Identification of Heterogeneous Cognitive Subgroups in Community-Dwelling Older Adults: A Latent Class Analysis of the Einstein Aging Study

Andrea R. Zammit[1,2], **Charles B. Hall**[1,2,3], **Richard B. Lipton**[1,2,3], **Mindy J. Katz**[1,2], and **Graciela Muniz-Terrera**[4]

[1]Saul B. Korey Department of Neurology, Albert Einstein College of Medicine, Bronx, New York [2]Einstein Aging Study, Albert Einstein College of Medicine, Bronx, New York [3]Department of Epidemiology and Population Health, Albert Einstein College of Medicine, Bronx, New York [4]The University of Edinburgh, Scotland

Abstract

Objectives: The aim of this study was to identify natural subgroups of older adults based on cognitive performance, and to establish each subgroup's characteristics based on demographic factors, physical function, psychosocial well-being, and comorbidity.

Methods: We applied latent class (LC) modeling to identify subgroups in baseline assessments of 1345 Einstein Aging Study (EAS) participants free of dementia. The EAS is a community-dwelling cohort study of 70 + years-old adults living in the Bronx, NY. We used 10 neurocognitive tests and 3 covariates (age, sex, education) to identify latent subgroups. We used goodness-of-fit statistics to identify the optimal class solution and assess model adequacy. We also validated our model using two-fold split-half cross-validation.

Results: The sample had a mean age of 78.0 (*SD*=5.4) and a mean of 13.6 years of education (*SD*=3.5). A 9-class solution based on cognitive performance at baseline was the best-fitting model. We characterized the 9 identified classes as (i) disadvantaged, (ii) poor language, (iii) poor episodic memory and fluency, (iv) poor processing speed and executive function, (v) low average, (vi) high average, (vii) average, (viii) poor executive and poor working memory, (ix) elite. The cross validation indicated stable class assignment with the exception of the average and high average classes.

Conclusions: LC modeling in a community sample of older adults revealed 9 cognitive subgroups. Assignment of subgroups was reliable and associated with external validators. Future work will test the predictive validity of these groups for outcomes such as Alzheimer's disease,

Correspondence and reprint requests to: Andrea R. Zammit, Saul B. Korey, Department of Neurology, Albert Einstein College of Medicine, 1225 Morris Park Avenue, Van Etten Building, Room 3C9A, Bronx, NY 10461. andrea.zammit@einstein.yu.edu.

Conflicts of Interest: All authors declare that there are no financial, personal, or other potential conflicts of interest.

vascular dementia and death, as well as markers of biological pathways that contribute to cognitive decline.

Keywords

Cognitive aging; Cognitive heterogeneity; Cognitive subtypes; Individual differences

INTRODUCTION

Cognitive impairment is a central characteristic of Alzheimer's disease (AD) and other forms of dementia. Cognitive impairment in individuals who go on to develop AD takes place long before dementia diagnosis (Elias et al., 2000; Hall, Lipton, Sliwinski, & Stewart, 2000; La Rue & Jarvik, 1987; Masur, Sliwinski, Lipton, Blau, & Crystal, 1994; Snowdon et al., 1996; Whalley et al., 2000). Identifying and characterizing cognitive subgroups in the pre-dementia phase is essential to the development of prevention and treatment strategies. Multiple factors contribute to a high level of between-person variation in cognitive subgroups contributing to the difficulties in characterizing subgroups of individuals at risk for future cognitive decline.

The parsing of cognitive heterogeneity in aging populations is a fundamental step in the identification of subgroups at risk for AD, vascular dementia (VaD), and other forms of dementia. Statistical methods can be used to identify distinct subgroups of cognitively more homogeneous older adults. Although several studies have applied statistical clustering methods based on neuropsychological data, most have been limited by the range of cognitive variability: studies have focused on mild cognitive impairment (Delano-Wood et al., 2009; Hanfelt et al., 2011; Libon et al., 2010), individuals with dementia (Davidson et al., 2010; Peter et al., 2014), or unimpaired individuals (Costa, Santos, Cunha, Palha, & Sousa, 2013; Zammit, Starr, Johnson, & Deary, 2014). In other studies, the neuropsychological tests are limited, the sample sizes are modest (n 120) (Libon et al., 2010; Ylikoski et al., 1999), or the statistical methods are suboptimal (Libon et al., 2010; Ylikoski et al., 1999).

Latent class analysis (LCA) is a group-based approach that uses general mixture modeling techniques to provide an empirical means of summarizing large amounts of data to study individual differences using a taxonomic dimension (McLachlan & Peel, 2000). This individualized multilevel approach allows each individual to be placed on a scale of prognostic probability (Nagin & Odgers, 2010). This approach may allow us to identify groups of participants at increased risk of different types of dementia. Furthermore, by defining cognitive subgroups and their phenotypic risk profile (e.g., comorbidity, depressive symptomatology, etc.), we may eventually be in a better position to tailor specific interventions. This approach will allow us to define the level of risk individuals are in by placing them on a spectrum from negligible to high risk.

The aims of this study were (i) to establish naturally occurring classes of cognitive function in community-dwelling older adults at cross-section; and (ii) characterize these classes using demographic features, physical function, psychosocial well-being, and comorbidity.

Several studies have provided useful information on the various profiles that exist in different populations in old age in an attempt to find out how the latent structure of cognition unfolds across domains and across individuals (Costa et al., 2013; Maxson, Berg, & McClearn, 1997; Ko, Berg, Butner, Uchino, & Smith, 2007; Morack, Ram, Fauth, & Gerstorf, 2013; Smith & Baltes, 1997; Zammit et al., 2014); however, they are purely descriptive, and their taxonomies capture just one model of reality. As suggested (Nagin & Odgers, 2010) we adopted a bottom-up approach within an exploratory and confirmatory framework and imposed no constraints on the number or make-up of the classes being generated.

