genotype according to the presence or absence of one or more APOE e4 alleles. APOE genotype was assessed using the Light Cycler APOE mutation detection method (Roche Diagnostics GmbH, Mannheim, Germany) in the ADC. For ADNI, APOE alleles were genotyped using DNA extracted by Cogenics. For the DCN, APOE genotype was assessed using Qiagen blood isolation kit (Qiagen, Hilden, Germany). For UCSF, APOE genotype was conducted using a TaqMan Allelic Discrimination Assay on an ABI 7900HT Fast Real-Time PCR system (Applied Biosystems, Foster City, CA, USA).

2.5 AD biomarkers

Cerebrospinal fluid (CSF) markers amyloid-beta$_{1-42}$ (Abeta$_{1-42}$), and total tau (tau) were available for ADC (n=389; 79%), ADNI (n=102; 27%), and DCN (n=193; 37%). CSF biomarkers were assessed using Sandwich ELISAs (Fujirebio, Gent, Belgium) in ADC and DCN[37], and Multiplex xMAP Luminex platform (Luminex Corp, Austin, TX) with Fujirebio immunoassay kit–based reagents (INNO-BIA Alzbio3; Fujirebio, Ghent, Belgium) in ADNI. CSF biomarkers were considered positive for AD when tau / Abeta$_{1-42}$ ratio was > 0.52 [38]. Pittsburgh compound B position emission tomography visual reading results (i.e. positive or negative) were available for the UCSF cohort (n=52; 9%).

2.6 Statistical analysis

For statistical analyses, we used RStudio for Mac version 3.2.2 (Integrated Development for R. RStudio, Inc., Boston, MA, http://www.rstudio.com). Clustering was performed with the R package

Nonnegative Matrix Factorization (NMF, version 0.20.6).[39] Clustering is a data-driven method to divide a heterogeneous set of objects (patients) in subgroups that are more homogeneous in terms of characteristics provided as input for the clustering (neuropsychological test results). NMF is a dual-clustering approach, meaning that clustering includes two parallel steps (illustrated in figure 1). Firstly, neuropsychological test results are grouped together into neuropsychological profiles ('components'), and neuropsychological tests that determine these components can be identified by their component loading. Secondly, patients are grouped together based on the fit of their neuropsychological profile to the neuropsychological summary component. The optimal number of clusters is based on the most consistent assignment of patients to the identified cognitive profiles in the multiple runs of NMF. The stability of the cluster solution is assessed with the cophenetic correlation coefficient, that measures how consistently tests and subjects are assigned to a given component, ranging between 0 and 1 (i.e. no stability to stable cluster solution). For each cohort, we determined the optimal number of clusters based on the highest cophenetic correlation coefficient for 2 to 9 cluster solutions.[15] We used the 'nonsmooth' NMF algorithm that introduces an intermediate smoothing matrix to enhance sparsity of the clusters.[40]

Characterization of identified clusters in terms of neuropsychological profile was based on the most strongly loading neuropsychological tests.[17] For characterization of identified clusters in terms of demographic, clinical, and neurobiological characteristics, we analysed age, sex, education, disease duration reported by the patient, MMSE, APOE e4 genotype, CSF biomarkers, MRI atrophy, and WMH measurements using χ^2, t-tests, or Kruskal Wallis tests where appropriate. These analyses were performed for each cohort separately. In addition, for each cluster we pooled the patient characteristics over the cohorts in order to compare them for the total sample. To this end we Z-transformed variables with different scales (i.e. education, CSF biomarkers, and MRI biomarkers) before pooling. When atrophy of the hippocampus and the posterior cortex was measured using a visual rating scale in which higher scores reflect more severe atrophy, the normalized scores were inverted so that higher scores reflect less atrophy in all cohorts and the pooled sample.

3. RESULTS

3.1 Cohort characteristics

Characteristics of all cohorts are summarized in table 1. On average, patients were 71±9 years old, with the ADC being the youngest cohort. 54% of patients were female, ranging from 44% (ADNI) to 60% (DCN). Patients were mildly-to moderately demented, with an average MMSE varying between cohorts from 22-24. Roughly two thirds of patients were APOE e4 carrier, ranging from 57% (UCSF) to 67% (ADC).

3.2 NMF clusters of cognitive subtypes

NMF is a dual-clustering approach; first, neuropsychological tests are grouped into components, and second patients are clustered based on the fit of their neuropsychological profiles to the profiles of the identified neuropsychological components, taking the load of each test to the component into account. NMF showed that within each cohort, the optimal number of clusters was two, as the solution with two clusters showed the strongest cophenetic correlation (>.90).

Results of the clustering of tests are shown in figure 2. In the ADC, one neuropsychological component mainly included memory tests Rey auditory verbal learning test (RAVLT) immediate and delayed recall. The other component included mainly non-memory tests trail making test (TMT)-A and TMT-B, and fragmented letters. In the ADNI cohort, one neuropsychological component mainly included memory tests logical memory immediate and delayed recall, and RAVLT delayed recall. The other component mainly included non-memory tests TMT-A and TMT-B. In the DCN cohort, one neuropsychological component included mainly memory tests logical memory immediate and delayed recall, word list immediate and delayed recall of the Consortium to Establish a Registry for Alzheimer's Disease (CERAD), and Rey figure recall. The other component included mainly non-memory tests TMT-A, TMT-B, clock drawing, and CERAD figure copy. In the UCSF cohort, one neuropsychological component included mainly memory tests figure recall, and the California verbal learning test (CVLT) delayed recall. The other component included mainly non-memory tests modified trails, design fluency, and stroop interference.

Patients were assigned to either of two clusters based on the fit of their neuropsychological test results to the memory or non-memory component in each cohort (figure 3). Across all cohorts, the memory clusters included on average 60% of patients, ranging from 48% (ADNI) to 71% (ADC), and the non-memory clusters included on average 40% of patients, ranging from 29% (ADC) to 52% (ADNI).

3.3 Cluster characterisation

9

We analysed cluster characteristics in terms of demographical, clinical and neurobiological characteristics. Associations for each cohort are shown in table 2. In the ADC, patients in the non-memory cluster were younger, had lower MMSE scores, and had more severe atrophy of the posterior cortex. In the ADNI cohort, non-memory patients tended to be younger and less often APOE e4 positive with relative hippocampal sparing, but these differences did not reach significance. In the DCN cohort, patients in the non-memory cluster had more severe atrophy of the posterior cortex. In the UCSF cohort, the non-memory cluster had lower MMSE scores, and patients were less often APOE e4 positive. There were no differences in WMH (ADC, ADNI, DCN). When we analysed pooled data, we found that across cohorts, patients in non-memory clusters were younger, less educated, reported shorter disease duration, had lower MMSE scores, were less often APOE e4 positive, had less severe hippocampal atrophy, and more severe atrophy of the posterior cortex than patients in memory clusters.

3.4 Validation of cognitive clusters after stratification for disease severity

We tested whether disease severity influenced the cluster solutions by repeating NMF analyses in each cohort after stratifying for disease severity according to cohort median MMSE value. Results appeared to be robust across severity subgroups, with a memory cluster and non-memory cluster appearing in both the mild and the moderately demented patient strata (supplementary figures 1-4). Cluster characteristics in terms of demographic and neurobiological characteristics remained largely the same (supplementary tables 2 and 3). Differences between the memory and non-memory clusters were most pronounced in the mildly demented stratum.

3.5 Validation of cognitive clusters in AD biomarker confirmed patients

Availability of AD biomarkers is given in table 1 and 2. Based on χ^2 analyses, no differences were found between the memory and non-memory clusters in terms of AD biomarker positivity ($p > 0.05$). Only for the ADC, enough data were available to repeat data-driven NMF analyses in AD biomarker confirmed patients (n=357 with CSF total tau/amyloid $\beta_{1-42} > 0.52$[38]). Results appeared to be consistent with findings of the total cohorts, with a memory component mainly including RAVLT immediate and delayed recall, and a non-memory component mainly including TMT-A, TMT-B, fragmented letters, and letter digit substitution test (LDST), shown in supplementary figure 5. The memory cluster included 73% of patients, and the non-memory cluster 27% (supplementary figure 6). Consistent with characteristics of the total ADC, the non-memory cluster of the biomarker confirmed subset was younger (64 ± 8 versus 67 ± 8 years old, $p < 0.01$), had lower MMSE scores (20.9 ± 3.1 versus 22.7 ± 3.0, $p < 0.001$), and had more severe posterior atrophy (1.51 ± 0.82 versus 1.13 ± 0.71, $p < 0.05$, rated according to a visual rating scale in which higher scores reflect more severe atrophy [29]) than the memory cluster. In addition, the non-memory cluster was less often APOE e4 positive than the memory cluster (61% versus 74%, $p < 0.05$). Results are provided in supplementary table 4.

4. DISCUSSION

Across four independent AD dementia cohorts, we robustly found two cognitive clusters using a data-driven dual-clustering approach. One cluster was characterized by more prominent memory impairment, and one cluster by more prominent impairment on non-memory tests. These memory and non-memory AD phenotypes were consistently found across cohorts, even though these cohorts differed in their patient populations (e.g. age, disease severity, mono-centre versus multi-centre, geographic location) and composition and extensiveness of neuropsychological test battery. Moreover, the clusters were associated with specific demographic, clinical, and neurobiological characteristics. These findings demonstrate the biological relevance of clinical heterogeneity in AD, as this may reflect variation in underlying disease mechanisms.

Of all included patients, 40% belonged to a non-memory cluster. Compared to the memory clusters, non-memory cluster patients were younger, less educated, more often APOE e4 negative, had more severe posterior atrophy and relatively spared hippocampi. In addition, patients assigned to the non-memory clusters reported shorter disease duration, but they had on average lower MMSE scores. Possibly, non-memory clusters are associated with a more aggressive disease progression.[41,42] Future studies should address the question whether cognitive subtypes are related to rate of decline, and hence suitable as putative prognostic marker.

Among the suggested disease mechanisms causing heterogeneity, is the influence of copathologies, e.g. vascular pathology. We did not find differences in severity of ischemic vascular pathology (WMH) however. Of note, cognitive heterogeneity is most prominent in early onset AD patients, where AD pathology is often pure, and copathologies are less present.[43] Biomarker support for AD pathology was not available for each patient, but the available data showed no difference in AD biomarker positivity between clusters, suggesting that misdiagnosis is not a major driver of our findings. Variation in disease mechanisms could also be sought in the origin and spreading of neurofibrillary tangles (typically characterized by origin in the entorhinal cortex, progressing through the hippocampus to the association cortex, and finally to the cortex [44]) since the medial temporal lobe was relatively spared and the posterior cortex most prominently affected in the non-memory phenotype. This idea fits with the hypothesis that early-onset, APOE e4 negative AD patients are predisposed to vulnerability of cerebral networks beyond the medial temporal lobe.[44,45] This hypothesis coincides with an autopsy study that identified an AD subtype with relative less tangles in the hippocampus that was associated with younger age at death, male sex, rapid disease progression, and more often focal cortical clinical syndromes.[46] However, this phenotype only tended to have less often an APOE e4 positive genotype (p=0.067). Probably, the APOE e4 allele is not the only AD risk factor that modifies clinical heterogeneity; the effect of genetic risk factors on clinical AD

11

phenotype is still poorly understood. An increasing number of promising genes however are discovered to be associated with higher or reduced risks for developing AD and with certain neuropathological pathways.[47]

A limitation of our study is that we did not have post mortem pathological confirmation of patients, so we cannot rule out the possibility of misdiagnosis. We think however that misdiagnoses will not have driven our results, as in all cohorts diagnoses were made according to careful application of clinical criteria, and repeating analyses in the CSF biomarker confirmed subset of ADC gave similar results. Furthermore, the present study could be biased since we excluded patients that were already severely demented (i.e. MMSE < 17) at diagnosis. We think that this selection bias could have resulted in underrepresentation of the non-memory phenotype, because atypical AD variants are less easily recognized as AD (especially at younger age, when other causes for cognitive complaints such as depression or burn-out are more common) and therefore probably more often associated with patients' and doctors' delay, and delay because of initial misdiagnoses. In addition, non-memory patients reporter shorter disease duration, while MMSE scores were already lower, possibly due to faster disease progression before diagnosis, suggesting higher risk for more severe dementia at time of diagnosis.

It could be argued that the substantial differences between cohorts in patient population (e.g. geographically, age, disease severity, degree of cognitive heterogeneity within cohorts), and in extensiveness of neuropsychological test battery could be a limitation. However, we see this as a strong point of our study since we were able to replicate our finding of two robust clusters with their corresponding clinical characterisation. This suggests that the clusters we identified are generalizable to other AD populations; an often-encountered limitation of data driven methods to cluster patients is that of limited generalizability.

