Journal of Alzheimer's Disease xx (2021) x–xx
DOI 10.3233/JAD-210719
IOS Press

Heterogeneity and Factorial Structure in Alzheimer's Disease: A Cognitive Perspective

Andrea Zangrossi[a,b,*], Sonia Montemurro[c], Gianmarco Altoè[d] and Sara Mondini[e,f]
[a]*Department of Neuroscience, University of Padua, Padua, Italy*
[b]*Padova Neuroscience Center (PNC), University of Padua, Padua, Italy*
[c]*IRCCS San Camillo Hospital, Venice, Italy*
[d]*Department of Developmental and Social Psychology, University of Padua, Padua, Italy*
[e]*Department of Philosophy, Sociology, Pedagogy and Applied Psychology, University of Padua, Padua, Italy*
[f]*Human Inspired Technology Research Centre, University of Padua, Padua, Italy*

Accepted 21 July 2021

Abstract.
Background: Alzheimer's disease (AD) patients show heterogeneous cognitive profiles which suggest the existence of cognitive subgroups. A deeper comprehension of this heterogeneity could contribute to move toward a precision medicine perspective.
Objective: In this study, we aimed 1) to investigate AD cognitive heterogeneity as a product of the combination of within- (factors) and between-patients (sub-phenotypes) components, and 2) to promote its assessment in clinical practice by defining a small set of critical tests for this purpose.
Methods: We performed factor mixture analysis (FMA) on neurocognitive assessment results of N = 230 patients with a clinical diagnosis of AD. This technique allowed to investigate the structure of cognitive heterogeneity in this sample and to characterize the core features of cognitive sub-phenotypes. Subsequently, we performed a tests selection based on logistic regression to highlight the best tests to detect AD patients in our sample. Finally, the accuracy of the same tests in the discrimination of sub-phenotypes was evaluated.
Results: FMA revealed a structure characterized by five latent factors and four groups, which were identifiable by means of a few cognitive tests and were mainly characterized by memory deficits with visuospatial difficulties ("Visuospatial AD"), typical AD cognitive pattern ("Typical AD"), less impaired memory ("Mild AD"), and language/praxis deficits with relatively spared memory ("Nonamnestic AD").
Conclusion: The structure of cognitive heterogeneity in our sample of AD patients, as studied by FMA, could be summarized by four sub-phenotypes with distinct cognitive characteristics easily identifiable in clinical practice. Clinical implications under the precision medicine framework are discussed.

Keywords: Alzheimer's disease, factor mixture analysis, heterogeneity, precision medicine

INTRODUCTION

Clinical heterogeneity in neurological practice is a critical and underestimated issue, highly impacting both diagnosis and prognosis (for a review, see [1]). Indeed, interindividual differences in clinical

*Correspondence to: Andrea Zangrossi, Padova Neuroscience Center (PNC), University of Padua, Via Giuseppe Orus 2, 35131 Padua, Italy. Tel.: +39 049 8212623; E-mail: andrea.zangrossi@unipd.it.

manifestation of a disease may reveal biological or epigenetic differences [2] which could strongly affect drug mechanism of action [3], and effectiveness of other treatments (e.g., cognitive training) [4]. Heterogeneity has been studied in many neurological disorders, including psychosis [5], schizophrenia [6], stroke [7, 8], Parkinson's disease [9], and multiple sclerosis [10]. Taken together, these findings highlight the need of a paradigm shift toward precision medicine. This issue is particularly relevant in neurodegenerative diseases where clinical phenotypes reflect the combination of heterogeneity in brain aging [11], age-related cognitive decline [12], and baseline individual differences. This would lead to high variance both in behavioral and *in vivo* biomarkers, seriously misguiding the disease understanding, as in the case of Alzheimer's disease (AD) [13]. According to the DSM-5, the core symptom for the diagnosis of neurocognitive disorder due to AD is a progressive decline in memory, with alteration of at least one other cognitive domain. In clinical practice, however, the pattern of cognitive deficits observed in AD patients is highly variable and, according to the Alzheimer Precision Medicine Initiative (APMI), there is a strong need for patient-tailored interventions accounting for individual-specific biological profiles [3, 14]. For this reason, it is of crucial importance that research on AD focus on clinical variability, in terms of possible sub-phenotypes [13].

Previous research has characterized cognitive heterogeneity in AD, through theory-driven approach (e.g., [15]). However, only a few studies have dealt with this issue in a data-driven manner. For example, Cappa and colleagues [16] suggested the existence of four sub-phenotypes mainly characterized by the differential impairment of visuospatial/perceptual abilities, memory, perception, calculation, and language. Other studies have shown AD patients either classifiable on eight clusters of cognitive features [17], or simply based on the presence/absence of memory impairment [18]. Taken together, these studies suggest the presence of cognitive AD subphenotypes, but the number of clusters explaining variability across profiles is not clear, yet. One of the reasons behind this lack of consensus is that no studies have combined the investigation of inter-individual differences with that of intra-individual latent factors, which could lead to a finer understanding of the structure of AD cognitive heterogeneity.

In the present study, we aimed at investigating cognitive sub-phenotypes in AD through a relatively new approach for the study of heterogeneity, namely the factor mixture analysis (FMA) [19, 20]; see Methods for details), whose effectiveness has been proven in different domains, including mild cognitive impairment (MCI) and dementia [21]. The strength of this method is that it fosters a finer-grained description of heterogeneity compared to standard approaches. Indeed, by employing a hybrid/combination of categorical and continuous latent variables, FMA allows both to study heterogeneity at the group-level (i.e., classifying individuals into subgroups) and to describe heterogeneity within subgroups [22]. This technique is specifically suitable for our purpose since it allows to identify both the latent factors (i.e., linear combination of cognitive scores) and the potential sub-phenotype of patients who share common cognitive patterns (see the methods section for more details). This approach could help mapping clinical heterogeneity in AD, thus contributing to the implementation of the precision medicine perspective in clinical neuropsychology practice. Furthermore, our second aim was to find the minimum set of cognitive tests to effectively and rapidly highlight such features in clinical routine, under the hypothesis that extensive neuropsychological batteries may be effectively reduced to a smaller set of critical and essential tests without losing diagnostic accuracy and quality in the description of the cognitive profile, and maximizing resources [23].

MATERIALS AND METHODS

Participants and procedure

The study group is a retrospective sample of N = 268 consecutive patients selected from a larger cohort of patients with neurological disorders referring to the neuropsychological service of the University of Padua (Italy). Inclusion criteria were: 1) clinical diagnosis of probable AD based on the criteria of the National Institute of Neurological and Communicative Disorders and Stroke and the Alzheimer's Disease and Related Disorders Association (NINCDS-ADRDA) [24]; 2) availability of the Mini-Mental State Examination (MMSE) [25] score within an extensive cognitive assessment (i.e., Esame Neuropsicologico Breve 2 - ENB2 [26]; see Supplementary Table 1). Pathophysiological biomarkers were not available in this retrospective sample, but all patients included in the final sample showed a clinical phenotype of AD, in line with the latest recommendations [27]. Patients showing comorbidity with psychiatric or other neurological diseases

were excluded. Only patients coming for the first time at the neuropsychological service for clinical assessment were included in the study. Thirty-eight patients were discarded due to missing data; thus, the final sample was composed of N = 230 patients (age range: 58–93, $M_{age} = 77.1$, $SD_{age} = 6.3$; $M_{education} = 7.3$, $SD_{education} = 3.9$; $M_{MMSE} = 21.5$, $SD_{MMSE} = 3.5$, 151 F). Furthermore, a sample of N = 326 age- and education-matched healthy controls (HC), who were administered the same neuropsychological assessment, was matched with the sample of patients. All participants gave their written consent for the anonymous use of the data. The study was conducted in accordance with the Declaration of Helsinki and was approved by the Ethical Committee for the Psychological Research of the University of Padova.