METHODS

Our methodological approach can be described in four steps.

1. Selecting the Study Population

We used the Einstein Aging Study (EAS) cohort for our analyses; a detailed description of the study design has been previously reported (Katz et al., 2012). Participants are 70 years and older, community-dwelling, English-speaking and reside in the Bronx, New York. Participants were systematically recruited from the Health Care Financing Administration/ Centers for Medicaid and Medicare Services rosters for Medicare-eligible between 1993 and 2004, and from New York City Board of Elections voter registration lists from 2004 onward. Written informed consent was obtained on their first clinical visit. The study protocol was approved by the local institutional review board.

Between 1993 and 2015, a total of 2262 participants had baseline evaluations, of those, 1395 had follow-up data. There were 50 participants with prevalent dementia who were excluded from the study. Participants were assigned a diagnosis of dementia at consensus case conferences, which included comprehensive review of cognitive test results, relevant neurological signs and symptoms, and functional status based on standardized clinical criteria from the Diagnostic and Statistical Manual, Fourth Edition (Diagnostic and Statistical Manual of Mental Disorders, DSM-IV, 1994); this required impairment in memory and at least one additional cognitive domain, as well as evidence of functional decline. We proceeded our analyses on the remaining 1345 participants.

Core Variables Applied to the Latent Class Analysis

Neurocognitive measures.: All EAS participants were administered the Free and Cued Selective Reminding Test (FCSRT) free recall test (Buschke, 1973, 1984; Grober et al., 1999); Logical Memory (LM) (Wechsler, 1987), Categories (CAT) (Benton & Hamsher, 1989), the Controlled Oral Word Fluency Test (FAS) (Monsch et al., 1992), Trail Making Tests A (TMTA) and B (TMTB) (Army Individual Test Battery, 1944), Digit Symbol Coding (Wechsler, 1997), Digit Span (Wechsler, 1997), Block Design (Wechsler, 1997), and the Boston Naming Test (BNT) (Kaplan, Goodglass, & Weintraub, 1983) during the in-person evaluation on the participant's clinic visit day. Baseline scores from these neuropsychological instruments were used to identify cognitive subgroups.

Demographic measures.: We included self-reported age, gender, and number of years of formal education as covariate predictors of class membership in the latent class model. The inclusion of at least one strong covariate positively affects model estimation and improves the fit of the class solution (Wurpts & Geiser, 2014).

2. Fitting the LCA model and Selecting the Optimal Class

We fitted models with an increasing number of classes from 2 to 10. We then applied recommended procedures to identify the most suitable model; these included comparison of fit indices, evaluation of class separation, and class interpretation. We specified 2-, 3-, 4-, 5-, 6-, 7-, 8-, 9-, and 10-class solutions. To assess fit between models with k *versus* k+1 classes, we used the Bayesian information criterion (BIC) (Raftery, 1995); we also used entropy (Celeux & Soromenho, 1996) to identify the solution that seemed to have the best precision in distinguishing amongst classes. To obtain appropriate model convergence and a robust solution we applied 500 random starts in the initial stage and 10 optimizations in the final stage.

Further recommendations in assessing model adequacy (Nagin & Odgers, 2010) included, obtaining a close correspondence between the estimated probability of group membership and the proportion assigned to that group based on the posterior probability of group membership for each class, and ensuring that the average of the posterior probabilities of group membership for individuals assigned to each group exceeded a minimum threshold of 0.7.

3. Cross-Replication Procedures

For replication and validation purposes, we applied two-fold cross-validation split-half procedures as described in Collins et al.(Collins, Graham, Long, & Hansen, 1994), Collins and Lanza (Collins & Lanza, 2010), and Grimm et al. (Grimm, Mazza, & Davoudzadeh, 2017). Specifically, we tested our model by applying two-fold cross-validation techniques by defining two random independent sub-samples, fitting models in each sample and then applying them to the other sample. In other words, each sample was in turn both used as a calibration (known class) sample and a cross-validation (training) sample. We first developed models ranging between 2-and 10-class solutions in our sample of 1345 individuals. For the cross-validation we randomly split the sample into two subsamples (*n*=672 and 673) and we applied these models onto each half. We then used each subsample to predict membership on the other half, hence calibrating the model on one subsample and cross-validating it on the other using training procedures in MPlus.

We then examined the cross-tabulated classes to determine the Kappa, and examine how closely the inter-class probabilities matched between participants who were subjected to the LCA under trained and untrained conditions. Thus, based on subsample 2's performance and subsample 1's class formation, we tested whether the model correctly assigned the participants from subsample 2 to similar classes of the known participants from subsample 1, and *vice-versa.* We ran the models for training using neuropsychological tests, age, gender, and education as predictor variables.

4. Characterization and Validation Procedures

To characterize and validate our model, we analyzed preexisting characteristics to determine if variables unrelated to the class-formation of the participants are distinguishable amongst the classes. We applied two sets of validators; those that were used to form the groups, that is, the core neurocognitive measures, and those that were only introduced to the classes after the model solution was finalized, that is, the external non-cognitive validators.

Core neurocognitive characteristics—A good first indication of whether the classes are substantially different from each other are differences among classes on the variables that were used to define the classes (Vermunt & Magidson, 2002). In our study, we had 10 neurocognitive variables and 3 covariates (described above).