Non-memory cluster patients had on average lower MMSE scores, and therefore we performed additional analyses to study whether clustering has been driven by disease severity. Repeating the clustering after stratification based on MMSE scores, a memory and a non-memory cluster were identified in each stratified cohort as well. Cluster differences in terms of clinical characteristics were largely similar for the strata, albeit more pronounced in the mildly demented stratum. This suggests that clinical heterogeneity is more prominently present in early stages of AD.

Our results emphasize the presence of non-memory phenotypes in AD. Being able to identify AD subtypes is important in a clinical setting for early and adequate diagnosis and personalized medicine. Also, cognitive profiling should be taken into account when including patients for clinical trials, or when choosing cognitive outcomes to analyse the effect of an intervention. Furthermore, the existence

of these clusters, with similar patient characteristics across independent cohorts suggests that cognitive heterogeneity is caused by different disease mechanisms.

In conclusion, we found two robust AD subtypes using a data-driven clustering approach in four AD cohorts. Identified clusters were associated with distinct demographical, clinical, and neurobiological characteristics, emphasizing the presence of cognitive heterogeneity in AD, and suggesting variation in underlying pathology.

REFERENCES

[1] McKhann GM, Knopman DS, Chertkow H, Hyman BT, Jack CR, Kawas CH, et al. The diagnosis of dementia due to Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. Alzheimers Dement, vol. 7, 2011, pp. 263–9. doi:10.1016/j.jalz.2011.03.005.

[2] Gorno-Tempini ML, Hillis AE, Weintraub S, Kertesz A, Mendez M, Cappa SF, et al. Classification of primary progressive aphasia and its variants. Neurology, vol. 76, Lippincott Williams & Wilkins; 2011, pp. 1006–14. doi:10.1212/WNL.0b013e31821103e6.

[3] Benson DF, Davis RJ, Snyder BD. Posterior cortical atrophy. Arch Neurol 1988;45:789–93.

[4] Crutch SJ, Schott JM, Rabinovici GD, Boeve BF, Cappa SF, Dickerson BC, et al. Shining a light on posterior cortical atrophy. Alzheimers Dement, vol. 9, 2013, pp. 463–5. doi:10.1016/j.jalz.2012.11.004.

[5] Ossenkoppele R, Pijnenburg YAL, Perry DC, Cohn-Sheehy BI, Scheltens NME, Vogel JW, et al. The behavioural/dysexecutive variant of Alzheimer's disease: clinical, neuroimaging and pathological features. Brain 2015;138:2732–49.

[6] Van Der Flier WM, Schoonenboom SNM, Pijnenburg YAL, Fox NC, Scheltens P. The effect of APOE genotype on clinical phenotype in Alzheimer disease. Neurology 2006;67:526–7. doi:10.1212/01.wnl.0000228222.17111.2a.

[7] van der Vlies AE, Pijnenburg YAL, Koene T, Klein M, Kok A, Scheltens P, et al. Cognitive impairment in Alzheimer's disease is modified by APOE genotype. Dement Geriatr Cogn Disord 2007;24:98–103. doi:10.1159/000104467.

[8] Smits LL, Tijms BM, Benedictus MR, Koedam ELGE, Koene T, Reuling IEW, et al. Regional atrophy is associated with impairment in distinct cognitive domains in Alzheimer's disease. Alzheimers Dement 2014;10:S299–305. doi:10.1016/j.jalz.2013.06.007.

[9] Ossenkoppele R, Cohn-Sheehy BI, La Joie R, Vogel JW, Möller C, Lehmann M, et al. Atrophy patterns in early clinical stages across distinct phenotypes of Alzheimer's disease. Hum Brain Mapp 2015;36:4421–37. doi:10.1002/hbm.22927.

[10] Ossenkoppele R, Mattsson N, Teunissen CE, Barkhof F, Pijnenburg Y, Scheltens P, et al. Cerebrospinal fluid biomarkers and cerebral atrophy in distinct clinical variants of probable Alzheimer's disease. Neurobiol Aging 2015;36:2340–7. doi:10.1016/j.neurobiolaging.2015.04.011.

[11] Stopford CL, Snowden JS, Thompson JC, Neary D. Variability in cognitive presentation of Alzheimer's disease. Cortex 2008;44:185–95. doi:10.1016/j.cortex.2005.11.002.

[12] Davidson JE, Irizarry MC, Bray BC, Wetten S, Galwey N, Gibson R, et al. An exploration of cognitive subgroups in Alzheimer's disease. J Int Neuropsychol Soc 2010;16:233–43. doi:10.1017/S1355617709991160.

[13] Libon DJ, Drabick DAG, Giovannetti T, Price CC, Bondi MW, Eppig J, et al. Neuropsychological syndromes associated with Alzheimer's/vascular dementia: a latent class analysis. J Alzheimers Dis 2014;42:999–1014. doi:10.3233/JAD-132147.

[14] Scheltens NME, Galindo-Garre F, Pijnenburg YAL, van der Vlies AE, Smits LL, Koene T, et al. The identification of cognitive subtypes in Alzheimer's disease dementia using latent class analysis. J Neurol Neurosurg Psychiatr 2015:jnnp–2014–309582. doi:10.1136/jnnp-2014-309582.

[15] Brunet J-P, Tamayo P, Golub TR, Mesirov JP. Metagenes and molecular pattern discovery using matrix factorization. Proc Natl Acad Sci USa 2004;101:4164–9. doi:10.1073/pnas.0308531101.

[16] Wang Y-X, Zhang Y-J. Nonnegative Matrix Factorization: A Comprehensive Review. Knowledge and Data Engineering, IEEE Transactions on 2013;25:1336–53. doi:10.1109/TKDE.2012.51.

[17] Carmona-Saez P, Pascual-Marqui RD, Tirado F, Carazo JM, Pascual-Montano A. Biclustering of gene expression data by Non-smooth Non-negative Matrix Factorization. BMC Bioinformatics 2006;7:78. doi:10.1186/1471-2105-7-78.

[18]   Shokrollahi M, Krishnan S, Dopsa DD, Muir RT, Black SE, Swartz RH, et al. Nonnegative matrix factorization and sparse representation for the automated detection of periodic limb movements in sleep. Med Biol Eng Comput 2016:1–14. doi:10.1007/s11517-015-1444-y.

[19]   Folstein MF, Folstein SE, McHugh PR. "Mini-mental state." A practical method for grading the cognitive state of patients for the clinician. J Psychiatr Res 1975;12:189–98.

[20]   van der Flier WM, Pijnenburg YAL, Prins N, Lemstra AW, Bouwman FH, Teunissen CE, et al. Optimizing patient care and research: the Amsterdam Dementia Cohort. J Alzheimers Dis 2014;41:313–27. doi:10.3233/JAD-132306.

[21]   Verhage F. Intelligence and religious persuasion. Ned Tijdschr Psychol 1964;19:247–54.

[22]   Kornhuber J, Schmidtke K, Frolich L, Perneczky R, Wolf S, Hampel H, et al. Early and differential diagnosis of dementia and mild cognitive impairment: design and cohort baseline characteristics of the German Dementia Competence Network. Dement Geriatr Cogn Disord 2009;27:404–17. doi:10.1159/000210388.

[23]   Kramer JH, Jurik J, Sha SJ, Rankin KP, Rosen HJ, Johnson JK, et al. Distinctive neuropsychological patterns in frontotemporal dementia, semantic dementia, and Alzheimer disease. Cogn Behav Neurol 2003;16:211–8.

[24]   Van Buuren S, Groothuis-Oudshoorn K. MICE: Multivariate Imputation by Chained Equations in R. Journal of Statistical Software 2011;45:1–67.

[25]   Van Buuren S. Flexible Imputation of Missing Data. vol. 20125245. CRC Press; 2012. doi:10.1201/b11826.

[26]   Rubin DB. Multiple Imputation for Nonresponse in Surveys. Hoboken, NJ, USA: John Wiley & Sons; 2009. doi:10.1002/9780470316696.

[27]   Marshall A, Altman DG, Holder RL, Royston P. Combining estimates of interest in prognostic modelling studies after multiple imputation: current practice and guidelines. BMC Med Res Methodol 2009;9:57. doi:10.1186/1471-2288-9-57.

[28]   Scheltens P, Launer LJ, Barkhof F, Weinstein HC, van Gool WA. Visual assessment of medial temporal lobe atrophy on magnetic resonance imaging: interobserver reliability. J Neurol 1995;242:557–60.

[29]   Koedam ELGE, Lehmann M, van der Flier WM, Scheltens P, Pijnenburg YAL, Fox N, et al. Visual assessment of posterior atrophy development of a MRI rating scale. Eur Radiol 2011;21:2618–25. doi:10.1007/s00330-011-2205-4.

[30]   Fazekas F, Chawluk JB, Alavi A, Hurtig HI, Zimmerman RA. MR signal abnormalities at 1.5 T in Alzheimer's dementia and normal aging. AJR Am J Roentgenol 1987;149:351–6. doi:10.2214/ajr.149.2.351.

[31]   Jack CR, Bernstein MA, Fox NC, Thompson P, Alexander G, Harvey D, et al. The Alzheimer's Disease Neuroimaging Initiative (ADNI): MRI methods. J Magn Reson Imaging 2008;27:685–91. doi:10.1002/jmri.21049.

[32]   Fischl B. FreeSurfer. Neuroimage 2012;62:774–81. doi:10.1016/j.neuroimage.2012.01.021.

[33]   Schwarz CG, Fletcher E, DeCarli CS, Carmichael O. Fully-Automated White Matter Hyperintensity Detection With Anatomical Prior Knowledge and Without FLAIRs. 2013.

[34]   Ewers M, Teipel SJ, Dietrich O, Schönberg SO, Jessen F, Heun R, et al. Multicenter assessment of reliability of cranial MRI. Neurobiol Aging 2006;27:1051–9. doi:10.1016/j.neurobiolaging.2005.05.032.

[35]   Patenaude B, Smith SM, Kennedy DN, Jenkinson M. A Bayesian model of shape and appearance for subcortical brain segmentation. Neuroimage 2011;56:907–22. doi:10.1016/j.neuroimage.2011.02.046.

[36]   Smith SM, Jenkinson M, Woolrich MW, Beckmann CF, Behrens TEJ, Johansen-Berg H, et al. Advances in functional and structural MR image analysis and implementation as FSL. Neuroimage 2004;23 Suppl 1:S208–19. doi:10.1016/j.neuroimage.2004.07.051.

[37]   Lewczuk P, Kornhuber J, Wiltfang J. The German Competence Net Dementias: standard operating procedures for the neurochemical dementia diagnostics. J Neural Transm (Vienna) 2006;113:1075–80. doi:10.1007/s00702-006-0511-9.

[38]   Duits FH, Teunissen CE, Bouwman FH, Visser PJ, Mattsson N, Zetterberg H, et al. The cerebrospinal fluid "Alzheimer profile": easily said, but what does it mean? Alzheimers Dement 2014;10:713–723.e2. doi:10.1016/j.jalz.2013.12.023.

[39]   Gaujoux R, Seoighe C. A flexible R package for nonnegative matrix factorization. BMC Bioinformatics 2010;11:367. doi:10.1186/1471-2105-11-367.

[40]   Pascual-Montano A, Carazo JM, Kochi K, Lehmann D, Pascual-Marqui RD. Nonsmooth nonnegative matrix factorization (nsNMF). IEEE Trans Pattern Anal Mach Intell 2006;28:403–15. doi:10.1109/TPAMI.2006.60.

[41]   van der Vlies AE, Koedam ELGE, Pijnenburg YAL, Twisk JWR, Scheltens P, Van Der Flier WM. Most rapid cognitive decline in APOE epsilon4 negative Alzheimer's disease with early onset. Psychol Med 2009;39:1907–11. doi:10.1017/S0033291709005492.

[42]   Smits LL, Pijnenburg YAL, van der Vlies AE, Koedam ELGE, Bouwman FH, Reuling IEW, et al. Early onset APOE E4-negative Alzheimer's disease patients show faster cognitive decline on non-memory domains. Eur Neuropsychopharmacol 2015;25:1010–7. doi:10.1016/j.euroneuro.2015.03.014.

[43]   van der Flier WM. Clinical heterogeneity in familial Alzheimer's disease. Lancet Neurol 2016;15:1296–8. doi:10.1016/S1474-4422(16)30275-7.

[44]   Braak H, Braak E. Neuropathological stageing of Alzheimer-related changes. Acta Neuropathol 1991;82:239–59.

[45]   van der Flier WM, Pijnenburg YA, Fox NC, Scheltens P. Early-onset versus late-onset Alzheimer's disease: the case of the missing APOE ε4 allele. Lancet Neurol 2011;10:280–8. doi:10.1016/S1474-4422(10)70306-9.