Statistical analysis

Factor mixture analysis (FMA) to study heterogeneity

Heterogeneity in our sample of AD patients was investigated by means of a statistical technique called FMA [19, 20], which allows to evaluation of the factorial structure of a phenomenon while simultaneously investigating the existence of sub-populations (i.e., clusters of participants) [20, 28], without assuming that all participants in a sample are representative of the same population, as traditional factor analysis models do. In particular, FMA goes beyond standard factor analysis since it does not rely on the assumption that factors are normally distributed. Furthermore, it assumes that correlations between latent factors could vary across subpopulations, thus allowing to identify clusters within a heterogeneous sample [6]. Finally, FMA assumes a parametric structure within each class and can be used to test a series of structural hypothesis: in this way it allows to understand complex phenotypic structures that are simultaneously categorical and dimensional [22, 29–31].

For these reasons, FMA is a suitable technique to model the underlying structure of psychological [19] and psychopathological [22] constructs. Recent studies have shown that FMA can be useful in the identification of sub-groups in HC, MCI, and dementia [21]. We thus decided to adopt the FMA to investigate the presence of cognitive sub-phenotypes in AD and to describe their key features. The FMA was applied on data from the whole cognitive battery except one test (i.e., token test) which was discarded due to its null variability.

Importantly, we adopted an exploratory approach in order to highlight the most reliable model of AD cognitive heterogeneity. To this end, AD cognitive scores were first scaled on HC data, then we estimated 49 FMA models by testing all the combinations from 1 up to 7 factors, and from 1 up to 7 groups to find the best combination fitting our data. The Bayesian Information Criterion (BIC) [32] was calculated for each model and the one with the lowest BIC was chosen as indicating the most plausible combination of latent factors and groups. Only the results relative to the best model will be reported and discussed.

Selection of the best subset of tests

Our second aim was to find the core set cognitive tests with the highest diagnostic accuracy (i.e., in discriminating AD versus HC). To this end, we ran a logistic regression model with participants' status (either AD or HC) as dependent variable, and the whole set of tests as predictors. Then, this model was used as input for a backward stepwise procedure based on the Akaike Information Criterion (AIC) [33], which returned the best set of tests for the prediction of participants' status. In order to control for the variability among HC data, and to match sample sizes, this procedure was repeated 1000 times, each time randomly selecting 230 out of 326 HC to match AD sample size, and the logistic model was run on a dataset of N = 460 (230 AD and 230 HC). Thus, each iteration resulted in a selection of tests providing the highest classification accuracy between AD and the random sample of HC. Tests selected in $\geq$95% iterations were included in the best subset. As a control analysis, we tested the efficacy of this subset in the discrimination between AD and HC, and, more importantly, in the detection of AD sub-phenotypes. In other words, the set of tests which best detected AD patients was tested also to identify individual cognitive sub-phenotypes. To this end, we employed three machine-learning classifiers, i.e., Random Forest (RF), Support Vector Machine (SVM), and Naïve Bayes (NB) with a 10-folds cross-validation design (see the Supplementary Material for details). Again, the procedure was repeated 1000 times employing random selections of HC. Finally, the accuracy resulted from the selected tests was compared to that of the whole battery. All analyses were performed by means of R Software [34] and custom coding. The FMA was performed by means of the FactMixtAnalysis R package [35]. Machine learning analyses were performed by means of RWeka [36] R package.

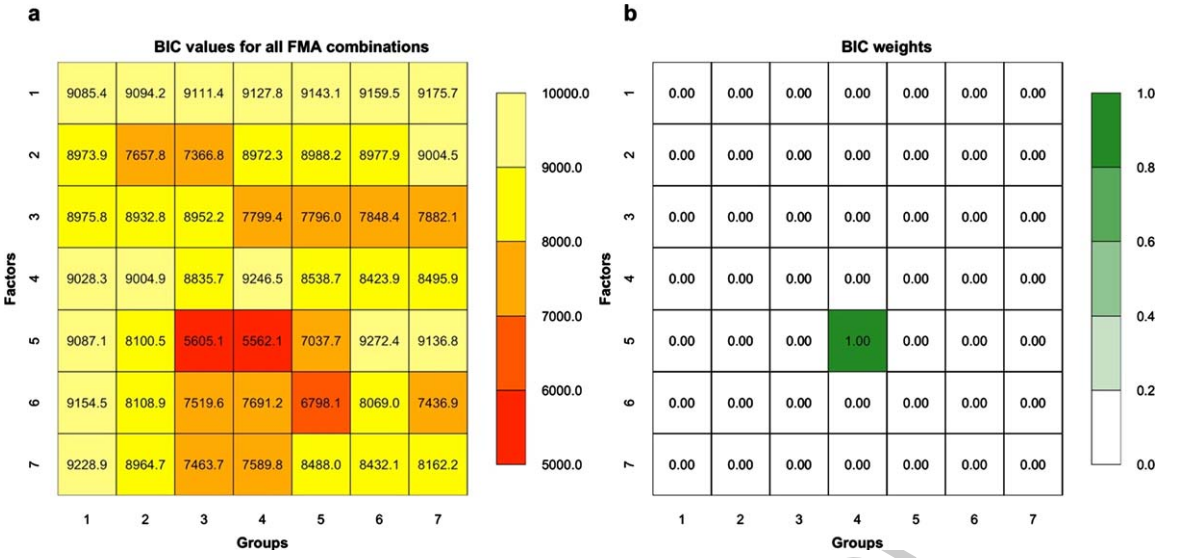

Fig. 1. Comparison of FMA models. a) Bayesian Information Criterion (BIC) for each FMA model. The minimum value of BIC indicates the best solution, i.e., 5 factors and 4 groups. b) BIC weights are computed in probability space, with 1 indicating 100% probability of being the best model compared to the alternatives.

RESULTS

Four cognitive sub-phenotypes of Alzheimer's disease

A BIC value was computed for all 49 FMA models (Fig. 1a) and the lowest BIC (i.e., best balance between model likelihood and parsimony) highlighted a model with five factors and four groups (i.e., clusters) as the solution that best fitted our data. We also computed BIC weights [37], a transformation of BIC values into a probability space (range: 0–1), which allows to quantify the evidence in favor of one model being better than the others. The model with 5 factors and 4 groups showed a rounded BIC weight close to 1, indicating a ∼100% probability of being the best solution within the set of tested models (Fig. 1b).