External non-cognitive validators—External validators are variables that have not been used in the class formation to determine if the classes are distinguishable on pre-existing characteristics and help in interpreting the results (Nagin & Tremblay, 2005; Nylund, Asparouhov, & Muthen, 2007). Table 1 lists in detail a comprehensive account of the external variables that were used to validate the final model. We grouped these variables in four categories of (i) demographics (ethnicity and the Wide Range Achievement Test), (ii) physical function (gait speed, grip strength, peak flow), (iii) psychological wellbeing (depression, stress, neuroticism), and (iv) health and disease (smoking, hypertension, diabetes, history of stroke).

We used multivariate analyses of variance for continuous variables, and Chi-square tests for categorical variables with group membership as the independent variables to learn how the classes differed from each other.

Statistical Software—For LCA modeling, we used MPlus version 7 (Muthen & Muthen, 1998–2016). Figures were generated in RStudio for Mac version 1.0.136 (Integrated Development for R. RStudio, Inc., Boston, MA, http://www.rstudio.com); the "ggplot2" package was used to generate the boxplots. All other analyses were done in SPSS version 24 (SPSS Inc., Released 2016).

RESULTS

Characteristics of the Study Population

Table 2 shows the characteristics of the whole sample, including subsamples 1 (n=672) and 2 (n=673). The average age of the cohort was 78.0 years (SD = 5.4) and 61.6% (n = 828) were female. Average years of education were 13.6 (SD = 3.5) and 68% (n=914) were non-Hispanic White. As can be seen in Table 2, the characteristics of both subsamples are similar to each other and the whole cohort.

Formation and Selection of the Optimal Class Solution

Model fit statistics for 2–10 class models are presented in Supplementary Table 1. Although the 3-class solution had the highest entropy (0.847), the fit kept improving markedly with every added class, up until the 9-class solution (BIC=90957.213),with an entropy estimating

that over 82% of research participants had a distinctive class assigned to them. With the 10-class solution, the fit did not improve much further, and the entropy started to decline (0.799), so we stopped fitting classes at that point. Models with 3, 4, and 5 solutions showed quantitative differences amongst the groups, and it was only when we started adding 6 + classes that qualitative differences started to emerge. When using continuous data, there is usually a trade-off between parsimony and fit, and typically at least two models will emerge as close approximation to the *true* model with some ambiguity as to which is the *best* model. It is up to the researcher to select the *best* model, that is, the closest approximation to the unknown *true* model (Collins et al., 1994).

Reliability and Adequacy of the Class Solution

We compared classes between the new-formed solutions in subsamples 1 and 2 and the cross-validated models on each split-half. Both demographic characteristics and cognitive subgroups were similar between the classes of subsamples 1 and 2 (results available on request). The cross-validated models showed that the model fit improved with the 5-class solution, and then deteriorated before improving again with the 9-class solution, after which it started declining again with the 10-class solution. The model that yielded the most favorable fit statistics depended upon whether subsample 1 or subsample 2 served as the cross-validation sample. Had there been no cross-validation, or had single cross-validation been used, the conclusions drawn from these results might have been inconclusive.

For example, without cross-validation, the BIC kept improving with every added class, leaving no clear identification of a solution (Supplementary Table 1); however, as can be seen in Supplementary Table 2, the BIC improved for the 5-class solution in both cross-validated subsamples 1 (BIC=91409.145) and 2 (BIC=91221.606), and then deteriorated up until it reached a better fit for the 9-class solution in both cross-validated subsamples 1 (BIC = 91229.389) and 2 (BIC = 9122.651). Entropy was   .9 for the cross-validated 9-class solutions and > .8 and .9 for the 5-class solutions. When mapped onto the trained solutions, participants generally fell into similar classes (Supplementary Table 3). The major misclassifications across classes were between classes 6 and 7 in both subsamples, which were assignments to similar classes (average and high average cognition). Cohen's kappa coefficient for the inter-class agreement was moderate to good at 0.640 and 0.724, for subsamples 1 and 2.

We, therefore, investigated model adequacy on the 9-class solution on the whole sample, which showed that the posterior probabilities of group membership exceeded the minimum threshold of 0.7, ranging from a minimum of 75.8% in class 7 to 99.1% in Class 8 Each class also obtained a close correspondence (probability > 0.70) between the estimated probability of group membership and the proportion assigned to that group, with probabilities ranging between 0.70 for Class 7 to 0.90 for Class 6 We further investigated the effect sizes of each of the cognitive measures on the latent classes by using the following formula, $\beta_{stdYX} = \beta*SD(X)/SD(Y)$ to calculate the standardized covariates These ranged from 1 4 for Block Design in Latent Class 8 to 9.3 for Trail Making Test A in Latent Class 8 Further details can be found in Supplementary Table 4.

These results provided evidence that our model was robust, and we proceeded to investigate further the 9-class solution.

Characteristics and Validity of the 9-Class Solution

Table 3 shows the demographic characteristics of each of the nine classes. In this table, we also included external demographic variables, such as ethnicity and the WRAT (as a proxy for premorbid IQ).

In Table 4, test performance on the neurocognitive battery that was used to form the classes is presented according to class membership.

External validators of the class-solution including indicators of physical fitness, psychological wellbeing, and health and other medical co-morbidities can be found in Table 5. As the Tables show, classes differed significantly amongst each other on all variables except for grip strength, current smokers, and hypertension (Distress in the PSS-14 almost reached significance) indicating distinguishable pre-existing characteristics that the model correctly identified.

Summary of the Classes

Figures 1 and 2 illustrate in box plots how each of the classes performed in relation to each other according to working (Digit Span) and episodic memory (FCSRT, LM) and language (BNT) and fluency (word fluency, Categories), and processing speed (TMTA, Digit Symbol Coding) and executive function (TMTB, Block Design).