[46]   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. doi:10.1016/S1474-4422(11)70156-9.

[47]   Lambert JC, Ibrahim-Verbaas CA, Harold D, Naj AC, Sims R, Bellenguez C, et al. Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. Nat Genet 2013;45:1452–8. doi:10.1038/ng.2802.

Table 1 Demographical and neurobiological characteristics of study cohorts

	ADC	ADNI	DCN	UCSF	Pooled sample
	n = 496	n = 376	n = 521	n = 589	n = 1.982
Demographics					
Age (yr)	67 ± 8	75 ± 8	72 ± 8	71 ± 10	71 ± 9
Female	258 (52%)	165 (44%)	313 (60%)	328 (56%)	1064 (54%)
Education[*]	5 (4-6)	16 (13-18)	11 (10-13)	16 (14-18)	0.00 ± 1.00
Duration complaints (yr)	2 (2-4)	-	2 (1-3)	-	2 (1-4)
Global cognition					
MMSE	22 ± 3	23 ± 2	23 ± 3	24 ± 4	23 ± 3
APOE e4 genotype					
APOE e4 positive	299 (67%)	128 (65%)	255 (64%)	100 (57%)	782 (64%)
AD biomarkers [†]					
AD biomarker available	389 (79%)	102 (27%)	193 (37%)	53 (9%)	737 (37%)
AD biomarker positive	358 (92%)	80 (78%)	164 (85%)	52 (98%)	654 (89%)
MRI					
Hippocampus[‡]	1.5 (1-2)	2882 ± 511	2933 ± 473	-	0.00 ± 1.00
Posterior cortex[§]	1 (1-2)	-	1 (0-2)	-	0.00 ± 1.00
WMH	1 (0-2)	0.41 (0.16-1.25)	0.5 (0-1)	-	0.02 ± 1.01

Data are presented in mean ± standard deviation, number (%), or median (2nd-4th quantile). Abbreviations: ADC = Amsterdam Dementia Cohort, ADNI = Alzheimer's Disease Neuroimaging Initiative, APOE = Apolipoprotein E, c.o. = cut-off, CSF = cerebrospinal fluid, DCN = German Dementia Competence Network, MMSE = Mini-Mental State Examination, UCSF = University of California – San Francisco, WMH = white matter hyperintensities. [*]Education is given according to the Verhage scale (1-7 resp. low-high education[21]) for ADC, years of education for ADNI, DCN, and UCSF, and normalized scores for the pooled sample. [†]AD biomarkers are available as cerebrospinal fluid total tau/amyloid β_{1-42} (abnormal when > 0.52 according to Duits e.a. [38]) in the ADC ADNI, and DCN cohorts, or as Pittsburgh compound B positron emission tomography positivity in the UCSF cohort. [‡]Hippocampal atrophy is measured according to the medial temporal lobe (MTA) visual rating scale for the ADC (0-4, higher scores reflect more severe atrophy [28]), and hippocampal volumes in mm^3 for ADNI and DCN, and z-scores (in which normalized MTA scores are inverted) for the pooled sample. [§]Posterior atrophy is scored using a visual rating scale for the ADC [29] and the DCN, inverted z-scores are given for the pooled sample. WMH are scored according to a visual rating scale for the ADC [30] and the DCN (0-3, higher scores reflect more sever pathology), WMH volumes for ADNI [33], and z-scores are given for the pooled sample.

Table 2 Demographical and neurobiological cluster characteristics

	ADC		ADNI		DCN		UCSF		Pooled sample	
	mem	**non-mem**	**mem**	**non-mem**	**mem**	**non-mem**	**mem**	**non-mem**	**mem**	**non-mem**
	n = 352	n =144	n = 182	n = 194	n = 335	n = 186	n = 326	n = 263	n = 1195	n = 787
	(71%)	(29%)	(48%)	(52%)	(64%)	(36%)	(55%)	(45%)	(60%)	(40%)
Demographics										
Age (years)	**67.7 ± 7.8**	**64.7 ± 8.4**[*]	76.2 ± 7.2	74.1 ± 8.3	72.3 ± 7.8	71.1 ± 9.0	72.1 ± 9.7	70.3 ± 11.0	**71.5 ± 8.7**	**70.4 ± 9.9**[‡]
Female	180 (51%)	78 (54%)	78 (43%)	87 (45%)	201 (60%)	112 (60%)	171 (52%)	157 (60%)	630 (53%)	434 (55%)
Education	0.03 ± 1.00	-0.06 ± 1.01	0.01 ± 0.97	-0.01 ± 1.03	0.03 ± 1.00	-0.05 ± 1.00	0.1 ± 1.00	-0.13 ± 0.99	**0.04 ± 0.99**	**-0.07 ± 1.01**[‡]
Duration complaints	3.1 ± 2.2	2.7 ± 1.7	-	-	2.7 ± 2.3	2.4 ± 1.8	-	-	**2.9 ± 2.3**	**2.5 ± 1.8**[‡]
Global cognition										
MMSE	**22.6 ± 3.0**	**21.5 ± 3.2**[†]	23.5 ± 1.9	23.0 ± 2.3	23.4 ± 2.6	22.9 ± 3.0	**24.4 ± 3.1**	**22.7 ± 3.8**[*]	**23.5 ± 2.8**	**22.6 ± 3.2**[*]
APOE ε4 genotype										
APOE ε4 positive	225 (70%)	74 (59%)	72 (67%)	56 (63%)	175 (67%)	80 (58%)	**61 (66%)**	**39 (48%)**[‡]	**533 (68%)**	**249 (58%)**[*]
AD biomarkers										
AD biomarker available	275 (78%)	114 (79%)	52 (29%)	50 (26%)	123 (38%)	70 (38%)	25 (8%)	28 (11%)	475 (40%)	262 (33%)
AD biomarker positive	253 (92%)	105 (92%)	40 (77%)	40 (80%)	106 (86%)	58 (83%)	24 (96%)	28 (100%)	423 (89%)	231 (88%)
MRI										
Hippocampus	-0.07 ± 1.00	0.18 ± 0.97	-0.05 ± 1.01	0.05 ± 1.00	-0.09 ± 0.95	0.17 ± 1.07	-	-	**-0.07 ± 1.00**	**0.12 ± 1.00**[†]
Posterior cortex	**0.11 ± 0.94**	**-0.29 ± 1.10**[†]	-	-	**0.08 ± 0.98**	**-0.15 ± 1.03**[‡]	-	-	**0.09 ± 0.96**	**-0.21 ± 1.06**[*]
WMH	0.10 ± -0.01	-0.07 ± -0.01	0.01 ± -0.32	-0.01 ± 0.95	-0.04 ± 0.96	0.07 ± 1.07	-	-	**0.03 ± -0.01**	**0.00 ± 1.05**

Data are presented in number (%) or mean ± standard deviation, also when not-normally distributed enabling clearer comparison between clusters. p-values are based on t-tests, χ^2, or Kruskal Wallis analyses when appropriate. Normalized values are given for education, and MRI characteristics. AD biomarkers are available as cerebrospinal fluid total tau/amyloid β_{1-42} (abnormal when > 0.52 according to Duits e.a. [38]) in the ADC ADNI, and DCN cohorts, or as Pittsburgh compound B positron emission tomography positivity in the UCSF cohort. When MRI atrophy characteristics were measured using a visual rating scale in which a higher score reflects more severe atrophy, results were inverted so that higher scores reflect less atrophy. Differences between memory and non-memory clusters are shown in bold and indicated as follows: [*]$p \le 0.001$, [†]$p \le 0.01$, [‡]$p \le 0.05$. Interpretation: The non-memory clusters were younger (pooled sample, ADC), less educated (pooled sample), had shorter duration of complaints (pooled sample), lower MMSE scores (ADC, UCSF, pooled sample), were more often APOE e4 negative (UCSF, pooled sample), had less hippocampal atrophy (pooled sample), but more severe atrophy of the posterior cortex (ADC, DCN, pooled sample) than the memory clusters.

FIGURE LEGENDS

Figure 1. Nonnegative Matrix Factorization is a dual-clustering approach, meaning that clustering includes two parallel steps. Firstly, neuropsychological (NP) test results are grouped together into neuropsychological profiles ('components'), illustrated in the upper half of the figure, in which each row represents one neuropsychological test, and each column an identified neuropsychological component. The warmer the colour, the higher the test loads to the component when test scores are high (relatively spared cognition); the colder the colour, the lower the test loads to the component when test scores are high (relatively impaired cognition). The optimal number of components is based on the cophenetic correlation coefficient (for this example n=2). Secondly, patients are grouped together (into 'clusters') based on the fit of their neuropsychological profile to the identified neuropsychological component, taking each test's load to the component into account. This step is illustrated in the lower half of the figure, in which each row represents one patient. The warmer the colour, the better the fit of patients' neuropsychological profile to the neuropsychological profile of the identified component.

Figure 2. Memory tests are indicated in dark blue. Abbreviations: ABCD = Arizona Battery for Communication Disorders of Dementia, ADC = Amsterdam Dementia Cohort, ADNI = Alzheimer's Disease Neuroimaging Initiative, comp questions = comparative questions, CVLT = California Verbal Learning Test, DCN = German Dementia Competence Network, DS = Digit Span, FAB = Frontal Assessment Battery, LDST = Letter Digit Substitution Test, LM = CERAD Logical Memory, mem = memory-impaired, non-mem = memory-spared, NMF = Nonnegative Matrix Factorization, RAVLT = Rey Auditory Verbal Learning Test, TMT = Trail Making Test, UCSF = University of California – San Francisco, VAT = Visual Association Test, WL = CERAD Word List. In all cohorts, the optimal number of test clusters was two. In this figure, each row represents one neuropsychological test. The two columns represent the two found neuropsychological components. The warmer the colour, the higher the test loads to the component when test scores are high (relatively spared cognition); the colder the colour, the lower the test loads to the component when test scores are high (relatively impaired cognition). Interpretation: In each cohort, one component is associated with relative impairment of memory tests, therefore called the memory component. The other component is associated with relative impairment of non-memory functions, therefore called the non-memory component.

Figure 3. Abbreviations: ADC = Amsterdam Dementia Cohort, ADNI = Alzheimer's disease Neuroimaging Initiative, DCN = German Dementia Competence Network, UCSF = University of California – San Francisco. Patients were assigned to either the memory or non-memory cluster based on the fit of their neuropsychological profile to the memory or non-memory component (figure 2). Each column represents one patient. The warmer the colour, the better the fit of patients' neuropsychological profile to the neuropsychological profile of that component.

19

ACKNOWLEDGEMENTS

CONFLICTS OF INTEREST

N. Scheltens, B. Tijms, T. Koene, S. Wolfsgruber, M. Wagner, B. Cohn-Sheehy, J. Kramer, B. Miller and W. van der Flier report no conflicts of interest. F. Barkhof serves as a consultant to Biogen-Idec, Janssen Alzheimer Immunotherapy, Bayer-Schering, Merck-Serono, Roche, Novartis, Genzyme, and Sanofi-aventis. He has received sponsoring from EU-H2020, NWO, SMSR, EU-FP7, TEVA, Novartis, Toshiba, he is supported by NIHR-BRC-UCL, and he is a member of editorial boards of Radiology, Brain, Neuroradiology, MSJ, and Neurology. C. Teunissen serves on the advisory board of Fujirebio and Roche, received research consumables from Euroimmun, IBL, Fujirebio, Invitrogen and Mesoscale Discovery, and performed contract research for IBL, Shire,

20

Cognitive subtypes of probable Alzheimer's disease robustly identified in four cohorts

N.M.E. Scheltens[1], B.M. Tijms[1], T. Koene[2], F. Barkhof[3,4], C.E. Teunissen[5], S. Wolfsgruber[6], M. Wagner[6], J. Kornhuber[7], O. Peters[8], B.I. Cohn-Sheehy[9], G.D. Rabinovici[9], B.L. Miller[9], J.H. Kramer[9], Ph. Scheltens[1], W.M. van der Flier[1,10];
Alzheimer's Disease Neuroimaging Initiative*, German Dementia Competence Network, University of California San Francisco Memory and Aging Center, and Amsterdam Dementia Cohort.