An oblique (Promax) rotation was applied to factor loadings (see Fig. 2) to improve their interpretation. The first factor (F1) loaded mainly on verbal memory tests, F2 on visuospatial abilities, F3 on working memory, while F4 mainly loaded on attention and executive functions, and F5 on language and praxis abilities.

According to this factorial structure, the sample of AD patients was split into four clusters including 20% (45/230), 18% (42/230), 46% (106/230), and 16% (37/230) patients, respectively (Fig. 3a), which loaded on different combinations of factors

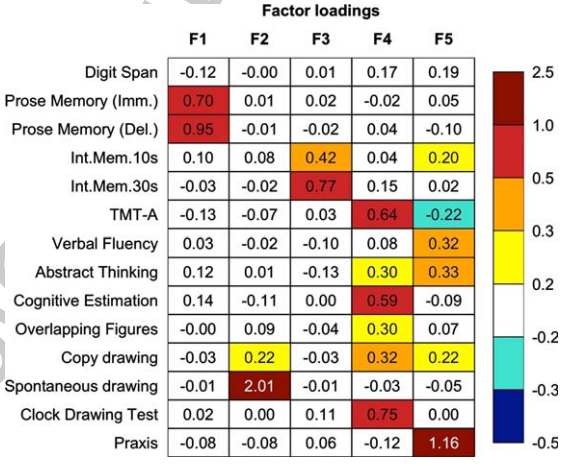

Fig. 2. Factor loadings of the latent components emerged in the FMA. The values were Promax rotated to improve interpretability. Columns from F1 to F5 correspond to the five-factor solution derived from FMA. Colored cells indicate the most important tests for each factor (i.e., loading values above a threshold of |0.2|). According to the highest loadings, each factor can be interpreted as follows. F1, verbal memory; F2, visuospatial abilities; F3, working memory; F4, attention and executive functions; F5, language and praxis abilities; Prose Memory (Imm.), immediate recall prose memory; Prose Memory (Del.), short-delayed recall prose memory; Int.Mem.10 s, interference memory (10 seconds); Int.Mem.30 s, interference memory (30 seconds); TMT-A, Trail Making Test A.

(Fig. 3b). A deeper look into clusters' cognitive profiles revealed that in profiles belonging to Cluster 1, memory difficulties were mainly accompanied by

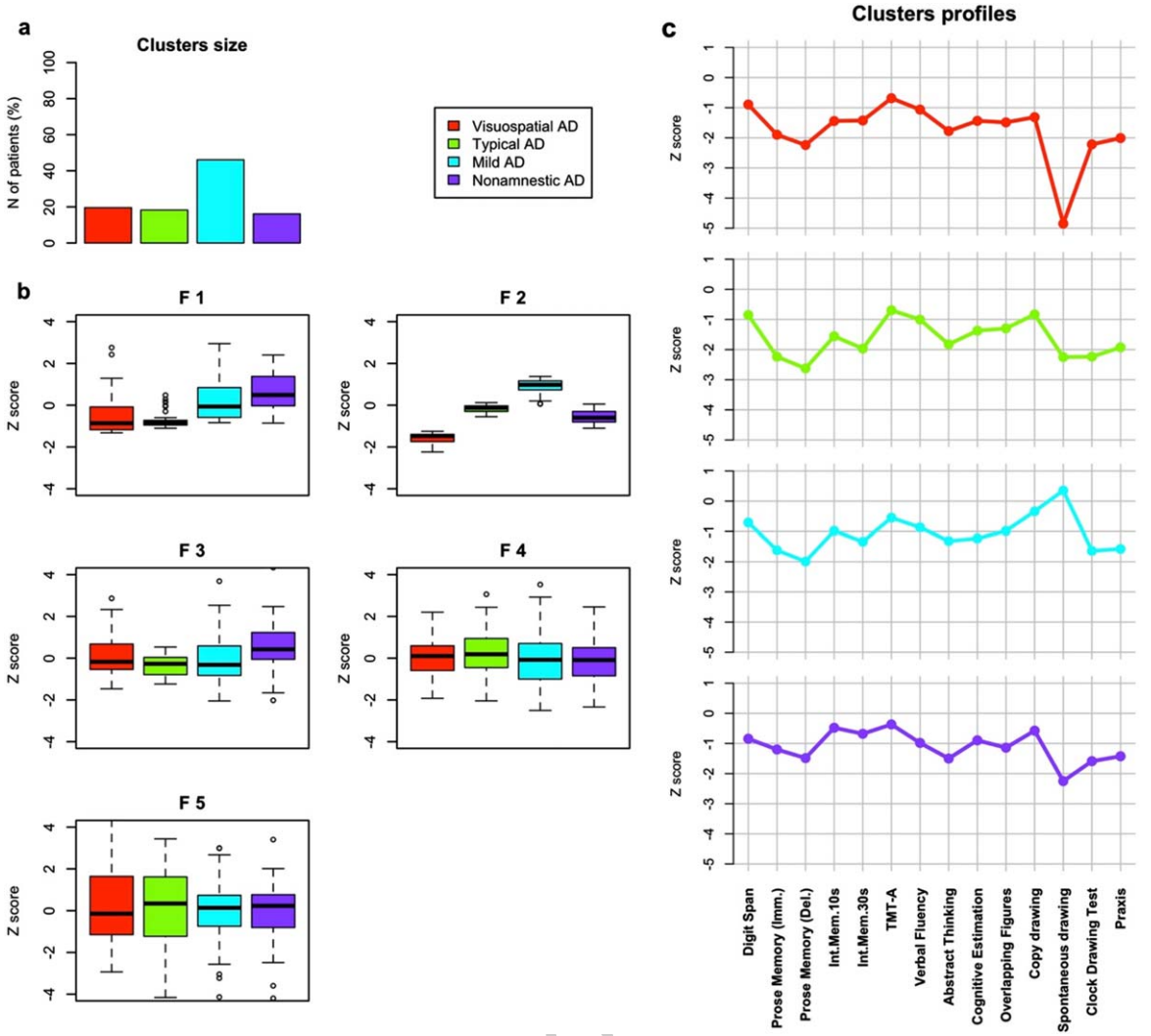

Fig. 3. Clusters (i.e., sub-phenotypes) characterization. a) Clusters size distribution. b) Clusters comparison across factors. F1, verbal memory; F2, visuospatial abilities; F3, working memory; F4, attention and executive functions; F5, language and praxis abilities. c) Mean normalized score obtained by each cluster in the different cognitive tests. Notably, TMT-A score (time in seconds) was transformed in a velocity measure (i.e., 25 items/time) to be comparable with the other measures (i.e., higher values indicate better performance). Each score was z-scored on the HC sample (N = 326). Prose Memory (Imm.), immediate recall prose memory; Prose Memory (Del.), short-delayed recall prose memory; Int.Mem.10 s, interference memory (10 seconds); Int.Mem.30 s, interference memory (30 seconds); TMT-A, Trail Making Test A.

visuospatial deficits (F2), thus we called this cluster "Visuospatial AD". In Cluster 2 patients showed the typical AD cognitive pattern, characterized by predominant memory deficits, for this reason this cluster can be labelled as "Typical AD". On the other hand, Cluster 3 showed a less impaired memory performance, thus can be called "Mild AD". Finally, Cluster 4 was mainly explained by deficits in language and praxis abilities (F5), with relatively spared memory, thus this cluster could be labelled as "Nonamnestic AD" (Fig. 3c). Noteworthy, clusters should not be considered as being associated with a single cognitive feature, but as patterns distinguishable from each other based on peculiar cognitive weaknesses. For a clearer clinical interpretation of clusters' cognitive profiles, summary statistics of cognitive scores are reported in Table 1.