Class 1 consisted of 77 (5.70%) participants. It had the highest percentage of females ($n = 58$; 75.3%). Class 1 seemed to be the overall *disadvantaged class* – participants had the poorest levels of education (mean = 8.4 years; $SD = 2.8$), scored lowest on premorbid IQ (mean=49.1; $SD = 15$), and lowest on Boston Naming, Digit Span, Block Design, and word fluency. It also scored second to lowest on Digit Symbol Coding and Logical Memory. In terms of physical fitness, Class 1 had the poorest peak flow and the slowest gait speed. Individuals in this class also scored highest on the GDS and, along with Class 3, on Neuroticism.

Class 2 consisted of 90 (6.7%) participants with an average age of 81.3 ($SD = 6.3$); this was one of the oldest classes. Strikingly, although scores on its cognitive measures were not as severe as Classes 1, 3, 4 and 8, compared to the well-performing classes (Classes 5, 6, 7, and 9), and even to Class 4, this class performed relatively worse on language and fluency measures. The two main elements that distinguished this class from the rest is that, despite the average scores on cognitive measures, it had poorer levels of education and worse WRAT scores than most classes, even those with poor cognition (except for Classes 1 and 8, which showed patterns of cognitive impairment), indicating possibly a low cognitive reserve threshold. To keep the classes strictly related to cognition, we termed this class *the poor language class* due to the low premorbid ability on the reading test. This class had the highest proportion of current smokers (11.5%), possibly reflecting the low education levels of its participants. Interestingly, this class had the lowest scores on Neuroticism (mean= 17.5; $SD = 4.6$).

Class 3 consisted of 81 (6.0%) participants with an average age of 79.0 (*SD*= 5.0). This class scored poorly on fluency measures; however, it performed the poorest on Free Recall (means= 23.4; *SD* = 5.8) and second to poorest on Logical Memory (Class 1 performed the worst). We termed Class 3 as the *poor episodic memory and fluency class.* Relatedly, this class also had relatively high scores on the GDS (mean= 3.2; *SD*= 2.7) compared to the rest of the classes; and scored the highest on Neuroticism (mean= 28.0; *SD* = 8.7).

Class 4 (*n* =51; 6.1%) was one of the smallest classes, displaying a very specific impairment in timed tasks of processing speed and executive function, namely, Digit Symbol Coding, Block Design, and Trail Making Tests A and B, but did better on tests of memory (FCSRT and LM) and premorbid IQ (NART and education were relatively high). The majority was African American (*n* =19; 51.4%), and 56.8% were female. Due to this domain-specific impairment, especially in light of the relatively good scores in other domains, we hypothesize that perhaps an underlying mechanism affecting speeded tasks and executive functions may be at play, and we thus termed this the *poor processing speed and executive function class.*

Classes 5 and 6 were the largest classes (*n* =346 and *n* =440) making up 25.7% and 32.7% of the population. These classes performed at low average and high average compared to the rest of the sample on the neurocognitive battery. Class 7 was also a relatively large class at 136 (10.1%) participants. This class had a high proportion of non-Hispanic White (*n*= 110; 80.9%) and one of the lowest of females (*n* =39; 28.7%). Classes 5, 6, and 7 all had relatively high levels of education ranging from 12.2 years (*SD* = 2.6) in Class 5 to 15.3 years (*SD* = 3.2) in Class 7. Overall, Classes 5, 6, and 7, outperformed each other across measures, with Class 6 doing better than Classes 5 and 7 overall, and Class 6 faring better than Class 5 on measures of peak flow, grip strength, and gait speed, while Class 7 showed highest peak flow and grip strength average across all classes. This pattern reflected their score on the neurocognitive measures. We labeled these classes as *low average cognition* (Class 5), *average cognition* (Class 7), and *high average cognition* (Class 6).

Class 8 was the smallest (*n*= 34; 2.5%) and oldest class (age = 81.4 years; *SD* = 5.9). It had the highest proportion of African Americans (*n* =19; 55.9%), along with Class 4 (*poor processing speed and executive function class).* This class also scored the poorest on TMTA and TMTB; along with Class 1, this class also had the lowest scores on Digit Symbol Coding, and Digit Span. We termed Class 8 as the *poor executive and poor working memory class* due to the noticeable co-impairment in measures reflecting these domains. This class, along with Class 1 had the slowest gait speed; it also had the highest proportion of individuals with a history of stroke (23.5%).

Class 9 (*n*= 104; 7.7%) was the youngest class (mean age=75.1; *SD*=3.9) and its participants had the highest number of years in education (mean=16.8; *SD*=2.7). This class consisted of 67.3% females and the majority (88.5%) was non-Hispanic White. This was the best performing class on all tests of the neurocognitive battery. In terms of physical fitness, Class 9 outperformed all other classes on gait speed, but not on peak flow and grip strength (Class 7 did better). Participants in class 9 were also generally more emotionally stable than the rest of the sample displaying lowest scores on the GDS, the PSS-14, and Neuroticism. Class 9,

along with Class 6 (high average) had the highest percentage of past smokers (57.7% and 59.7%). We termed this class as the *elite* class.

DISCUSSION

Cognitive performance in older adults is determined by a multiplicity of factors including age and illness-related cognitive decline, including Alzheimer's pathology, vascular disease and Lewy body disease, among others. As these factors have sites of predilection, patterns of cognitive performance may be linked to various patterns of disease-related decline (Gelber, Launer, & White, 2012; Ritchie et al., 2016; White et al., 2016). In this study, we aimed to identify groups of older adults differing in their patterns of cognitive performance, and to link these groups to profiles of physical function, psychological well-being, and comorbidity before the development of dementia.