[1] Department of Neurology, Alzheimer Center, Amsterdam Neuroscience, VU University Medical Center, Amsterdam, The Netherlands
[2] Department of Medical Psychology, VU University Medical Center, Amsterdam, The Netherlands
[3] Department of Radiology and Nuclear Medicine, Amsterdam Neuroscience, VU University Medical Center, Amsterdam, The Netherlands
[4] Institutes of Neurology and Healthcare Engineering, UCL, London, UK
[5] Neurochemistry Laboratory and Biobank, Department of Clinical Chemistry, Amsterdam Neuroscience, VU University Medical Centre, Amsterdam, The Netherlands
[6] Department of Psychiatry, University of Bonn, Bonn, Germany, and German Center for Neurodegenerative Diseases, Bonn, Germany
[7] Department of Psychiatry, Friedrich-Alexander-University Erlangen, Erlangen, Germany
[8] Department of Psychiatry, Charité Berlin, Campus Benjamin Franklin, Berlin, Germany
[9] Memory and Aging Center, Department of Neurology, University of California San Francisco, San Francisco, USA
[10] Department of Epidemiology and Biostatistics, VU University Medical Center, Amsterdam, The Netherlands

Title characters incl. spaces:	86	
Abstract word count:	150	150 allowed
Research in context	141	150 allowed
Text word count:	3651	3.500 allowed (excluding the abstract, references, figures, and tables)
Tables:	2	+ 4 supplementary
Figures:	3	+ 4 supplementary
References:	47	50 allowed

ABSTRACT

Introduction Patients with Alzheimer's disease (AD) show heterogeneity in profile of cognitive impairment. We aimed to identify cognitive subtypes in four large AD cohorts using a data-driven clustering approach.

Methods We included probable AD dementia patients from the Amsterdam Dementia Cohort (ADC, n=496), Alzheimer's Disease Neuroimaging Initiative (ADNI, n=376), German Dementia Competence Network (DCN, n=521), and University of California, San Francisco (UCSF, n=589). Neuropsychological data were clustered using Nonnegative Matrix Factorization. We explored clinical and neurobiological characteristics of identified clusters.

Results In each cohort, a two-clusters solution best fitted the data (cophenetic correlation >.9): One cluster was memory-impaired, and the other relatively memory-spared. Pooled analyses showed that the memory-spared clusters (29-52% of patients) were younger, more often APOE e4 negative, and had more severe posterior atrophy compared to the memory-impaired clusters (all $p<.05$).

Conclusions We could identify two robust cognitive clusters in four, independent large cohorts with distinct clinical characteristics.

ABBREVIATIONS ADC = Amsterdam Dementia Cohort, ADNI = Alzheimer's Disease Neuroimaging Initiative, APOE = Apolipoprotein E, CSF = cerebrospinal fluid, DCN = German Dementia Competence Network, MMSE = Mini-Mental State Examination, MRI = magnetic resonance imaging, MTA = medial temporal lobe atrophy, NMF = nonnegative matrix factorization, PA = atrophy of the posterior cortex, UCSF = University of California, San Francisco.

HIGHLIGHTS (max 85 characters including spaces per point)

- Alzheimer's disease (AD) is a heterogeneous disorder (54)
- We identified two cognitive AD subtypes in four cohorts with a data-driven approach (82)
- Non-amnestic AD is associated with distinct neurobiological characteristics (77)

RESEARCH IN CONTEXT

1.Systematic review: Alzheimer's disease (AD) is characterized by cognitive heterogeneity. We searched PubMed for clinical and neurobiological heterogeneity in AD profiles, and for data-driven approaches used to identify AD subtypes based on neuropsychological test scores. Several studies demonstrated the potential of clustering methods to identify cognitive AD subtypes. Identified subtypes showed distinct clinical characteristics. However none of the previous studies tested the generalizability of the cluster solutions, since those results were based on single-cohort studies.

2. Interpretation: In four large AD cohorts, we consistently identified two cognitive clusters using Nonnegative Matrix Factorization of neuropsychological test scores. In each cohort one cluster most

prominently showed memory-impairment, and the other cluster was relatively memory-spared. These clusters were associated with distinct clinical characteristics.

3. Future directions: Future research should aim to further study the underlying biological disease mechanisms that cause a non-memory AD phenotype.

1. Introduction

Alzheimer's disease (AD) dementia is characterized by progressive cognitive impairment in multiple cognitive domains, e.g. memory, language, visuospatial and executive functioning, and attention. Typically, AD is characterized by early and prominent memory loss.[1] A minority of patients has a prominent and relatively focal cognitive presentation, such as logopenic-variant primary progressive aphasia (lvPPA), posterior cortical atrophy (PCA), or a behavioural/dysexecutive subtype.[2-5] Atypical variants have been associated with specific demographic, genetic, and neuroimaging/biomarker findings that are distinct from those of typical amnestic patients (e.g. age at onset, apolipoprotein E [APOE] genotype, distribution of cortical atrophy, hypometabolism, tau deposition, cerebrospinal fluid biomarker concentrations, and pathological findings).[6-10] However, even patients who do not display a defined subtype also show considerable variation in patterns of cognitive impairment. Earlier studies demonstrated the potential to capture cognitive heterogeneity in AD using a data-driven clustering approach.[11-14] Studies differed in sample size, clinical diagnosis of included patients, available neuropsychological test results, available neurobiological characteristics to compare clusters with, and clustering technique. This has resulted in different numbers of clusters, with different cognitive and neurobiological characteristics. Although those studies were clearly suggestive of variability in underlying pathological mechanisms, it is difficult to generalise the findings, since they result from single studies that show considerable variability in patient population and methodological approaches.

In the present study, we aimed to identify cognitive subtypes and to study whether these subtypes could be replicated in three independent AD dementia cohorts. For the identification of cognitive AD subtypes, we used Nonnegative Matrix Factorization (NMF).[15-18] Based on earlier descriptions of cognitive heterogeneity, we expected NMF to identify at least a cluster including patients with typical amnestic AD, and one or more other clusters including patients with non-amnestic features.[15-18]

2. Methods

2.1. Patients

We selected AD patients from four large cohorts: The Amsterdam Dementia Cohort (ADC), the Alzheimer's Disease Neuroimaging Initiative (ADNI), the German Dementia Competence Network (DCN), and the University of California, San Francisco Memory and Aging Center research cohort (UCSF). Patients were selected based on 1) clinical diagnosis of probable AD dementia, 2) availability of neuropsychological test results, and 3) Mini-Mental State Examination (MMSE) score >16/30.[19] In the ADC and UCSF cohort, patients with focal presentations lvPPA, PCA, and the behavioural/dysexecutive subtype of probable AD dementia were included, while such subjects were explicitly excluded from participation in the ADNI and DCN studies.

From the ADC we selected 496 patients with probable AD.[20] Patients visited the outpatient memory clinic of the VU University Alzheimer Center between 2008 and 2013. Standard dementia screening included for most patients medical history and medication use, physical and neurological examination, extensive neuropsychological evaluation, screening laboratory tests, APOE genotyping, magnetic resonance imaging (MRI), and lumbar puncture (LP). In the ADC, level of education was defined according to a rating scale ranging from 1 (low, primary school not finished) to 7 (high, university degree).[21] All participants provided written informed consent to use their clinical data for research purposes. The local ethical committee approved the study.

From the ADNI database (adni.loni.usc.edu) we selected 376 probable AD patients. Patients were recruited in over 50 sites across the U.S. and Canada (www.adni-info.org). Standard workup included medical history, physical and neurological examination, extensive neuropsychological evaluation, screening laboratory tests, APOE genotyping, neuroimaging including MRI, and LP. For the present study, we used data of screening and baseline visits, acquired for ADNI-1 or ADNI-2 between 2005 and 2013. All patients gave written informed consent at screening.

From the DCN cohort database (http://www.kompetenznetz-demenzen.de) we selected 521 probable AD patients.[22] The DCN is a collaboration of fourteen specialized German memory clinics from university hospitals. All patients were offered a uniform dementia screening at first visit between 2003 and 2007, including medical history, physical and neurological examination, extensive neuropsychological evaluation, screening laboratory tests, MRI scan, and LP. The DCN study protocol was approved by the Institutional Review Boards of all participating study centres.[22] All patients, or their legal guardians, provided written informed consent.

From the UCSF research cohort we selected 589 probable AD patients.[23] Patients were either seen in the outpatient memory clinic, or for a research assessment in the UCSF Alzheimer's Disease Research Center. All patients were assessed at first visit between 1998 and 2013. Standardized dementia screening included medical history, physical and neurological examination, neuropsychological evaluation, screening laboratory tests, APOE genotyping, and neuroimaging including MRI. A core screening neuropsychological battery was performed in both the clinical and research settings. All patient and informants provided written informed consent. Surrogate consent was accepted when patients lacked capacity to provide consent themselves. The local medical ethical committee approved the study.

2.2 Neuropsychological tests

Neuropsychological data included tests covering the major cognitive domains in each cohort, but the exact composition of neuropsychological test batteries differed across cohorts. Neuropsychological tests included for analysis in this study are shown in supplementary table 1. The number of missing neuropsychological values differed across cohorts and within neuropsychological test batteries (on average 20% in ADC, 27% in ADNI, 1% in DCN, 12% in UCSF). Main reasons for missingness are practical reasons unrelated to the data (random). In part of the cases however, tests could not be finished due to cognitive impairment, whereas scoring differed across cohorts and between tests (i.e. assignment of missing value or minimum score). The clustering technique NMF does not allow for missing data or negative values. In order to reduce selection bias, we did not select patients based on completeness of datasets, but we completed the datasets using a multiple imputation approach that is commonly used as a reliable method to estimate missing data.

We imputed missing neuropsychological data using R package Multivariate Imputation by Chained Equations (MICE, version 2.25).[24,25] MICE estimates missing neuropsychological values by predicting these values from the relationships with other neuropsychological variables. We also included predictors age, gender, MMSE, and when available education, duration of complaints, Alzheimer's Disease Assessment Scale – Cognitive subscale (ADAS-cog) or Cambridge Cognitive Examination (CAMCOG) or Cognitive Dementia Rating (CDR) sum of boxes in the imputation model. We ran MICE 50 times per cohort, resulting in 50 imputed datasets for each cohort. For further analyses, we included pooled measures over the 50 derived imputed datasets per cohort.[26,27] We inverted values when appropriate, so that for all tests lower scores reflect worse cognitive impairment. Next, the imputed neuropsychological data were normalized and scaled to include only positive values (0-1).

2.3 MRI characteristics

MRI characteristics were available for patients from the ADC, ADNI, and DCN cohorts.

For the ADC cohort, imaging data was obtained on a 1.5T or 3T scan. Visual ratings of medial temporal lobe atrophy (MTA; range 0-4[28]), posterior atrophy (PA; range 0-3[29]), and white matter hyperintensities (WMH; range 0-3[30]) were performed by an experienced neuroradiologist. For ADNI, a structural MRI 1.5T scan was performed on screening or baseline visit.[31] Image processing has been done with cortical reconstruction and volumetric segmentations using FreeSurfer 4.3 (surfer.nmr.mgh.harvard.edu/).[32] Hippocampal and WMH volumes were downloaded from the ADNI website (ida.loni.usc.edu/).[33] For the DCN cohort, MRI data was obtained from multiple 1.5T scanners with standardized MRI acquisition guidelines across centres.[22,34] PA and WMH were scored by experienced neuroradiologists using a visual rating scale in which higher scores reflect more severe abnormalities (range 0-3). Furthermore, hippocampal volumes were measured using FMRIB's Integrated Registration and Segmentation Tool (FIRST) from the FMRIB Software Library (FSL) package of tools.[35,36] For all atrophy measurements, we included the mean of left and right hemisphere.

2.4 APOE e4 genotype

APOE e4 genotype was available in the ADC (n=448; 90%), ADNI (n=196; 52%), DCN (n=397; 76%), and UCSF (n=175; 30%) cohorts. We dichotomised APOE e4 genotype according to the presence or absence of one or more APOE e4 alleles. APOE genotype was assessed using the Light Cycler APOE mutation detection method (Roche Diagnostics GmbH, Mannheim, Germany) in the ADC. For ADNI, APOE alleles were genotyped using DNA extracted by Cogenics. For the DCN, APOE genotype was assessed using Qiagen blood isolation kit (Qiagen, Hilden, Germany). For UCSF, APOE genotype was conducted using a TaqMan Allelic Discrimination Assay on an ABI 7900HT Fast Real-Time PCR system (Applied Biosystems, Foster City, CA, USA).

2.5 AD biomarkers

Cerebrospinal fluid (CSF) markers amyloid-beta$_{1-42}$ (Abeta$_{1-42}$), and total tau (tau) were available for ADC (n=389; 79%), ADNI (n=102; 27%), and DCN (n=193; 37%). CSF biomarkers were assessed using Sandwich ELISAs (Fujirebio, Gent, Belgium) in ADC and DCN[37], and Multiplex xMAP Luminex platform (Luminex Corp, Austin, TX) with Fujirebio immunoassay kit–based reagents (INNO-BIA Alzbio3; Fujirebio, Ghent, Belgium) in ADNI. CSF biomarkers were considered positive for AD when tau / Abeta$_{1-42}$ ratio was > 0.52 [38]. Pittsburgh compound B position emission tomography visual reading results (i.e. positive or negative) were available for the UCSF cohort (n=52; 9%).