Previous findings have shown that age, sex, and education might impact AD heterogeneity and drive diverging pathophysiologic paths across subtypes [38]. Thus, we checked whether these variables, as well as MMSE score, could explain our clusters

Table 1
The table reports Mean, SD and 95% Confidence Interval (CI) of cognitive scores for each cluster of patients (z-scored on N = 326 HC). Prose Memory (Imm.), immediate recall prose memory; Prose Memory (Del.), short-delayed recall prose memory; Int.Mem.10 s, interference memory (10 seconds); Int.Mem.30 s, interference memory (30 seconds); TMT-A, Trail Making Test A

Test	Cluster 1 Visuospatial AD		Cluster 2 Typical AD		Cluster 3 Mild AD		Cluster 4 Nonamnestic AD	
	M (SD)	CI 95%	M (SD)	CI 95%	M (SD)	CI 95%	M (SD)	CI 95%
Digit Span	–0.9 (0.8)	[–1.1,–0.6]	–0.9 (0.9)	[–1.1,–0.6]	–0.7 (0.8)	[–0.9,–0.6]	–0.8 (0.8)	[–1.1,–0.6]
Prose Memory (Imm.)	–1.9 (1)	[–2.2,–1.6]	–2.2 (0.5)	[–2.4,–2.1]	–1.6 (0.9)	[–1.8,–1.4]	–1.2 (0.9)	[–1.5,–0.9]
Prose Memory (Del.)	–2.2 (0.9)	[–2.5,–2]	–2.6 (0.4)	[–2.8,–2.5]	–2 (1)	[–2.2,–1.8]	–1.5 (1)	[–1.8,–1.1]
Int.Mem.10s	–1.4 (0.9)	[–1.7,–1.2]	–1.6 (0.6)	[–1.7,–1.4]	–1 (0.9)	[–1.2,–0.8]	–0.5 (0.8)	[–0.7,–0.2]
Int.Mem.30s	–1.4 (0.8)	[–1.7,–1.2]	–2 (0.1)	[–2,–1.9]	–1.3 (0.8)	[–1.5,–1.2]	–0.7 (0.8)	[–0.9,–0.4]
TMT-A	–0.7 (0.5)	[–0.8,–0.5]	–0.7 (0.5)	[–0.8,–0.6]	–0.5 (0.6)	[–0.7,–0.4]	–0.4 (0.7)	[–0.6,–0.1]
Verbal Fluency	–1.1 (0.8)	[–1.3,–0.8]	–1 (0.7)	[–1.2,–0.8]	–0.9 (0.6)	[–1,–0.8]	–1 (0.5)	[–1.2,–0.8]
Abstract Thinking	–1.8 (1.3)	[–2.2,–1.4]	–1.8 (0.9)	[–2.1,–1.5]	–1.3 (1.2)	[–1.5,–1.1]	–1.5 (1)	[–1.8,–1.2]
Cognitive Estimation	–1.4 (1.4)	[–1.8,–1]	–1.4 (1.3)	[–1.8,–1]	–1.2 (1.3)	[–1.5,–1]	–0.9 (1.2)	[–1.3,–0.5]
Overlapping Figures	–1.5 (0.5)	[–1.6,–1.3]	–1.3 (0.6)	[–1.5,–1.1]	–1 (0.6)	[–1.1,–0.9]	–1.1 (0.4)	[–1.3,–1]
Copy drawing	–1.3 (1.4)	[–1.7,–0.9]	–0.8 (1.3)	[–1.2,–0.4]	–0.3 (1.3)	[–0.6,–0.1]	–0.6 (1)	[–0.9,–0.2]
Spontaneous drawing	–4.8 (0)	[–4.8,–4.8]	–2.2 (0)	[–2.2,–2.2]	0.4 (0)	[0.4,0.4]	–2.2 (0)	[–2.2,–2.2]
Clock Drawing Test	–2.2 (1.5)	[–2.7,–1.8]	–2.2 (1.5)	[–2.7,–1.8]	–1.6 (1.7)	[–2,–1.3]	–1.6 (1.7)	[–2.1,–1]
Praxis	–2 (2.4)	[–2.7,–1.3]	–1.9 (2.9)	[–2.8,–1]	–1.6 (2.2)	[–2,–1.1]	–1.4 (1.8)	[–2,–0.8]

by means of a logistic regression model. Significant main effects of sex ($\chi^2 = 18.3$, $p < 0.001$) and MMSE ($\chi^2 = 13.9$, $p < 0.001$) emerged. More specifically, between-clusters *post-hoc* comparisons suggested that the Nonamnestic AD patients (Cluster 4) where characterized by a better global cognition (MMSE score) than Visuospatial (Cluster 1; t[77.5] = –5.04, $p < 0.001$) and Typical AD patients (Cluster 2; t[72] = –3.9, $p < 0.001$), while Mild AD patients (Cluster 3) had significantly higher MMSE as compared to Visuospatial AD patients (Cluster 1; t[69.4] = –3.6; $p = 0.003$). All *p*-values were Bonferroni-corrected for multiple comparisons. Moreover, the proportion of females in Mild AD patients (Cluster 3) was significantly higher than in the other clusters (Cluster 1: $\chi^2[1] = 18.6$; Cluster 2: $\chi^2[1] = 37$; Cluster 4: $\chi^2[1] = 29.5$; all Bonferroni-corrected ps<0.001). This result indicates that, to some extent, AD cognitive heterogeneity might partially reflect gender-related and global cognitive functioning differences (see Supplementary Figure 1).

Precision medicine in clinical practice: cognitive sub-phenotypes are captured by few tests

The stepwise procedure (see Methods section) run on a logistic model for the discrimination of N = 230 AD versus N = 230 HC over 1000 iterations highlighted nine tests as the most critical for the diagnosis of AD (i.e., without distinguishing between sub-phenotypes; Fig. 4). This set of core tests included Digit span, Prose memory (delayed), TMT-A, Verbal fluency, Abstract thinking, Overlapping figures, Spontaneous drawing, Clock drawing test, and Praxis abilities.

As a control analysis, we checked whether the diagnostic accuracy (i.e., AD versus HC) of the subset of tests was comparable to that of the whole battery by means of three machine-learning algorithms using a 10-folds cross-validation design. All algorithms showed a mean accuracy > 87% (i.e., 90.7%, 89.2%, 87.7%, respectively), and the difference in the classification performance between the subset of tests versus the whole battery was negligible (see Fig. 5a), indicating that using the selected 9 tests did not have a negative impact on diagnostic accuracy.