We fit latent class models to two groups of older adults and applied models developed in one group to the other. Our results showed that the older population is highly diverse with some classes of individuals showing dimensional patterns of cognitive function (e.g., the high average, average, and low average cognition classes), while others cluster at the high end of the spectrum (e.g., the elite class) or at the lower end (e.g., the disadvantaged class), and yet others display discontinuous patterns of scores across measures of cognition (e.g., the poor language, the poor episodic memory and fluency, the poor processing speed and executive function, and the poor executive and poor working memory class).

The reliability of classification in the cross-validation subsamples illustrated that most participants were grouped into similar classes. We were thus interested in determining why some cases had low posterior-probabilities compared to others whose class assignment was more clear-cut. Upon investigating the posterior probabilities of the high-risk classes (i.e., the disadvantaged, the poor language, the poor episodic memory and fluency, the poor processing speed and executive function, and the poor executive and poor working memory classes), we found that the individuals whose posterior probabilities were < .70 were also assigned probabilities that belonged to equally at-risk groups.

For example, participants from the disadvantaged class were more likely to be assigned to the poor processing speed and executive function and the poor executive and poor memory classes if they had not been assigned primarily to their class. One participant, for instance was assigned a posterior probability of .42 for his final assignment in the disadvantaged class, and .34 for the poor processing speed and executive function class, while another was given a posterior probability of .41 to his final class assignment in the disadvantaged class, and .37 for the poor executive and poor working memory class, indicating that possibly these three classes share an underlying pattern, that could indicate low cognitive reserve (all three classes had low education and poor WRAT scores).

Similarly, most individuals who received a posterior probability <.70 for the poor episodic memory and fluency class, were more likely to be assigned to the low average class than any other class, indicating that poor cognition overall may also be a risk factor for AD, especially if this class was not reflective of a life-time trait of poor cognition but a slow

general decline across all measures of cognition. Literature does, in fact, point out that with AD pathology, it is possible that one faulty network is likely to affect the breakdown of other networks responsible for various different cognitive functions (Mortamais et al., 2017), and that a dedifferentiation of the structure of individuals' cognitive abilities occurs as one slowly moves toward a clinical diagnosis (Sliwinski, Hofer, & Hall, 2003) and terminal decline (Hulur, Ram, Willis, Schaie, & Gerstorf, 2015).

Similar trends were found for the poor processing speed and executive function, the poor executive and poor working memory, and the poor language classes. Participants who were assigned posterior probabilities of < .70 for their eventual class assignment to the poor executive and poor working memory class ($n = 3$), were assigned moderately high (ranging from 0.27 to 0.35) posterior probabilities to the poor episodic memory class and one high probability (0.50) to the disadvantaged class. This result is interesting because these participants were more likely to be classed with poor episodic memory and fluency, or with the overall disadvantaged participants, than in the poor processing speed and executive function class.

Furthermore, the pattern of scores of these participants is more severe and reflective of imminent pathology than of simply scoring low on the same tests. Research shows that tests of attention/working memory and executive function have the highest discriminatory power between individuals who convert to AD than those who do not (Rajan, Wilson, Weuve, Barnes, & Evans, 2015; Rapp & Reischies, 2005).

This pattern of tentative assignment is insightful; these posterior probabilities may in fact be telling us something more than what the final class solution presents. Learning that individuals who made the cut for one class but posteriorly finding out about the ambiguities the model had in assigning them, reminds us that LCA is not so much about pigeonholing but more about illustrating the heterogeneity of individuals pertaining, in this case, not only to episodic memory, but also to executive function, language, fluency, processing speed, and working memory. Our results suggest that these domains seem to start showing patterns of impairment even before dementia has been diagnosed, which further adds evidence to existing research trying to move away from the heavy focus usually given solely to episodic memory, while neglecting other domains that seem to be equally predictive of AD (Backman, Jones, Berger, Laukka, & Small, 2005; Mortamais et al., 2017; Ritchie et al., 2017).

Our results also provide information on the latent structure of cognition; research shows that underlying cognitive measures are correlated (Salthouse, 2013), and it is thus sometimes assumed that classifications of cognitive measures will only result in quantitative patterns. Although to an extent this is true, especially in healthy populations, results from this study have also shown that this may not always be the case, and that, latent composites such as executive function and working memory, within the cognitive ability structure, may reveal specific information about particular groups of individuals. Lastly, although our study's objective was not about classifying individuals per se, but about taking a snapshot of the heterogeneity present in an older population, our results suggest that possibly more than one or two trajectory classes are present in the onset of AD and related dementias.

Future Directions

Our results have clinical implications. Fields of cognitive aging and cognitive epidemiology aim to diminish health inequalities and provide tailored support that come with increased awareness of individual differences relating to cognitive and personological traits (Deary & Batty, 2007). We realize that a simpler model may be more attractive for clinical-applicability; however, in reality, cognitive aging is a complex and diverse phenomenon. Furthermore, in the 9-class solution we captured more variance than we would have captured in a simpler solution; for example, the distributions for the 2- to 5-class solutions mainly illustrated quantitative differences amongst the classes (results available on request), and the bigger the sample is, the more heterogeneity is likely to be captured, especially when using continuous variables. Understandably, the identification of a simpler solution with fewer (clinic-friendly) measures that would still explain the same amount of variance as our 9-class model would be more practical.

Since in our cross-validation subsamples, a 5-class solution expressed favorable fit criteria, we plan on evaluating the extent to which this model explains the heterogeneity found in this sample, and if its classes are predictive of distal outcomes for clinical purposes. A simple solution would also raise some questions regarding our model. For example, does the model need two episodic memory tests? And would the addition of, or a different, fluency test be more helpful in stabilizing the groups? Suppose one group is defined by a particular measure, for example, BNT, but that measure is excluded from the model? Would the stability of the groups hold? We plan to address these questions in future work that will be aimed more at the clinical applicability of these solutions and further look into patterns that emerge with different measures, different models, and different cohorts.