2.6 Statistical analysis

For statistical analyses, we used RStudio for Mac version 3.2.2 (Integrated Development for R. RStudio, Inc., Boston, MA, http://www.rstudio.com). Clustering was performed with the R package

Nonnegative Matrix Factorization (NMF, version 0.20.6).[39] Clustering is a data-driven method to divide a heterogeneous set of objects (patients) in subgroups that are more homogeneous in terms of characteristics provided as input for the clustering (neuropsychological test results). NMF is a dual-clustering approach, meaning that clustering includes two parallel steps (illustrated in figure 1). Firstly, neuropsychological test results are grouped together into neuropsychological profiles ('components'), and neuropsychological tests that determine these components can be identified by their component loading. Secondly, patients are grouped together based on the fit of their neuropsychological profile to the neuropsychological summary component. The optimal number of clusters is based on the most consistent assignment of patients to the identified cognitive profiles in the multiple runs of NMF. The stability of the cluster solution is assessed with the cophenetic correlation coefficient, that measures how consistently tests and subjects are assigned to a given component, ranging between 0 and 1 (i.e. no stability to stable cluster solution). For each cohort, we determined the optimal number of clusters based on the highest cophenetic correlation coefficient for 2 to 9 cluster solutions.[15] We used the 'nonsmooth' NMF algorithm that introduces an intermediate smoothing matrix to enhance sparsity of the clusters.[40]

Characterization of identified clusters in terms of neuropsychological profile was based on the most strongly loading neuropsychological tests.[17] For characterization of identified clusters in terms of demographic, clinical, and neurobiological characteristics, we analysed age, sex, education, disease duration reported by the patient, MMSE, APOE e4 genotype, CSF biomarkers, MRI atrophy, and WMH measurements using χ^2, t-tests, or Kruskal Wallis tests where appropriate. These analyses were performed for each cohort separately. In addition, for each cluster we pooled the patient characteristics over the cohorts in order to compare them for the total sample. To this end we Z-transformed variables with different scales (i.e. education, CSF biomarkers, and MRI biomarkers) before pooling. When atrophy of the hippocampus and the posterior cortex was measured using a visual rating scale in which higher scores reflect more severe atrophy, the normalized scores were inverted so that higher scores reflect less atrophy in all cohorts and the pooled sample.

3. RESULTS

3.1 Cohort characteristics

Characteristics of all cohorts are summarized in table 1. On average, patients were 71±9 years old, with the ADC being the youngest cohort. 54% of patients were female, ranging from 44% (ADNI) to 60% (DCN). Patients were mildly-to moderately demented, with an average MMSE varying between cohorts from 22-24. Roughly two thirds of patients were APOE e4 carrier, ranging from 57% (UCSF) to 67% (ADC).

3.2 NMF clusters of cognitive subtypes

NMF is a dual-clustering approach; first, neuropsychological tests are grouped into components, and second patients are clustered based on the fit of their neuropsychological profiles to the profiles of the identified neuropsychological components, taking the load of each test to the component into account. NMF showed that within each cohort, the optimal number of clusters was two, as the solution with two clusters showed the strongest cophenetic correlation (>.90).

Results of the clustering of tests are shown in figure 2. In the ADC, one neuropsychological component mainly included memory tests Rey auditory verbal learning test (RAVLT) immediate and delayed recall. The other component included mainly non-memory tests trail making test (TMT)-A and TMT-B, and fragmented letters. In the ADNI cohort, one neuropsychological component mainly included memory tests logical memory immediate and delayed recall, and RAVLT delayed recall. The other component mainly included non-memory tests TMT-A and TMT-B. In the DCN cohort, one neuropsychological component included mainly memory tests logical memory immediate and delayed recall, word list immediate and delayed recall of the Consortium to Establish a Registry for Alzheimer's Disease (CERAD), and Rey figure recall. The other component included mainly non-memory tests TMT-A, TMT-B, clock drawing, and CERAD figure copy. In the UCSF cohort, one neuropsychological component included mainly memory tests figure recall, and the California verbal learning test (CVLT) delayed recall. The other component included mainly non-memory tests modified trails, design fluency, and stroop interference.

Patients were assigned to either of two clusters based on the fit of their neuropsychological test results to the memory or non-memory component in each cohort (figure 3). Across all cohorts, the memory clusters included on average 60% of patients, ranging from 48% (ADNI) to 71% (ADC), and the non-memory clusters included on average 40% of patients, ranging from 29% (ADC) to 52% (ADNI).

3.3 Cluster characterisation

We analysed cluster characteristics in terms of demographical, clinical and neurobiological characteristics. Associations for each cohort are shown in table 2. In the ADC, patients in the non-memory cluster were younger, had lower MMSE scores, and had more severe atrophy of the posterior cortex. In the ADNI cohort, non-memory patients tended to be younger and less often APOE e4 positive with relative hippocampal sparing, but these differences did not reach significance. In the DCN cohort, patients in the non-memory cluster had more severe atrophy of the posterior cortex. In the UCSF cohort, the non-memory cluster had lower MMSE scores, and patients were less often APOE e4 positive. There were no differences in WMH (ADC, ADNI, DCN). When we analysed pooled data, we found that across cohorts, patients in non-memory clusters were younger, less educated, reported shorter disease duration, had lower MMSE scores, were less often APOE e4 positive, had less severe hippocampal atrophy, and more severe atrophy of the posterior cortex than patients in memory clusters.

3.4 Validation of cognitive clusters after stratification for disease severity

We tested whether disease severity influenced the cluster solutions by repeating NMF analyses in each cohort after stratifying for disease severity according to cohort median MMSE value. Results appeared to be robust across severity subgroups, with a memory cluster and non-memory cluster appearing in both the mild and the moderately demented patient strata (supplementary figures 1-4). Cluster characteristics in terms of demographic and neurobiological characteristics remained largely the same (supplementary tables 2 and 3). Differences between the memory and non-memory clusters were most pronounced in the mildly demented stratum.

3.5 Validation of cognitive clusters in AD biomarker confirmed patients

Availability of AD biomarkers is given in table 1 and 2. Based on χ^2 analyses, no differences were found between the memory and non-memory clusters in terms of AD biomarker positivity ($p > 0.05$). Only for the ADC, enough data were available to repeat data-driven NMF analyses in AD biomarker confirmed patients (n=357 with CSF total tau/amyloid ß$_{1-42}$ > 0.52[38]). Results appeared to be consistent with findings of the total cohorts, with a memory component mainly including RAVLT immediate and delayed recall, and a non-memory component mainly including TMT-A, TMT-B, fragmented letters, and letter digit substitution test (LDST), shown in supplementary figure 5. The memory cluster included 73% of patients, and the non-memory cluster 27% (supplementary figure 6). Consistent with characteristics of the total ADC, the non-memory cluster of the biomarker confirmed subset was younger (64 ± 8 versus 67 ± 8 years old, $p < 0.01$), had lower MMSE scores (20.9 ± 3.1 versus 22.7 ± 3.0, $p < 0.001$), and had more severe posterior atrophy (1.51 ± 0.82 versus 1.13 ± 0.71, $p < 0.05$, rated according to a visual rating scale in which higher scores reflect more severe atrophy [29]) than the memory cluster. In addition, the non-memory cluster was less often APOE e4 positive than the memory cluster (61% versus 74%, $p < 0.05$). Results are provided in supplementary table 4.

4. DISCUSSION

Across four independent AD dementia cohorts, we robustly found two cognitive clusters using a data-driven dual-clustering approach. One cluster was characterized by more prominent memory impairment, and one cluster by more prominent impairment on non-memory tests. These memory and non-memory AD phenotypes were consistently found across cohorts, even though these cohorts differed in their patient populations (e.g. age, disease severity, mono-centre versus multi-centre, geographic location) and composition and extensiveness of neuropsychological test battery. Moreover, the clusters were associated with specific demographic, clinical, and neurobiological characteristics. These findings demonstrate the biological relevance of clinical heterogeneity in AD, as this may reflect variation in underlying disease mechanisms.

Of all included patients, 40% belonged to a non-memory cluster. Compared to the memory clusters, non-memory cluster patients were younger, less educated, more often APOE e4 negative, had more severe posterior atrophy and relatively spared hippocampi. In addition, patients assigned to the non-memory clusters reported shorter disease duration, but they had on average lower MMSE scores. Possibly, non-memory clusters are associated with a more aggressive disease progression.[41,42] Future studies should address the question whether cognitive subtypes are related to rate of decline, and hence suitable as putative prognostic marker.

Among the suggested disease mechanisms causing heterogeneity, is the influence of copathologies, e.g. vascular pathology. We did not find differences in severity of ischemic vascular pathology (WMH) however. Of note, cognitive heterogeneity is most prominent in early onset AD patients, where AD pathology is often pure, and copathologies are less present.[43] Biomarker support for AD pathology was not available for each patient, but the available data showed no difference in AD biomarker positivity between clusters, suggesting that misdiagnosis is not a major driver of our findings. Variation in disease mechanisms could also be sought in the origin and spreading of neurofibrillary tangles (typically characterized by origin in the entorhinal cortex, progressing through the hippocampus to the association cortex, and finally to the cortex [44]) since the medial temporal lobe was relatively spared and the posterior cortex most prominently affected in the non-memory phenotype. This idea fits with the hypothesis that early-onset, APOE e4 negative AD patients are predisposed to vulnerability of cerebral networks beyond the medial temporal lobe.[44,45] This hypothesis coincides with an autopsy study that identified an AD subtype with relative less tangles in the hippocampus that was associated with younger age at death, male sex, rapid disease progression, and more often focal cortical clinical syndromes.[46] However, this phenotype only tended to have less often an APOE e4 positive genotype (p=0.067). Probably, the APOE e4 allele is not the only AD risk factor that modifies clinical heterogeneity; the effect of genetic risk factors on clinical AD

phenotype is still poorly understood. An increasing number of promising genes however are discovered to be associated with higher or reduced risks for developing AD and with certain neuropathological pathways.[47]

A limitation of our study is that we did not have post mortem pathological confirmation of patients, so we cannot rule out the possibility of misdiagnosis. We think however that misdiagnoses will not have driven our results, as in all cohorts diagnoses were made according to careful application of clinical criteria, and repeating analyses in the CSF biomarker confirmed subset of ADC gave similar results. Furthermore, the present study could be biased since we excluded patients that were already severely demented (i.e. MMSE < 17) at diagnosis. We think that this selection bias could have resulted in underrepresentation of the non-memory phenotype, because atypical AD variants are less easily recognized as AD (especially at younger age, when other causes for cognitive complaints such as depression or burn-out are more common) and therefore probably more often associated with patients' and doctors' delay, and delay because of initial misdiagnoses. In addition, non-memory patients reporter shorter disease duration, while MMSE scores were already lower, possibly due to faster disease progression before diagnosis, suggesting higher risk for more severe dementia at time of diagnosis.

It could be argued that the substantial differences between cohorts in patient population (e.g. geographically, age, disease severity, degree of cognitive heterogeneity within cohorts), and in extensiveness of neuropsychological test battery could be a limitation. However, we see this as a strong point of our study since we were able to replicate our finding of two robust clusters with their corresponding clinical characterisation. This suggests that the clusters we identified are generalizable to other AD populations; an often-encountered limitation of data driven methods to cluster patients is that of limited generalizability.

Non-memory cluster patients had on average lower MMSE scores, and therefore we performed additional analyses to study whether clustering has been driven by disease severity. Repeating the clustering after stratification based on MMSE scores, a memory and a non-memory cluster were identified in each stratified cohort as well. Cluster differences in terms of clinical characteristics were largely similar for the strata, albeit more pronounced in the mildly demented stratum. This suggests that clinical heterogeneity is more prominently present in early stages of AD.

Our results emphasize the presence of non-memory phenotypes in AD. Being able to identify AD subtypes is important in a clinical setting for early and adequate diagnosis and personalized medicine. Also, cognitive profiling should be taken into account when including patients for clinical trials, or when choosing cognitive outcomes to analyse the effect of an intervention. Furthermore, the existence

of these clusters, with similar patient characteristics across independent cohorts suggests that cognitive heterogeneity is caused by different disease mechanisms.

In conclusion, we found two robust AD subtypes using a data-driven clustering approach in four AD cohorts. Identified clusters were associated with distinct demographical, clinical, and neurobiological characteristics, emphasizing the presence of cognitive heterogeneity in AD, and suggesting variation in underlying pathology.