We then tested the accuracy of the full and the reduced set of tests in the classification of cognitive sub-phenotypes (i.e., clusters) using the same classification approach. Accuracy obtained using the whole battery versus the selected tests is shown in Fig. 5b. Importantly, ceiling accuracy (i.e., overfitting) was expected when using the whole battery, since the phenotypes (clusters) were found on the same tests, thus the performance using the whole battery should be considered as a reference, while our focus was on the performance of the selected tests, which maintained a good classification accuracy (RF = 89.6%, SVM = 91.7%, NB = 90.4%) with a relatively small drop (7% on average) compared to the whole battery (See Supplementary Table 2 for further details on classification performance).

These results indicate that the selected tests can both identify critical core deficits for AD detection and capture cognitive sub-phenotypes (i.e., clusters).

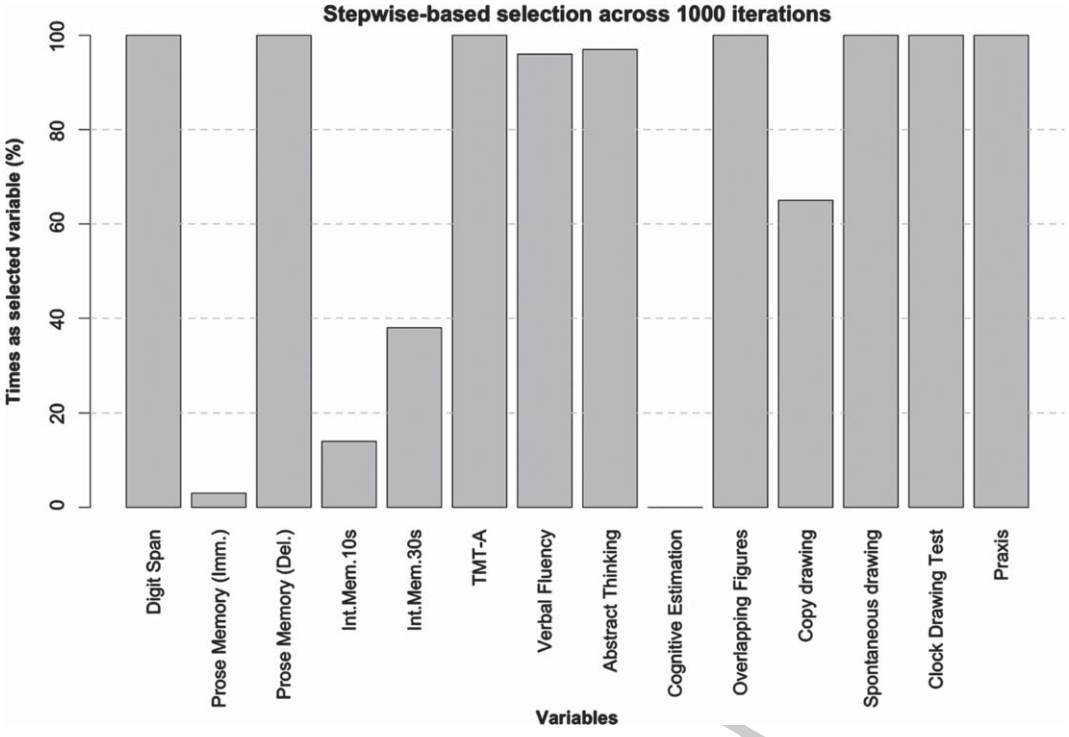

Fig. 4. Selection of tests for the classification between AD patients and healthy controls. A logistic regression model was built for 1000 iterations, each time on N = 230 AD and a random selection of N = 230 (out of 326) HC. A stepwise regression procedure was run for each iteration and the best predictors (i.e., tests) in the discrimination between AD and HC were highlighted. The tests which resulted as the best predictors in > 95% of iterations were selected. Prose Memory (Imm.), Immediate recall prose memory; Prose Memory (Del.), Delayed recall prose memory (5 minutes delay); Int.Mem.10 s, Interference memory (10 seconds); Int.Mem.30 s, Interference memory (30 seconds); TMT-A, Trail Making Test A.

This suggests that a quick cognitive assessment based on a few tests might potentially be useful in clinical practice for identification and cognitive phenotypization of AD patients.

DISCUSSION

Precision medicine is a field of medicine which aims to optimize effectiveness of disease treatment (or prevention) by taking into account specific individual characteristics. In this study, we contribute to this approach by studying heterogeneity of cognitive profiles in a sample of patients with a clinical diagnosis of AD. We first aimed at investigating the presence of latent factors and how they combine to create clusters of patients with similar profiles (i.e., cognitive sub-phenotypes). Secondly, we aimed at supporting the implementation of this approach in the clinical routine by identifying a core set of cognitive tests able to characterize sub-phenotypes at the individual level.

We evaluated the cognitive heterogeneity in AD patients by means of FMA, a relatively novel approach which could allow to overcome some of the limitations of dimensionality reduction and cluster-analysis techniques adopted in previous research on this topic. This approach suggested a model with five factors and four cognitive clusters as the most suitable towards explaining our data. The latent factors were mainly grounded on memory (F1), visuospatial abilities (F2), working memory (F3), attention and executive functions (F4), and language and praxis abilities (F5). Along these dimensions, four cognitive sub-phenotypes were shown. The most represented (46% of patients) was called Mild AD (Cluster 3) since it was characterized by a mild general impairment. The Visuospatial AD cluster (Cluster 1) included 20% of patients, whose cognitive profile was mainly characterized by visuospatial deficits. Then, 18% of patients belonged to the Typical AD cluster (Cluster 2) which was characterized by a homogeneous cognitive profile with deficits primarily affecting memory performance. Finally, 16% of