At this early stage of investigation, our results suggested that the individuals who would benefit most from disease-modifying interventions are those who fell in the high-risk classes (i.e., the disadvantaged, poor language, poor episodic memory and fluency, poor processing speed and executive function, and poor executive and poor working memory classes); in other words, individuals who displayed a highrisk profile of neurocognitive function, evidence of specific underlying disease processes, presence of MCI symptomatology, or are on the brink of expressing clinical symptoms. In our future work, we will expect that the course of late-life aging can lead to improved parsing of cognitive heterogeneity and early diagnosis.

Discovering and identifying specific sub-populations that inform dementia-risk may become phenotypes for further imaging and genetic testing, which we plan to do by mapping our classes on genetic and imaging data. We also plan to follow up our class solution to find out how individuals moved across classes from one time point to the next, and determine if, and which, classes have more mobility than others due to inherent risks and characteristics. It is only a matter of time when results of complex modeling of presymptomatic AD will be interpretable with more precision for researchers to assess an individual's risk of conversion over a defined period of time, possibly even estimating this before preclinical disease. Consequently, this will open possibilities that will help place individuals on a risk spectrum ranging from negligible to imminent risk.

Strengths—The EAS is a large and rich dataset of a community-based sample consisting of a population diverse across several dimensions, including, ethnic, educational, occupational, and socio-economical background, health-status, and cognitive function. By segmenting a large and diverse dataset into smaller classes, a group-based approach to studying small subsets of homogeneous groups provides a pragmatic approach to summarizing large amounts of information to examine long-standing developmental theories using a taxonomic dimension. LCA modeling for these purposes has several strengths: it does not conform to model assumptions, such as normal distribution or homogeneity; it allows mixed scale variables in the same analysis; it allows the relationship between latent classes and covariates to be assessed simultaneously; and it provides posterior membership probabilities to classify participants in their appropriate class. Finally, LCA classifies individuals using a more sophisticated approach using proportion estimates than other approaches that force equal subgroup sizes or that adopt a quartile approach (Gelman, Carlin, Stern, & Rubin, 2004; Vermunt & Magidson, 2002).

The EAS also has an extensive battery of neuropsychological tests; in this study we used 10 neurocognitive variables to class our participants; although additional measures have added benefit to classification since more heterogeneous patterns are likely to emerge, we realize that other studies may not administer the same measures, may not have the same number or type of variables; they may not even have any variables relating to a particular cognitive domain, for example, language, which are challenges for replication purposes. Although the identification of cognitive composites allows results to be more generalizable (Salthouse, 2010), in future work we plan to harmonize our data with other aging cohorts in an effort to maximize the potential these results have, and to avoid the many limitations replication studies have (for a review see Hofer & Piccinin, 2010).

Limitations—This is a cohort of older adults; thus, the information that we have captured at this time-point may have been influenced by factors leading up to this time; datasets that enroll participants at a younger age may be able to identify early development of diseases, before the brain is significantly affected and clinical symptoms become evident. It is important to note that in this study we did not intend to develop a "one size fits all" model of cognition in older adults, but rather to demonstrate the complexity and diversity that exists in a cohort that is still living independently in the community. Latent groups can be thought of as latent strata in the data that distinguish clusters of individuals displaying distinctive patterns that help in making sense of a bigger and more complex reality; like all statistical models, trajectory groups are not literal depictions of reality, are not immutable, and over time they do not follow the group-level trajectory in lock step, but they are only meant as a convenient statistical approximation (Nagin & Odgers, 2010; Nagin & Tremblay, 2005). It was within our judgment that entirely distinct classes of individuals do not exist, and that these boundaries will likely shift from sample to sample; however, we are confident that our results are a good illustration that may represent the heterogeneity of the general population.

Although this study was performed at cross-section, the EAS is an ongoing study, with opportunities for follow-up. In future work, we plan to find out if, and how, the trajectories of the individuals assigned to these groups progress, and if trends of accelerated cognitive decline and outcomes of Alzheimer's disease and dementia are present. We also plan to

harmonize these results across other cohort studies to establish if the groups are indeed replicable and if they represent meaningful strata. Possibly, with more participants included in the model, our classes may become more meaningful, and established patterns in this cohort become more distinguished. Despite the many challenges, for example, our groups may be statistical artifacts; they may be sample-specific; and they may not be replicable, these results are still useful in providing descriptive information on the cognitive profiles that exist in an older diverse cohort.

CONCLUSION

n this study, we sought a group-based model to capture the heterogeneity in cognitive function in 70 + year-old community-dwelling older adults with the overarching aim to identify specific groups with risks that are more homogeneous. In brief, our results showed that the EAS, which is reflective of an urban community, is made up of a highly diverse cohort of individuals reflected in the number of classes generated by our LCA. Our descriptive analyses further showed that some of these classes may be at risk of clinical outcomes, such as AD, VaD, and early mortality. In our future work, we plan to follow these individuals/classes to determine if their class-assignments are predictive of these outcomes.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGMENTS

REFERENCES

Army Individual Test Battery. (1944). Manual of directions and scoring. Washington, DC: War Department, Adjutant General's Office.

Backman L, Jones S, Berger AK, Laukka EJ, & Small BJ (2005). Cognitive impairment in preclinical Alzheimer' s disease: a meta-analysis. Neuropsychology, 19(4), 520–531. doi: 10.1037/0894-4105.19.4.520 [PubMed: 16060827]

Benton AL, & Hamsher K (1989). Multilingual Aphasia Examination. Iowa City: AJA Assoc.

Buschke H (1973). Selective reminding for analysis of memory and learning. Journal of Verbal Learning and Verbal Behavior, 12, 543–550.