REFERENCES

[1]   McKhann GM, Knopman DS, Chertkow H, Hyman BT, Jack CR, Kawas CH, et al. The diagnosis of dementia due to Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. Alzheimers Dement, vol. 7, 2011, pp. 263–9. doi:10.1016/j.jalz.2011.03.005.

[2]   Gorno-Tempini ML, Hillis AE, Weintraub S, Kertesz A, Mendez M, Cappa SF, et al. Classification of primary progressive aphasia and its variants. Neurology, vol. 76, Lippincott Williams & Wilkins; 2011, pp. 1006–14. doi:10.1212/WNL.0b013e31821103e6.

[3]   Benson DF, Davis RJ, Snyder BD. Posterior cortical atrophy. Arch Neurol 1988;45:789–93.

[4]   Crutch SJ, Schott JM, Rabinovici GD, Boeve BF, Cappa SF, Dickerson BC, et al. Shining a light on posterior cortical atrophy. Alzheimers Dement, vol. 9, 2013, pp. 463–5. doi:10.1016/j.jalz.2012.11.004.

[5]   Ossenkoppele R, Pijnenburg YAL, Perry DC, Cohn-Sheehy BI, Scheltens NME, Vogel JW, et al. The behavioural/dysexecutive variant of Alzheimer's disease: clinical, neuroimaging and pathological features. Brain 2015;138:2732–49.

[6]   Van Der Flier WM, Schoonenboom SNM, Pijnenburg YAL, Fox NC, Scheltens P. The effect of APOE genotype on clinical phenotype in Alzheimer disease. Neurology 2006;67:526–7. doi:10.1212/01.wnl.0000228222.17111.2a.

[7]   van der Vlies AE, Pijnenburg YAL, Koene T, Klein M, Kok A, Scheltens P, et al. Cognitive impairment in Alzheimer's disease is modified by APOE genotype. Dement Geriatr Cogn Disord 2007;24:98–103. doi:10.1159/000104467.

[8]   Smits LL, Tijms BM, Benedictus MR, Koedam ELGE, Koene T, Reuling IEW, et al. Regional atrophy is associated with impairment in distinct cognitive domains in Alzheimer's disease. Alzheimers Dement 2014;10:S299–305. doi:10.1016/j.jalz.2013.06.007.

[9]   Ossenkoppele R, Cohn-Sheehy BI, La Joie R, Vogel JW, Möller C, Lehmann M, et al. Atrophy patterns in early clinical stages across distinct phenotypes of Alzheimer's disease. Hum Brain Mapp 2015;36:4421–37. doi:10.1002/hbm.22927.

[10]   Ossenkoppele R, Mattsson N, Teunissen CE, Barkhof F, Pijnenburg Y, Scheltens P, et al. Cerebrospinal fluid biomarkers and cerebral atrophy in distinct clinical variants of probable Alzheimer's disease. Neurobiol Aging 2015;36:2340–7. doi:10.1016/j.neurobiolaging.2015.04.011.

[11]   Stopford CL, Snowden JS, Thompson JC, Neary D. Variability in cognitive presentation of Alzheimer's disease. Cortex 2008;44:185–95. doi:10.1016/j.cortex.2005.11.002.

[12]   Davidson JE, Irizarry MC, Bray BC, Wetten S, Galwey N, Gibson R, et al. An exploration of cognitive subgroups in Alzheimer's disease. J Int Neuropsychol Soc 2010;16:233–43. doi:10.1017/S1355617709991160.

[13]   Libon DJ, Drabick DAG, Giovannetti T, Price CC, Bondi MW, Eppig J, et al. Neuropsychological syndromes associated with Alzheimer's/vascular dementia: a latent class analysis. J Alzheimers Dis 2014;42:999–1014. doi:10.3233/JAD-132147.

[14]   Scheltens NME, Galindo-Garre F, Pijnenburg YAL, van der Vlies AE, Smits LL, Koene T, et al. The identification of cognitive subtypes in Alzheimer's disease dementia using latent class analysis. J Neurol Neurosurg Psychiatr 2015:jnnp–2014–309582. doi:10.1136/jnnp-2014-309582.

[15]   Brunet J-P, Tamayo P, Golub TR, Mesirov JP. Metagenes and molecular pattern discovery using matrix factorization. Proc Natl Acad Sci USa 2004;101:4164–9. doi:10.1073/pnas.0308531101.

[16]   Wang Y-X, Zhang Y-J. Nonnegative Matrix Factorization: A Comprehensive Review. Knowledge and Data Engineering, IEEE Transactions on 2013;25:1336–53. doi:10.1109/TKDE.2012.51.

[17]   Carmona-Saez P, Pascual-Marqui RD, Tirado F, Carazo JM, Pascual-Montano A. Biclustering of gene expression data by Non-smooth Non-negative Matrix Factorization. BMC Bioinformatics 2006;7:78. doi:10.1186/1471-2105-7-78.

[18]   Shokrollahi M, Krishnan S, Dopsa DD, Muir RT, Black SE, Swartz RH, et al. Nonnegative matrix factorization and sparse representation for the automated detection of periodic limb movements in sleep. Med Biol Eng Comput 2016:1–14. doi:10.1007/s11517-015-1444-y.

[19]   Folstein MF, Folstein SE, McHugh PR. "Mini-mental state." A practical method for grading the cognitive state of patients for the clinician. J Psychiatr Res 1975;12:189–98.

[20]   van der Flier WM, Pijnenburg YAL, Prins N, Lemstra AW, Bouwman FH, Teunissen CE, et al. Optimizing patient care and research: the Amsterdam Dementia Cohort. J Alzheimers Dis 2014;41:313–27. doi:10.3233/JAD-132306.

[21]   Verhage F. Intelligence and religious persuasion. Ned Tijdschr Psychol 1964;19:247–54.

[22]   Kornhuber J, Schmidtke K, Frolich L, Perneczky R, Wolf S, Hampel H, et al. Early and differential diagnosis of dementia and mild cognitive impairment: design and cohort baseline characteristics of the German Dementia Competence Network. Dement Geriatr Cogn Disord 2009;27:404–17. doi:10.1159/000210388.

[23]   Kramer JH, Jurik J, Sha SJ, Rankin KP, Rosen HJ, Johnson JK, et al. Distinctive neuropsychological patterns in frontotemporal dementia, semantic dementia, and Alzheimer disease. Cogn Behav Neurol 2003;16:211–8.

[24]   Van Buuren S, Groothuis-Oudshoorn K. MICE: Multivariate Imputation by Chained Equations in R. Journal of Statistical Software 2011;45:1–67.

[25]   Van Buuren S. Flexible Imputation of Missing Data. vol. 20125245. CRC Press; 2012. doi:10.1201/b11826.

[26]   Rubin DB. Multiple Imputation for Nonresponse in Surveys. Hoboken, NJ, USA: John Wiley & Sons; 2009. doi:10.1002/9780470316696.

[27]   Marshall A, Altman DG, Holder RL, Royston P. Combining estimates of interest in prognostic modelling studies after multiple imputation: current practice and guidelines. BMC Med Res Methodol 2009;9:57. doi:10.1186/1471-2288-9-57.

[28]   Scheltens P, Launer LJ, Barkhof F, Weinstein HC, van Gool WA. Visual assessment of medial temporal lobe atrophy on magnetic resonance imaging: interobserver reliability. J Neurol 1995;242:557–60.

[29]   Koedam ELGE, Lehmann M, van der Flier WM, Scheltens P, Pijnenburg YAL, Fox N, et al. Visual assessment of posterior atrophy development of a MRI rating scale. Eur Radiol 2011;21:2618–25. doi:10.1007/s00330-011-2205-4.

[30]   Fazekas F, Chawluk JB, Alavi A, Hurtig HI, Zimmerman RA. MR signal abnormalities at 1.5 T in Alzheimer's dementia and normal aging. AJR Am J Roentgenol 1987;149:351–6. doi:10.2214/ajr.149.2.351.

[31]   Jack CR, Bernstein MA, Fox NC, Thompson P, Alexander G, Harvey D, et al. The Alzheimer's Disease Neuroimaging Initiative (ADNI): MRI methods. J Magn Reson Imaging 2008;27:685–91. doi:10.1002/jmri.21049.

[32]   Fischl B. FreeSurfer. Neuroimage 2012;62:774–81. doi:10.1016/j.neuroimage.2012.01.021.

[33]   Schwarz CG, Fletcher E, DeCarli CS, Carmichael O. Fully-Automated White Matter Hyperintensity Detection With Anatomical Prior Knowledge and Without FLAIRs. 2013.

[34]   Ewers M, Teipel SJ, Dietrich O, Schönberg SO, Jessen F, Heun R, et al. Multicenter assessment of reliability of cranial MRI. Neurobiol Aging 2006;27:1051–9. doi:10.1016/j.neurobiolaging.2005.05.032.

[35]   Patenaude B, Smith SM, Kennedy DN, Jenkinson M. A Bayesian model of shape and appearance for subcortical brain segmentation. Neuroimage 2011;56:907–22. doi:10.1016/j.neuroimage.2011.02.046.

[36]   Smith SM, Jenkinson M, Woolrich MW, Beckmann CF, Behrens TEJ, Johansen-Berg H, et al. Advances in functional and structural MR image analysis and implementation as FSL. Neuroimage 2004;23 Suppl 1:S208–19. doi:10.1016/j.neuroimage.2004.07.051.

[37]   Lewczuk P, Kornhuber J, Wiltfang J. The German Competence Net Dementias: standard operating procedures for the neurochemical dementia diagnostics. J Neural Transm (Vienna) 2006;113:1075–80. doi:10.1007/s00702-006-0511-9.

[38]   Duits FH, Teunissen CE, Bouwman FH, Visser PJ, Mattsson N, Zetterberg H, et al. The cerebrospinal fluid "Alzheimer profile": easily said, but what does it mean? Alzheimers Dement 2014;10:713–723.e2. doi:10.1016/j.jalz.2013.12.023.

[39]   Gaujoux R, Seoighe C. A flexible R package for nonnegative matrix factorization. BMC Bioinformatics 2010;11:367. doi:10.1186/1471-2105-11-367.

[40]   Pascual-Montano A, Carazo JM, Kochi K, Lehmann D, Pascual-Marqui RD. Nonsmooth nonnegative matrix factorization (nsNMF). IEEE Trans Pattern Anal Mach Intell 2006;28:403–15. doi:10.1109/TPAMI.2006.60.

[41]   van der Vlies AE, Koedam ELGE, Pijnenburg YAL, Twisk JWR, Scheltens P, Van Der Flier WM. Most rapid cognitive decline in APOE epsilon4 negative Alzheimer's disease with early onset. Psychol Med 2009;39:1907–11. doi:10.1017/S0033291709005492.

[42]   Smits LL, Pijnenburg YAL, van der Vlies AE, Koedam ELGE, Bouwman FH, Reuling IEW, et al. Early onset APOE E4-negative Alzheimer's disease patients show faster cognitive decline on non-memory domains. Eur Neuropsychopharmacol 2015;25:1010–7. doi:10.1016/j.euroneuro.2015.03.014.

[43]   van der Flier WM. Clinical heterogeneity in familial Alzheimer's disease. Lancet Neurol 2016;15:1296–8. doi:10.1016/S1474-4422(16)30275-7.

[44]   Braak H, Braak E. Neuropathological stageing of Alzheimer-related changes. Acta Neuropathol 1991;82:239–59.

[45]   van der Flier WM, Pijnenburg YA, Fox NC, Scheltens P. Early-onset versus late-onset Alzheimer's disease: the case of the missing APOE ε4 allele. Lancet Neurol 2011;10:280–8. doi:10.1016/S1474-4422(10)70306-9.

[46]   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. doi:10.1016/S1474-4422(11)70156-9.

[47]   Lambert JC, Ibrahim-Verbaas CA, Harold D, Naj AC, Sims R, Bellenguez C, et al. Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. Nat Genet 2013;45:1452–8. doi:10.1038/ng.2802.