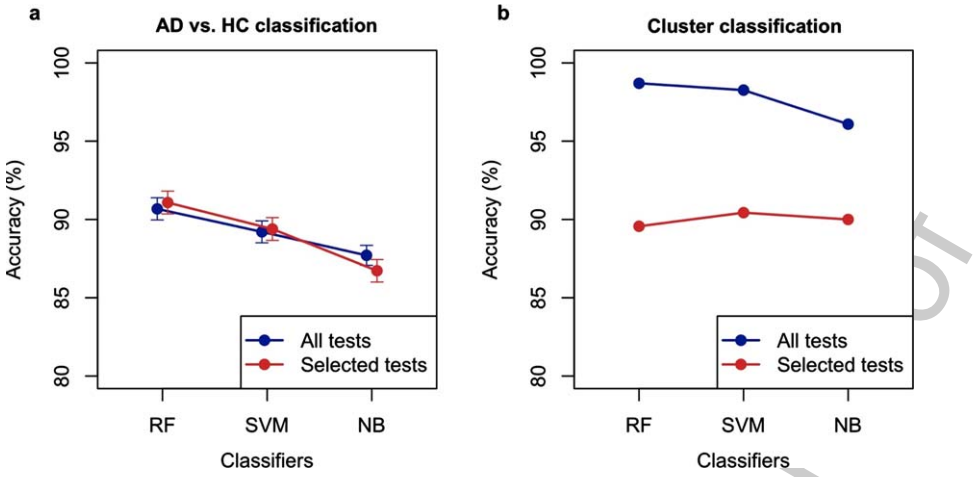

Fig. 5. Classification of AD versus HC and identification of clusters (i.e., sub-phenotypes). a) Accuracy obtained by three machine-learning algorithms in the discrimination between AD and HC (error bars indicate SD computed across 1000 iterations, in each one the classification was performed between N = 230 AD and a random subsample of N = 230 HC from the whole HC sample of N = 326), both using the whole cognitive battery (blue line) and a subset of selected subtests (red line). This subset was selected by means of a recursive stepwise procedure across 1000 iterations (see Methods). Error bars refer to SD calculated on accuracy values obtained across 1000 iterations. b) Accuracy obtained by three machine-learning algorithms in the classification of the four cognitive sub-phenotypes emerged from FMA. The classification was performed using the selected tests (red line). The blue line indicates the reference (overfitted) classification using the whole battery of tests. Despite the diminished accuracy using the subset of tests, all algorithms still showed a good classification performance, suggesting that the four cluster (i.e., sub-phenotypes) could be identified also by means of a few tests. RF, Random Forest; SVM, Support Vector Machine; NB, Naïve Bayes.

patients were labelled as Nonamnestic AD (Cluster 4) since they showed more deficits in language and praxis abilities, and relatively spared memory. Patients in the latter cluster also showed higher MMSE score and a relatively younger age (despite age difference was not significant) compared to other clusters. The contrast between profiles characterized by memory versus non-memory deficits is consistent with recent studies [18] and confirms memory involvement as one of the main dimensions explaining interindividual cognitive variability in AD. Moreover, the Visuospatial AD is consistent with recent findings [39] suggesting that such profile may be explained by a predominant right temporoparietal pattern of brain atrophy and hypoperfusion [16]. Our findings are also consistent with a previous study [40] on heterogeneity in patterns of global cognitive measures (i.e., MMSE and Dementia Rating Scale) employing Latent Class Analysis (LCA). The application of FMA in our work would allow to overcome some of the limits of LCA; moreover, we faced heterogeneity of cognition in AD across many domains, thus providing a characterization of clusters' cognitive profile.

To date, the literature on cognitive heterogeneity in AD has led to spurious results, with some studies agreeing on the existence of four clusters [16], while others suggesting different solutions [17, 18]. This weak consensus might be explained by a lack of ground-truth, e.g., out-of-sample validation of findings or relation between cognitive profiles and known neuroanatomical patterns.

A recent study on patients with mild to moderate AD reliably identified typical and atypical cognitive profiles describing 79.6% and 20% of patients, respectively [41]. Here, applying a similar approach we found similar results, but with a more fine-grained description of patients with atypical cognitive profile. For instance, we highlighted patients with characteristic visuospatial deficits which were not identified by Qiu et al.'s study since visuospatial measures were unavailable in the sample used for clusters' identification. Furthermore, the results of the present study showed a substantial convergence with findings on neuroanatomical heterogeneity in AD and MCI patients (for a review, see [42]). For instance, a recent work by Dong and colleagues [43] analyzed MRI data of 314 AD and 530 MCI patients, and identified in both samples a four-dimensional categorization of neuroanatomical alterations, mainly characterized by: 1) a largely normal anatomy; 2) classical AD-like neuroanatomical pattern; 3) diffuse pattern of atrophy mainly involving parietal and dorsolateral regions with relatively spared medial temporal lobe (MTL);

and 4) predominant involvement of MTL. Other studies on AD and prodromal AD patients have found neuroanatomical subtypes mainly characterized by right temporoparietal [39] or parieto-occipital [44] atrophy, clinically related to visuospatial difficulties.

Aside from the contribution of the present study to the controversial literature on cognitive subtypes in AD, the main take home message of this work is that AD cognitive heterogeneity should be taken into account in the clinical routine. Many protocols of cognitive interventions on AD patients have proven their efficacy at the group-level [45]. However, one of the main goals of neurocognitive assessment is to highlight cognitive strengths and weaknesses at the individual level, and design tailored cognitive trainings accordingly, with a positive impact on patients' global functioning and quality of life. Moreover, some authors [46] have suggested that phase II pharmacological trials would benefit from taking into account finer neurocognitive descriptions of AD patients, since these features may dramatically change drug effect [3]. The identification of individual AD cognitive sub-phenotypes could to some extent improve accuracy and precision in the estimation of prognosis, with different clinical phenotypes being plausibly related to different neurobiological patterns [43]. Future investigations should also shed light on heterogeneity in early-onset AD patients, since pure AD pathology is more frequent in this population and comorbidities are more rarely present [18]. Taken together, the present findings highlight the necessity of further investigating the complex association between cognitive profiles and relative neurobiological features, and potentially lead to the development of finer-grained behavioral biomarkers of disease and disease progression, in a precision medicine perspective [13]. The approach adopted in this study is pivotal in clinical contexts, as it sheds light on the possibility to provide clinicians of quick toolbox, able to identify individual cognitive sub-phenotypes. This was the main reason behind the second aim of our paper, i.e., to highlight the minimum set of cognitive tasks able to accurately identify AD patients, as well as their cognitive sub-phenotype. First, we identified the best set of tests for the discrimination between AD and HC by means of a recursive stepwise procedure (Digit span, Delayed prose memory, TMT-A, Verbal fluency, Abstract thinking, Overlapping figures, Spontaneous drawing, Clock drawing test, and Praxis abilities). This reduced cognitive battery allowed to identify AD patients and their cognitive sub-phenotypes (i.e., clusters) with an accuracy of about 87%. This implies that a few critical cognitive tests can replace the administration of a full neuropsychological battery, not only for a diagnostic purpose, but also for a fine-grained description of AD cognitive profile. Indeed, our results demonstrated that a subset of tests performed as the whole cognitive battery, both in the discrimination of AD versus HC and in the identification of AD cognitive sub-phenotypes. A quicker assessment is more suitable for clinical practice since clinicians are required to evaluate patient's cognitive profile in short time-windows. Moreover, the probability to measure mental fatigue instead of proper cognitive deficits is reduced when less tests are employed.

A main limitation of the present study is the lack of biomarkers of AD and postmortem confirmation about pathology. Despite in principle we cannot rule out the possibility of misdiagnosis (which would affect cognitive heterogeneity), diagnoses were made by expert physicians through careful application of standard clinical criteria, thus we do not believe that our results were driven by potential misdiagnoses. However, future studies will rise from the present findings through the recruitment of a prospective sample of patients diagnosed with AD also by means of standard biomarkers.

A further limitation is the absence of an independent sample to test the generalization of our findings. In future studies we will investigate the relation between cognitive sub-phenotypes and brain anatomical/functional patterns.