Buschke H (1984). Cued recall in amnesia. Journal of Clinical Neuropsychology, 6(4), 433–440. [PubMed: 6501581]

Celeux G, & Soromenho G (1996). An entropy criterion for assessing the number of clusters in a mixture model. Journal of Classification, 13(2), 195–212. doi: 10.1007/bf01246098

Cohen S, Kamarck T, & Mermelstein R (1983). A Global Measure of Perceived Stress. Journal of Health and Social Behavior, 24(4), 385–396. [PubMed: 6668417]

Collins LM, Graham JW, Long JD, & Hansen WB (1994). Crossvalidation of latent class models of early substance use onset. Multivariate Behavioral Research, 29(2), 165–183. doi: 10.1207/s15327906mbr2902_3 [PubMed: 26745026]

Collins LM, & Lanza ST (2010). Latent class and latent transition analysis: With applications in the social, behavioral, and health sciences. Hoboken, NJ: John Wiley & Sons.

Costa PS, Santos NC, Cunha P, Palha JA, & Sousa N (2013). The use of bayesian latent class cluster models to classify patterns of cognitive performance in healthy ageing. PLoS One, 8 (8), e71940. doi: 10.1371/journal.pone.0071940 [PubMed: 23977183]

Davidson JE, Irizarry MC, Bray BC, Wetten S, Galwey N, Gibson R, ... Monsch AU. (2010). An exploration of cognitive subgroups in Alzheimer's disease. Journal of the International Neuropsychological Society, 16(2), 233–243. doi: 10.1017/S1355617709991160 [PubMed: 19958568]

Deary IJ, & Batty GD (2007). Cognitive epidemiology. Journal of Epidemiology and Community Health, 61(5), 378–384. doi: 10.1136/jech.2005.039206 [PubMed: 17435201]

Delano-Wood L, Bondi MW, Sacco J, Abeles N, Jak AJ, Libon DJ, & Bozoki A (2009). Heterogeneity in mild cognitive impairment: Differences in neuropsychological profile and associated white matter lesion pathology. Journal of the International Neuropsychological Society, 15(6), 906–914. doi: 10.1017/S1355617709990257 [PubMed: 19891820]

Diagnostic and Statistical Manual of Mental Disorders, DSM-IV. (1994). (pp. 133). Washington, DC: American Psychiatric Association.

Elias MF, Beiser A, Wolf PA, Au R, White RF, & D'Agostino RB (2000). The preclinical phase of alzheimer disease: A 22-year prospective study of the Framingham Cohort. Archives of Neurology, 57(6), 808–813. [PubMed: 10867777]

Executive Summary of The Third Report of The National Cholesterol Education Program (NCEP) Expert Panel on Detection, Evaluation, And Treatment of High Blood Cholesterol In Adults (Adult Treatment Panel III). (2001). JAMA, 285(19), 2486–2497. [PubMed: 11368702]

Ezzati A, Jiang J, Katz MJ, Sliwinski MJ, Zimmerman ME, &Lipton RB. (2014). Validation of the Perceived Stress Scale in a community sample of older adults. International Journal of Geriatric Psychiatry, 29(6), 645–652. doi: 10.1002/gps.4049 [PubMed: 24302253]

Franzen MD, Burgess EJ, & Smith-Seemiller L (1997). Methods of estimating premorbid functioning. Archives of Clinical Neuropsychology, 12(8), 711–738. doi: 10.1016/S0887-6177(97)00046-2 [PubMed: 14590649]

Gelber RP, Launer LJ, & White LR (2012). The Honolulu- Asia Aging Study: Epidemiologic and neuropathologic research on cognitive impairment. Current Alzheimer Research, 9(6), 664–672. [PubMed: 22471866]

Gelman A, Carlin J, Stern H, & Rubin D (2004). Bayesian Data Analysis (2nd ed.), Boca Raton, FL: Chapman and Hall/CRC.

Goldberg LR, Johnson JA, Eber HW, Hogan R, Ashton MC, Cloninger CR, & Gough HC (2006). The International Personality Item Pool and the future of public-domain personality measures. Journal of Research in Personality, 40, 84–96.

Grimm KJ, Mazza GL, & Davoudzadeh P (2017). Model selection in finite mixture models: A k-fold cross-validation approach. Structural Equation Modeling: A Multidisciplinary Journal, 24(2), 246–256. doi: 10.1080/10705511.2016.1250638

Grober E, Dickson D, Sliwinski MJ, Buschke H, Katz M, Crystal H, & Lipton RB (1999). Memory and mental status correlates of modified Braak staging. Neurobiology of Aging, 20(6), 573–579. [PubMed: 10674422]

Hall CB, Lipton RB, Sliwinski M, & Stewart WF (2000). A change point model for estimating the onset of cognitive decline in preclinical Alzheimer's disease. Statistics in Medicine, 19(11–12), 1555–1566. [PubMed: 10844718]

Hanfelt JJ, Wuu J, Sollinger AB, Greenaway MC, Lah JJ, Levey AI, & Goldstein FC (2011). An exploration of subgroups of mild cognitive impairment based on cognitive, neuropsychiatric and functional features: Analysis of data from the National Alzheimer' s Coordinating Center. American Journal of Geriatric Psychiatry, 19(11), 940–950. doi: 10.1097/JGP.0b013e31820ee9d2 [PubMed: 22024618]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Hewitt PL, Flett GL, & Mosher SW (1992). The Perceived Stress Scale: Factor structure and relation to depression symptoms in a psychiatric sample. Journal of Psychopathology and Behavioral Assessment, 14(3), 247–257. doi: 10.1007/BF00962631

Hofer SM, & Piccinin AM (2010). Toward an integrative science of life-span development and aging. The Journals of Gerontology Series B: Psychological Sciences and Social Sciences, 65B(3), 269–278. doi: 10.1093/geronb/gbq017

Hulur G, Ram N, Willis SL, Schaie KW, & Gerstorf D (2015). Cognitive dedifferentiation with increasing age and proximity of death: Within-person evidence from the Seattle Longitudinal Study. Psychology of Aging, 30(2), 311–323. doi: 10.1037/a0039260

Kaplan EF, Goodglass H, & Weintraub S (1983). The Boston Naming Test (2nd ed.), Philadelphia: Lea & Febiger.