Table 1 Demographical and neurobiological characteristics of study cohorts

	ADC	ADNI	DCN	UCSF	Pooled sample
	n = 496	n = 376	n = 521	n = 589	n = 1.982
Demographics					
Age (yr)	67 ± 8	75 ± 8	72 ± 8	71 ± 10	71 ± 9
Female	258 (52%)	165 (44%)	313 (60%)	328 (56%)	1064 (54%)
Education*	5 (4-6)	16 (13-18)	11 (10-13)	16 (14-18)	0.00 ± 1.00
Duration complaints (yr)	2 (2-4)	-	2 (1-3)	-	2 (1-4)
Global cognition					
MMSE	22 ± 3	23 ± 2	23 ± 3	24 ± 4	23 ± 3
APOE e4 genotype					
APOE e4 positive	299 (67%)	128 (65%)	255 (64%)	100 (57%)	782 (64%)
AD biomarkers[†]					
AD biomarker available	389 (79%)	102 (27%)	193 (37%)	53 (9%)	737 (37%)
AD biomarker positive	358 (92%)	80 (78%)	164 (85%)	52 (98%)	654 (89%)
MRI					
Hippocampus[‡]	1.5 (1-2)	2882 ± 511	2933 ± 473	-	0.00 ± 1.00
Posterior cortex[§]	1 (1-2)	-	1 (0-2)	-	0.00 ± 1.00
WMH	1 (0-2)	0.41 (0.16-1.25)	0.5 (0-1)	-	0.02 ± 1.01

Data are presented in mean ± standard deviation, number (%), or median (2nd-4th quantile). Abbreviations: ADC = Amsterdam Dementia Cohort, ADNI = Alzheimer's Disease Neuroimaging Initiative, APOE = Apolipoprotein E, c.o. = cut-off, CSF = cerebrospinal fluid, DCN = German Dementia Competence Network, MMSE = Mini-Mental State Examination, UCSF = University of California – San Francisco, WMH = white matter hyperintensities. *Education is given according to the Verhage scale (1-7 resp. low-high education[21]) for ADC, years of education for ADNI, DCN, and UCSF, and normalized scores for the pooled sample. [†]AD biomarkers are available as cerebrospinal fluid total tau/amyloid β_{1-42} (abnormal when > 0.52 according to Duits e.a. [38]) in the ADC ADNI, and DCN cohorts, or as Pittsburgh compound B positron emission tomography positivity in the UCSF cohort. [‡]Hippocampal atrophy is measured according to the medial temporal lobe (MTA) visual rating scale for the ADC (0-4, higher scores reflect more severe atrophy [28]), and hippocampal volumes in mm^3 for ADNI and DCN, and z-scores (in which normalized MTA scores are inverted) for the pooled sample. [§]Posterior atrophy is scored using a visual rating scale for the ADC [29] and the DCN, inverted z-scores are given for the pooled sample. WMH are scored according to a visual rating scale for the ADC [30] and the DCN (0-3, higher scores reflect more sever pathology), WMH volumes for ADNI [33], and z-scores are given for the pooled sample.

Table 2 Demographical and neurobiological cluster characteristics

	ADC		ADNI		DCN		UCSF		Pooled sample	
	mem	**non-mem**	**mem**	**non-mem**	**mem**	**non-mem**	**mem**	**non-mem**	**mem**	**non-mem**
	n = 352	n =144	n = 182	n = 194	n = 335	n = 186	n = 326	n = 263	n = 1195	n = 787
	(71%)	(29%)	(48%)	(52%)	(64%)	(36%)	(55%)	(45%)	(60%)	(40%)
Demographics										
Age (years)	**67.7 ± 7.8**	**64.7 ± 8.4***	76.2 ± 7.2	74.1 ± 8.3	72.3 ± 7.8	71.1 ± 9.0	72.1 ± 9.7	70.3 ± 11.0	**71.5 ± 8.7**	**70.4 ± 9.9‡**
Female	180 (51%)	78 (54%)	78 (43%)	87 (45%)	201 (60%)	112 (60%)	171 (52%)	157 (60%)	630 (53%)	434 (55%)
Education	0.03 ± 1.00	-0.06 ± 1.01	0.01 ± 0.97	-0.01 ± 1.03	0.03 ± 1.00	-0.05 ± 1.00	0.1 ± 1.00	-0.13 ± 0.99	**0.04 ± 0.99**	**-0.07 ± 1.01‡**
Duration complaints	3.1 ± 2.2	2.7 ± 1.7	-	-	2.7 ± 2.3	2.4 ± 1.8	-	-	**2.9 ± 2.3**	**2.5 ± 1.8‡**
Global cognition										
MMSE	**22.6 ± 3.0**	**21.5 ± 3.2†**	23.5 ± 1.9	23.0 ± 2.3	23.4 ± 2.6	22.9 ± 3.0	**24.4 ± 3.1**	**22.7 ± 3.8***	**23.5 ± 2.8**	**22.6 ± 3.2***
APOE ε4 genotype										
APOE ε4 positive	225 (70%)	74 (59%)	72 (67%)	56 (63%)	175 (67%)	80 (58%)	**61 (66%)**	**39 (48%)‡**	**533 (68%)**	**249 (58%)***
AD biomarkers										
AD biomarker available	275 (78%)	114 (79%)	52 (29%)	50 (26%)	123 (38%)	70 (38%)	25 (8%)	28 (11%)	475 (40%)	262 (33%)
AD biomarker positive	253 (92%)	105 (92%)	40 (77%)	40 (80%)	106 (86%)	58 (83%)	24 (96%)	28 (100%)	423 (89%)	231 (88%)
MRI										
Hippocampus	-0.07 ± 1.00	0.18 ± 0.97	-0.05 ± 1.01	0.05 ± 1.00	-0.09 ± 0.95	0.17 ± 1.07	-	-	**-0.07 ± 1.00**	**0.12 ± 1.00†**
Posterior cortex	**0.11 ± 0.94**	**-0.29 ± 1.10†**	-	-	**0.08 ± 0.98**	**-0.15 ± 1.03‡**	-	-	**0.09 ± 0.96**	**-0.21 ± 1.06***
WMH	0.10 ± -0.01	-0.07 ± -0.01	0.01 ± -0.32	-0.01 ± 0.95	-0.04 ± 0.96	0.07 ± 1.07	-	-	0.03 ± -0.01	0.00 ± 1.05

Data are presented in number (%) or mean ± standard deviation, also when not-normally distributed enabling clearer comparison between clusters. p-values are based on t-tests, χ^2, or Kruskal Wallis analyses when appropriate. Normalized values are given for education, and MRI characteristics. AD biomarkers are available as cerebrospinal fluid total tau/amyloid β_{1-42} (abnormal when > 0.52 according to Duits e.a. [38]) in the ADC ADNI, and DCN cohorts, or as Pittsburgh compound B positron emission tomography positivity in the UCSF cohort. When MRI atrophy characteristics were measured using a visual rating scale in which a higher score reflects more severe atrophy, results were inverted so that higher scores reflect less atrophy. Differences between memory and non-memory clusters are shown in bold and indicated as follows: *p ≤ 0.001, †p ≤ 0.01, ‡p ≤ 0.05. Interpretation: The non-memory clusters were younger (pooled sample, ADC), less educated (pooled sample), had shorter duration of complaints (pooled sample), lower MMSE scores (ADC, UCSF, pooled sample), were more often APOE e4 negative (UCSF, pooled sample), had less hippocampal atrophy (pooled sample), but more severe atrophy of the posterior cortex (ADC, DCN, pooled sample) than the memory clusters.

FIGURE LEGENDS

Figure 1. Nonnegative Matrix Factorization is a dual-clustering approach, meaning that clustering includes two parallel steps. Firstly, neuropsychological (NP) test results are grouped together into neuropsychological profiles ('components'), illustrated in the upper half of the figure, in which each row represents one neuropsychological test, and each column an identified neuropsychological component. The warmer the colour, the higher the test loads to the component when test scores are high (relatively spared cognition); the colder the colour, the lower the test loads to the component when test scores are high (relatively impaired cognition). The optimal number of components is based on the cophenetic correlation coefficient (for this example n=2). Secondly, patients are grouped together (into 'clusters') based on the fit of their neuropsychological profile to the identified neuropsychological component, taking each test's load to the component into account. This step is illustrated in the lower half of the figure, in which each row represents one patient. The warmer the colour, the better the fit of patients' neuropsychological profile to the neuropsychological profile of the identified component.

Figure 2. Memory tests are indicated in dark blue. Abbreviations: ABCD = Arizona Battery for Communication Disorders of Dementia, ADC = Amsterdam Dementia Cohort, ADNI = Alzheimer's Disease Neuroimaging Initiative, comp questions = comparative questions, CVLT = California Verbal Learning Test, DCN = German Dementia Competence Network, DS = Digit Span, FAB = Frontal Assessment Battery, LDST = Letter Digit Substitution Test, LM = CERAD Logical Memory, mem = memory-impaired, non-mem = memory-spared, NMF = Nonnegative Matrix Factorization, RAVLT = Rey Auditory Verbal Learning Test, TMT = Trail Making Test, UCSF = University of California – San Francisco, VAT = Visual Association Test, WL = CERAD Word List. In all cohorts, the optimal number of test clusters was two. In this figure, each row represents one neuropsychological test. The two columns represent the two found neuropsychological components. The warmer the colour, the higher the test loads to the component when test scores are high (relatively spared cognition); the colder the colour, the lower the test loads to the component when test scores are high (relatively impaired cognition). Interpretation: In each cohort, one component is associated with relative impairment of memory tests, therefore called the memory component. The other component is associated with relative impairment of non-memory functions, therefore called the non-memory component.

Figure 3. Abbreviations: ADC = Amsterdam Dementia Cohort, ADNI = Alzheimer's disease Neuroimaging Initiative, DCN = German Dementia Competence Network, UCSF = University of California – San Francisco. Patients were assigned to either the memory or non-memory cluster based on the fit of their neuropsychological profile to the memory or non-memory component (figure 2). Each column represents one patient. The warmer the colour, the better the fit of patients' neuropsychological profile to the neuropsychological profile of that component.

ACKNOWLEDGEMENTS

CONFLICTS OF INTEREST

Revisions cognitive AD clusters robustly identified in four cohorts   *19/01/2017*

We thank the reviewers for their careful reading and thoughtful comments. We have revised the manuscript according to their suggestions and feel that the paper has improved considerably. Please find below our response in which we describe how we dealt with each comment in a point-by-point fashion. Changes made to the manuscript have been written in blue.

REVIEWER #1

This is a very well written manuscript that looks at clustering groups of AD patients according to their clustered performance on neuropsychological tasks. Although I am unfamiliar with this statistical technique, I could understand the method as described in the manuscript. The authors then compared the memory and non-memory clusters of patients to determine their differences. Broadly, the non-memory group tended to be younger, non-APOE4 carriers and showed posterior atrophy.

1. I only have one comment, and this is more a musing that I wonder if the authors have thought about: investigations of AD populations are difficult due to the issue of survivor bias (both in terms of the AD patients that are left to study, and the AD patients that are too unwell to be studied). While this is a large grouping of participants across many studies, I still always wonder about what inevitable bias we cannot avoid. For instance, is it the fact that there are more APOE4 carriers in the memory group, because of some curiosity to do with the non-memory group rather than a "real" biological distinction between groups of AD patients? Again, this is not something that can adequately be addressed here, but something that is often worried about by epidemiologists and I am curious to know whether the authors have thought about this.

   We thank the editor for his/her friendly words and for the interesting consideration. Indeed, it is very important to consider possible sources of bias. We excluded patients with MMSE < 17/30 to limit the influences of floor effects on neuropsychological tests that were used for clustering, and because cognitive heterogeneity is most prominently recognizable in the early stages. Please note that we included data of patients who received a diagnosis of probable AD at their' first visits, therefore data of earlier stages were not available. We think that patients' and doctors' delay is probably shorter in the more typical (well-known) memory AD than in the less typical non-memory AD, as the latter is probably less well recognized in early stages and more often misdiagnosed (e.g. as burn-out, depression). We therefore think that – if anything - this selection bias probably underestimated the presence of non-memory AD phenotype in our study. Suggested underestimation of the non-memory phenotype is strengthened by the finding that non-memory patients reported shorter disease duration, while MMSE scores were already lower at first visit, suggesting faster disease progression before diagnosis and higher risk for more severe dementia at time of diagnosis. We added considerations regarding bias in the discussion (page 12).

REVIEWER #2

This is a paper on clusters of cognitive profiles in AD. The authors report that there are two clusters in patients with a clinical diagnosis of AD dementia, one with predominance of memory impairment and one with predominance of impairment in other cognitive domains. In general, the non-memory cluster was younger, had shorter disease duration, lower MMSE, were more often APOE e4 negative, and had more posterior and less hippocampal atrophy. Together this suggests that there are at least two different phenotypes of AD, with differences both in terms of clinical presentation, demographics, and brain changes. The study is strengthened by the fact that main findings were replicated in 4 independent cohorts.

Major comments:

1. This paper has already been revised. One of the revisions was a validation of the findings in the ADC subgroup with pathological CSF T-tau/AB42 ratio. However, in the copy that I have reviewed this result is only mentioned briefly in the discussion. I believe that an important limitation of this paper is the risk of clinical misdiagnosis of AD. The biomarker-based validation is therefore very important and needs to be thoroughly described in the results section.