As a final remark, the use of data-driven models to study behavioral heterogeneity have some limitations, e.g., the possibility that results are not always generalizable beyond the data they are trained on [47]. In the present work, given our aims and taking into account the limited retrospective sample size, we decided to employ a data-driven method (i.e., FMA) to foster the interpretability of results. However, other valuable methods could be adopted, such as computational models (e.g., [2, 48]). We hope that the present study could be a starting point for the generation of hypotheses that could be tested in future studies applying computational models to larger datasets. We believe that the combination of data-driven and theory-driven approaches could boost the study of clinical heterogeneity in AD and in other diseases.

In conclusion, our sample of AD patients was best described by four cognitive sub-phenotypes which could be detected even by means of a few tests, making this investigation suitable for clinical practice.

The mapping of AD cognitive heterogeneity is important for two main reasons. First, it allows a more fine-grained description of the individual disease, which is desirable in a precision medicine framework. Second, it improves our understanding of AD pathology, by characterizing which features contribute more to the interindividual variability in the clinical manifestation of the disease.

DISCLOSURE STATEMENT

Authors' disclosures available online (https://www.j-alz.com/manuscript-disclosures/21-0719r1).

SUPPLEMENTARY MATERIAL

The supplementary material is available in the electronic version of this article: https://dx.doi.org/10.3233/JAD-210719.

REFERENCES

[1] Lam B, Masellis M, Freedman M, Stuss DT, Black SE (2013) Clinical, imaging, and pathological heterogeneity of the Alzheimer's disease syndrome. *Alzheimers Res Ther* **5**, 1.

[2] Mukherjee S, Mez J, Trittschuh EH, Saykin AJ, Gibbons LE, Fardo DW, Wessels M, Bauman J, Moore M, Choi S-E, Gross AL, Rich J, Louden DKN, Sanders RE, Grabowski TJ, Bird TD, McCurry SM, Snitz BE, Kamboh MI, Lopez OL, De Jager PL, Bennett DA, Keene CD, Larson EB, Crane PK (2020) Genetic data and cognitively defined late-onset Alzheimer's disease subgroups. *Mol Psychiatry* **25**, 2942-2951.

[3] Hampel H, Vergallo A, Aguilar LF, Benda N, Broich K, Cuello AC, Cummings J, Dubois B, Federoff HJ, Fiandaca M, Genthon R, Haberkamp M, Karran E, Mapstone M, Perry G, Schneider LS, Welikovitch LA, Woodcock J, Baldacci F, Lista S (2018) Precision pharmacology for Alzheimer's disease. *Pharmacol Res* **130**, 331-365.

[4] Keine D, Walker JQ, Kennedy BK, Sabbagh MN (2018) Development, application, and results from a precision-medicine platform that personalizes multi-modal treatment plans for mild Alzheimer's disease and at-risk individuals. *Curr Aging Sci* **11**, 173-181.

[5] Lewandowski KE, Sperry SH, Cohen BM, Öngür D (2014) Cognitive variability in psychotic disorders: A cross-diagnostic cluster analysis. *Psychol Med* **44**, 3239-3248.

[6] Picardi A, Viroli C, Tarsitani L, Miglio R, de Girolamo G, Dell'Acqua G, Biondi M (2012) Heterogeneity and symptom structure of schizophrenia. *Psychiatry Res* **198**, 386-394.

[7] Corbetta M, Ramsey L, Callejas A, Baldassarre A, Hacker CD, Siegel JS, Astafiev S V, Rengachary J, Zinn K, Lang CE, Connor LT, Fucetola R, Strube M, Carter AR, Shulman GL (2015) Common behavioral clusters and subcortical anatomy in stroke. *Neuron* **85**, 927-941.

[8] Ramsey LE, Siegel JS, Lang CE, Strube M, Shulman GL, Corbetta M (2017) Behavioural clusters and predictors of performance during recovery from stroke. *Nat Hum Behav* **1**, 38.

[9] Szeto JY, O'Callaghan C, Shine JM, Walton CC, Mowszowski L, Naismith SL, Halliday GM, Lewis SJ (2015) The relationships between mild cognitive impairment and phenotype in Parkinson's disease. *NPJ Park Dis* **1**, 15015.

[10] Disanto G, Berlanga AJ, Handel AE, Para AE, Burrell AM, Fries A, Handunnetthi L, De Luca GC, Morahan JM (2011) Heterogeneity in multiple sclerosis: Scratching the surface of a complex disease. *Autoimmune Dis* **2011**, 932351.

[11] Eavani H, Habes M, Satterthwaite TD, An Y, Hsieh M-K, Honnorat N, Erus G, Doshi J, Ferrucci L, Beason-Held LL, Resnick SM, Davatzikos C (2018) Heterogeneity of structural and functional imaging patterns of advanced brain aging revealed via machine learning methods. *Neurobiol Aging* **71**, 41-50.

[12] Ylikoski R, Ylikoski A, Keskivaara P, Tilvis R, Sulkava R, Erkinjuntti T (1999) Heterogeneity of congnitive profiles in aging: Successful aging, normal aging, and individuals at risks for cognitive decline. *Eur J Neurol* **6**, 645-652.

[13] Au R, Piers RJ, Lancashire L (2015) Back to the future: Alzheimer's disease heterogeneity revisited. *Alzheimers Dement (Amst)* **1**, 368-370.

[14] Hampel H, O'Bryant SE, Durrleman S, Younesi E, Rojkova K, Escott-Price V, Corvol J-C, Broich K, Dubois B, Lista S, Alzheimer Precision Medicine Initiative (2017) A Precision Medicine Initiative for Alzheimer's disease: The road ahead to biomarker-guided integrative disease modeling. *Climacteric* **20**, 107-118.

[15] Crane PK, Trittschuh E, Mukherjee S, Saykin AJ, Sanders RE, Larson EB, McCurry SM, McCormick W, Bowen JD, Grabowski T, Moore M, Bauman J, Gross AL, Keene CD, Bird TD, Gibbons LE, Mez J (2017) Incidence of cognitively defined late-onset Alzheimer's dementia subgroups from a prospective cohort study. *Alzheimers Dement* **13**, 1307-1316.

[16] Cappa A, Ciccarelli N, Baldonero E, Martelli M, Silveri MC (2014) Posterior AD-type pathology: Cognitive subtypes emerging from a cluster analysis. *Behav Neurol* **2014**, 259358.

[17] Scheltens NME, Galindo-Garre F, Pijnenburg YAL, van der Vlies AE, Smits LL, Koene T, Teunissen CE, Barkhof F, Wattjes MP, Scheltens P, van der Flier WM (2016) The identification of cognitive subtypes in Alzheimer's disease dementia using latent class analysis. *J Neurol Neurosurg Psychiatry* **87**, 235-243.

[18] Scheltens NME, Tijms BM, Koene T, Barkhof F, Teunissen CE, Wolfsgruber S, Wagner M, Kornhuber J, Peters O, Cohn-Sheehy BI, Rabinovici GD, Miller BL, Kramer JH, Scheltens P, van der Flier WM (2017) Cognitive subtypes of probable Alzheimer's disease robustly identified in four cohorts. *Alzheimers Dement* **13**, 1226-1236.