    We are not sure what the reviewer means with 'already revised', because this paper has not been sent out for review before; we have however answered some questions on request of the editor before the manuscript was sent out for review. Among these considerations was an exploration of biomarker confirmed cases in our cohorts. Sufficient numbers of CSF biomarkers were only available for the ADC cohort. When we repeated clustering in the subset of biomarker confirmed ADC patients, we found comparable cognitive clusters, with similar demographic and neurobiological characteristics. On the reviewer's suggestion, we have added these results to the manuscript (page 10).

2. I think it is promising that you could replicate the main findings when restricting the analysis to the ADC cohort with an AD-pattern of CSF biomarkers. But to really address the question of possible misdiagnosis of subjects I think you need to report the percentages of people with pathological biomarkers in the different clusters and centres. For example, what was the percentage of pathological CSF Ab42 in memory and non-memory clusters in the different cohorts? The title of your paper implies that you are only testing patients with AD. I therefore think it is necessary to take every possible step to show that there was no systematic bias caused by clinical misdiagnosis. For example, was the percentage of normal CSF AB42 higher in the non-memory group in any cohort where this could be tested?

    We see the reviewer's point on the misleading title and lack of biomarker confirmation. Since our study focused on clinical AD diagnosis (i.e. probable AD according to the NINCDS-ADRDA criteria[1]), we have tried to limit the expectation that our study is based on biomarker confirmed data and rephrased the title (page 1). In addition, we have summarized requested information in the manuscript (page 10), and provided requested data in table 1 and 2. Please note that for cohorts other than ADC, availability of CSF biomarkers was not optimal.

3. Another potential bias that you need to control for is differences in vascular pathology/white matter pathology between the different clusters. Please include data on for example white matter lesions and vascular risk factors and show how these differ between clusters. If one of the clusters has less prevalence of amyloid pathology but higher prevalence of white matter pathology there is a risk that clinical misdiagnosis affected the results.

   We compared cognitive clusters in terms of white matter hyperintensities in the ADC, ADNI, and DCN cohorts (not available for UCSF). No differences were found between the memory and non-memory clusters ($p > 0.05$). We added results to tables 1 and 2.

4. The discussion section largely reiterates the results, and should be expanded with details on how these novel findings relate to previous findings of clinical and pathological heterogeneity in AD. For example, there is a rich literature on co-pathologies in patients with a clinical AD diagnosis. To what degree can that help explain your findings? You could also incorporate more discussions about pathogenic mechanisms. You mention this very briefly ("different underlying disease anatomy") but I think the paper would really benefit from a discussion about how these different clinicopathological phenotypes may arise.

   We thank the reviewer for the opportunity to extend our discussion section (page 11-12). We think this has enriched our manuscript.

Minor comments:

1. I agree with the authors that it was surprising that the non-memory cluster was so prevalent in ADNI. In fact, at 52 % it was even bigger than the memory cluster. You should therefore correct the sentence in the discussion which says that "In each cohort, the majority of patients belonged to the memory cluster". There is also a typo in section 3.3 where it says that "In the ADNI cohort, non-memory patients tended to be younger and less often APOE e4 negative". This should probably be "less often APOE e4 positive".

   We thank the reviewer for noticing and changed the text according to the suggestion (page 10, 11).

2. Depending on the results of this revision you can consider changing the title to highlight that you used a clinical AD diagnosis.

   We agree with the reviewer and rephrased the title (page 1), including the clinical diagnosis 'probable AD' that we used as inclusion criterion [1].

REFERENCES

[1]   McKhann G, Drachman D, Folstein M, Katzman R, Price D, Stadlan EM. Clinical diagnosis of Alzheimer"s disease: report of the NINCDS-ADRDA Work Group under the auspices of Department of Health and Human Services Task Force on Alzheimer"s Disease. Neurology 1984;34:939–44.
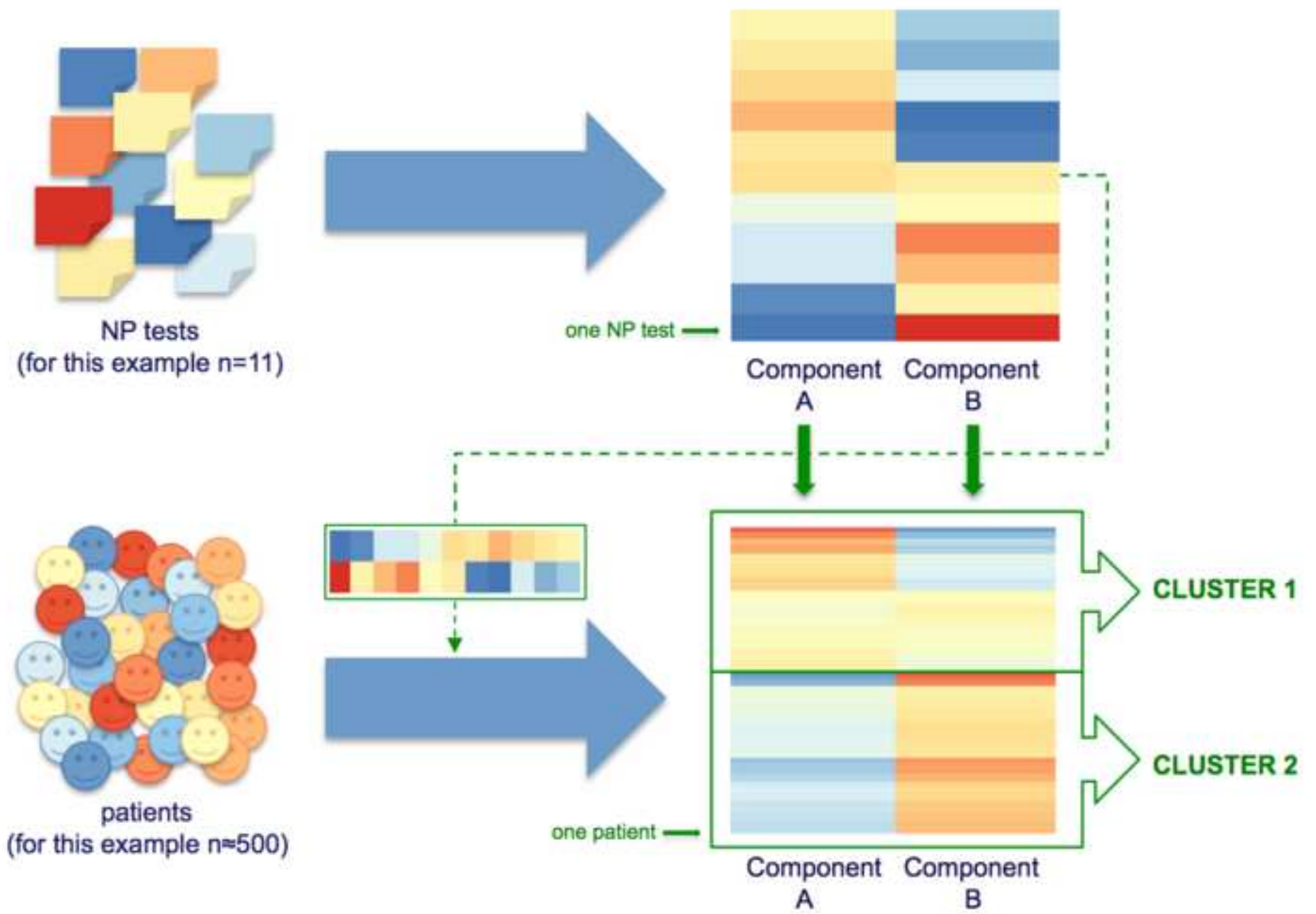
RESEARCH IN CONTEXT

1.Systematic review: Alzheimer's disease (AD) is characterized by cognitive heterogeneity. We searched PubMed for clinical and neurobiological heterogeneity in AD profiles, and for data-driven approaches used to identify AD subtypes based on neuropsychological test scores. Several studies demonstrated the potential of clustering methods to identify cognitive AD subtypes. Identified subtypes showed distinct clinical characteristics. However none of the previous studies tested the generalizability of the cluster solutions, since those results were based on single-cohort studies.

2. Interpretation: In four large AD cohorts, we consistently identified two cognitive clusters using Nonnegative Matrix Factorization of neuropsychological test scores. In each cohort one cluster most prominently showed memory-impairment, and the other cluster was relatively memory-spared. These clusters were associated with distinct clinical characteristics.

3. Future directions: Future research should aim to further study the underlying biological disease mechanisms that cause a non-memory AD phenotype.
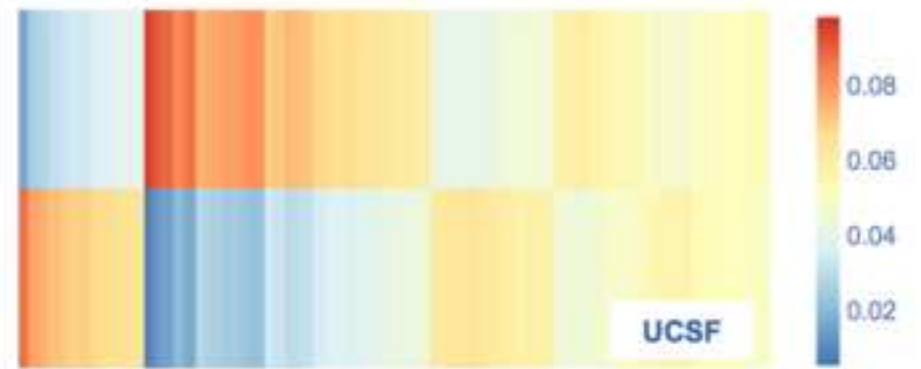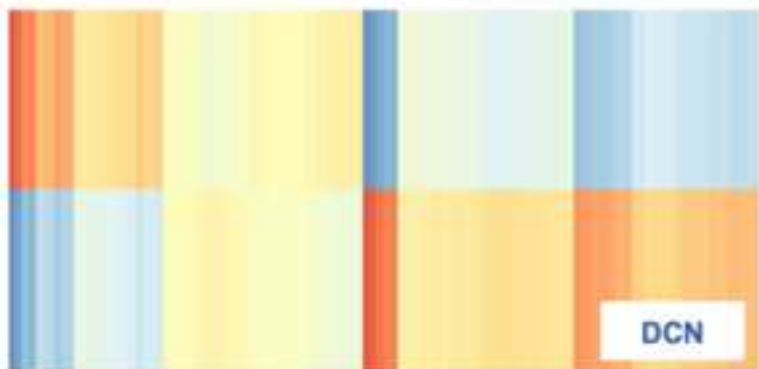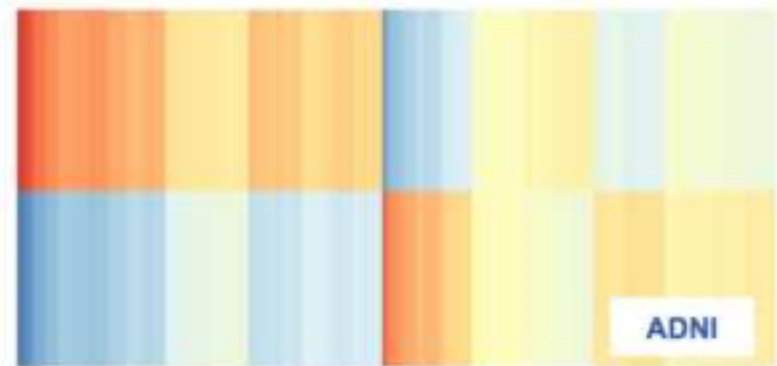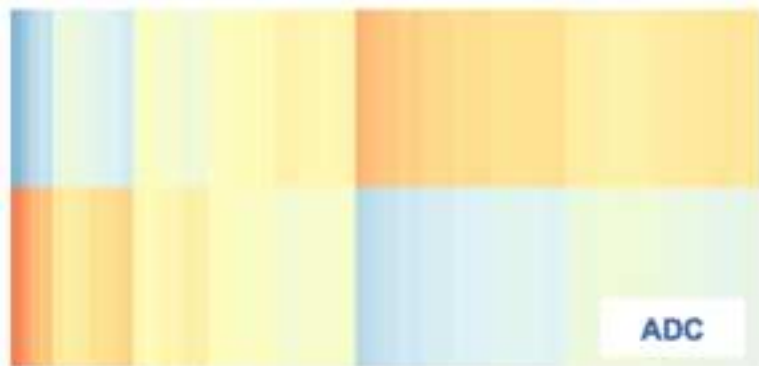
NP tests
(for this example n=11)

one NP test —

Component   Component
A   B

patients
(for this example n≈500)

one patient —

Component   Component
A   B

CLUSTER 1

CLUSTER 2

memory

non-memory

ADC

ADNI

memory

non-memory

DCN

UCSF

0.08

0.06

0.04

0.02

Supplementary files

Click here to download Supplementary files: Supplementary.docx

Supplementary files marked revisions