[19] Lubke GH, Muthén B (2005) Investigating population heterogeneity with factor mixture models. *Psychol Methods* **10**, 21-39.

[20] Montanari A, Viroli C (2010) Heteroscedastic factor mixture analysis. *Stat Model An Int J* **10**, 441-460.

[21] Viroli C (2012) Using factor mixture analysis to model heterogeneity, cognitive structure, and determinants of dementia: An application to the Aging, Demographics, and Memory Study. *Stat Med* **31**, 2110-2122.

[22] Clark SL, Muthén B, Kaprio J, D'Onofrio BM, Viken R, Rose RJ (2013) Models and strategies for factor mixture analysis: An example concerning the structure underlying

psychological disorders. *Struct Equ Model A Multidiscip J* **20**, 681-703.

[23] Chapman RM, Mapstone M, Porsteinsson AP, Gardner MN, McCrary JW, DeGrush E, Reilly LA, Sandoval TC, Guillily MD (2010) Diagnosis of Alzheimer's disease using neuropsychological testing improved by multivariate analyses. *J Clin Exp Neuropsychol* **32**, 793-808.

[24] McKhann GM, Knopman DS, Chertkow H, Hyman BT, Jack CR, Kawas CH, Klunk WE, Koroshetz WJ, Manly JJ, Mayeux R, Mohs RC, Morris JC, Rossor MN, Scheltens P, Carrillo MC, Thies B, Weintraub S, Phelps CH (2011) 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* **7**, 263-269.

[25] Folstein MF, Folstein SE, McHugh PR (1975) "Mini-mental state". A practical method for grading the cognitive state of patients for the clinician. *J Psychiatr Res* **12**, 189-198.

[26] Mondini S, Mapelli D, Vestri A, Arcara G, Bisiacchi PS (2011) *Esame neuropsicologico breve 2: Una batteria di test per lo screening neuropsicologico*.

[27] Dubois B, Villain N, Frisoni GB, Rabinovici GD, Sabbagh M, Cappa S, Bejanin A, Bombois S, Epelbaum S, Teichmann M, Habert M-O, Nordberg A, Blennow K, Galasko D, Stern Y, Rowe CC, Salloway S, Schneider LS, Cummings JL, Feldman HH (2021) Clinical diagnosis of Alzheimer's disease: Recommendations of the International Working Group. *Lancet Neurol* **20**, 484-496.

[28] Cagnone S, Viroli C (2012) A factor mixture analysis model for multivariate binary data. *Stat Model An Int J* **12**, 257-277.

[29] Frazier TW, Youngstrom EA, Sinclair L, Kubu CS, Law P, Rezai A, Constantino JN, Eng C (2010) Autism spectrum disorders as a qualitatively distinct category from typical behavior in a large, clinically ascertained sample. *Assessment* **17**, 308-320.

[30] Georgiades S, Szatmari P, Boyle M, Hanna S, Duku E, Zwaigenbaum L, Bryson S, Fombonne E, Volden J, Mirenda P, Smith I, Roberts W, Vaillancourt T, Waddell C, Bennett T, Thompson A (2013) Investigating phenotypic heterogeneity in children with autism spectrum disorder: A factor mixture modeling approach. *J Child Psychol Psychiatry* **54**, 206-215.

[31] Lubke GH, Miller PJ (2015) Does nature have joints worth carving? A discussion of taxometrics, model-based clustering and latent variable mixture modeling. *Psychol Med* **45**, 705-715.

[32] Schwarz GE (1978) Estimating the dimension of a model. *Ann Stat* **6**, 461-464.

[33] Akaike H (1987) Factor analysis and AIC. *Psychometrika* **52**, 317-332.

[34] R Core Team (2019) R: A language and environment for statistical computing.

[35] Viroli C (2012) FactMixtAnalysis: Factor Mixture Analysis with covariates. R package version 1.0.

[36] Hornik K, Buchta C, Zeileis A (2009) Open-source machine learning: R meets Weka. *Comput Stat* **24**, 225-232.

[37] Wagenmakers E-J, Farrell S (2004) AIC model selection using Akaike weights. *Psychon Bull Rev* **11**, 192-196.

[38] Ferreira D, Nordberg A, Westman E (2020) Biological subtypes of Alzheimer disease. *Neurology* **94**, 436-448.

[39] Kim KW, Park S, Jo H, Cho SH, Kim SJ, Kim Y, Jang H, Na DL, Seo SW, Kim HJ (2020) Identifying a subtype of Alzheimer's disease characterised by predominant right focal cortical atrophy. *Sci Rep* **10**, 7256.

[40] Davidson JE, Irizarry MC, Bray BC, Wetten S, Galwey N, Gibson R, Borrie M, Delisle R, Faldman HH, Hsiung G-Y, Fornazzari L, Gauthier S, Guzman D, Loy-English I, Keren R, Kertesz A, Geirge-Hyslop P St., Wherrett J, Monsch AU (2010) An exploration of cognitive subgroups in Alzheimer's disease. *J Int Neuropsychol Soc* **16**, 233.

[41] Qiu Y, Jacobs DM, Messer K, Salmon DP, Feldman HH (2019) Cognitive heterogeneity in probable Alzheimer disease. *Neurology* **93**, e778-e790.

[42] Habes M, Grothe MJ, Tunc B, McMillan C, Wolk DA, Davatzikos C (2020) Disentangling heterogeneity in Alzheimer's disease and related dementias using data-driven methods. *Biol Psychiatry* **88**, 70-82.

[43] Dong A, Toledo JB, Honnorat N, Doshi J, Varol E, Sotiras A, Wolk D, Trojanowski JQ, Davatzikos C (2017) Heterogeneity of neuroanatomical patterns in prodromal Alzheimer's disease: Links to cognition, progression and biomarkers. *Brain* **140**, 735-747.

[44] ten Kate M, Dicks E, Visser PJ, van der Flier WM, Teunissen CE, Barkhof F, Scheltens P, Tijms BM (2018) Atrophy subtypes in prodromal Alzheimer's disease are associated with cognitive decline. *Brain* **141**, 3443-3456.

[45] Bahar-Fuchs A, Clare L, Woods B (2013) Cognitive training and cognitive rehabilitation for persons with mild to moderate dementia of the Alzheimer's or vascular type: A review. *Alzheimers Res Ther* **5**, 35.

[46] Devi G, Scheltens P (2018) Heterogeneity of Alzheimer's disease: Consequence for drug trials? *Alzheimers Res Ther* **10**, 122.

[47] Birkenbihl C, Salimi Y, Fröhlich H (2021) Unraveling the heterogeneity in Alzheimer's disease progression across multiple cohorts and the implications for data-driven disease modeling. *Alzheimers Dement*, doi:10.1002/alz.12387.

[48] Caligiore D, Silvetti M, D'Amelio M, Puglisi-Allegra S, Baldassarre G (2020) Computational modeling of catecholamines dysfunction in Alzheimer's disease at preplaque stage. *J Alzheimers Dis* **77**, 275-290.