Katz MJ, Lipton RB, Hall CB, Zimmerman ME, Sanders AE, Verghese J, ... Derby CA. (2012). Age-specific and sex-specific prevalence and incidence of mild cognitive impairment, dementia, and Alzheimer dementia in blacks and whites: a report from the Einstein Aging Study. Alzheimer Disease and Associated Disorders, 26(4), 335–343. doi: 10.1097/WAD.0b013e31823dbcfc [PubMed: 22156756]

Katzman R, Brown T, Fuld P, Peck A, Schechter R, & Schimmel H (1983). Validation of a short Orientation-Memory- Concentration Test of cognitive impairment. American Journal of Psychiatry, 140(6), 734–739. doi: 10.1176/ajp.140.6.734 [PubMed: 6846631]

Ko KJ, Berg CA, Butner J, Uchino BN, & Smith TW (2007). Profiles of successful aging in middle-aged and older adult married couples. Psychology of Aging, 22(4), 705–718. doi: 10.1037/0882-7974.22.4.705

La Rue A, & Jarvik LF (1987). Cognitive function and prediction of dementia in old age. International Journal of Aging & Human Development, 25(2), 79–89. [PubMed: 3436685]

Libon DJ, Xie SX, Eppig J, Wicas G, Lamar M, Lippa C, ... Wambach DM. (2010). The heterogeneity of mild cognitive impairment: A neuropsychological analysis. Journal of the International Neuropsychological Society, 16(1), 84–93. doi: 10.1017/s1355617709990993 [PubMed: 19887015]

Masur DM, Sliwinski M, Lipton RB, Blau AD,& Crystal HA (1994). Neuropsychological prediction of dementia and the absence of dementia in healthy elderly persons. Neurology, 44(8), 1427–1432. [PubMed: 8058143]

Maxson PJ, Berg S, & McClearn G (1997). Multidimensional patterns of aging: A cluster-analytic approach. Experimental Aging Research, 23(1), 13–31. doi: 10.1080/03610739708254024 [PubMed: 9049610]

McLachlan G, & Peel D (2000). Finite mixture models. New York: John Wiley & Sons.

Monsch AU, Bondi MW, Butters N, Salmon DP, Katzman R, & Thal LJ (1992). Comparisons of verbal fluency tasks in the detection of dementia of the Alzheimer type. Archives of Neurology, 49(12), 1253–1258. [PubMed: 1449404]

Morack J, Ram N, Fauth EB, & Gerstorf D (2013). Multidomain trajectories of psychological functioning in old age: A longitudinal perspective on (uneven) successful aging. Developmental Psychology, 49(12), 2309–2324. doi: 10.1037/a0032267 [PubMed: 23527494]

Mortamais M, Ash JA, Harrison J, Kaye J, Kramer J, Randolph C, ... Ritchie K. (2017). Detecting cognitive changes in preclinical Alzheimer's disease: A review of its feasibility. Alzheimer's & Dementia. doi: 10.1016/j.jalz.2016.06.2365

Muthen LK, & Muthen BO (1998-2016). MPlus User's Guide (7th ed.), Los Angeles, CA: Muthen & Muthén.

Nagin DS, & Odgers CL (2010). Group-based trajectory modeling (nearly) two decades later. Journal of Quantitative Criminology, 26(4), 445–453. doi: 10.1007/s10940-010-9113-7 [PubMed: 21132047]

Nagin DS, & Tremblay RE (2005). Developmental trajectory groups: Fact or a useful statistical fiction?*. Criminology, 43(4), 873–904. doi: 10.1111/j.1745-9125.2005.00026.x

Nylund KL, Asparouhov T,& Muthén BO (2007). Deciding on the number of classes in latent class analysis and growth mixture modeling: A Monte Carlo simulation study. Structural Equation Modeling: A Multidisciplinary Journal, 14(4), 535–569. doi: 10.1080/10705510701575396

Peter J, Abdulkadir A, Kaller C, Kummerer D, Hull M, Vach W, & Kloppel S (2014). Subgroups of Alzheimer's disease: stability of empirical clusters over time. Journal of Alzheimers Disease, 42(2), 651–661. doi: 10.3233/jad-140261

RStudio. Integrated development environment for R (Version 1.0.136). Boston, MA Retrieved from http://www.rstudio.org/.

Raftery AE (1995). Bayesian model selection in social research. Sociological Methodology, 111–163.

Rajan KB, Wilson RS, Weuve J, Barnes LL, & Evans DA (2015). Cognitive impairment 18 years before clinical diagnosis of Alzheimer disease dementia. Neurology, 85, 898–904. doi: 10.1212/wnl.0000000000001774 [PubMed: 26109713]

Rapp MA, & Reischies FM (2005). Attention and executive control predict Alzheimer disease in late life: results from the Berlin Aging Study (BASE). American Journal of Geriatric Psychiatry, 13(2), 134–141. doi: 10.1176/appi.ajgp.13.2.134 [PubMed: 15703322]

Ritchie CW, Molinuevo JL, Truyen L, Satlin A, Van der Geyten S, & Lovestone S (2016). Development of interventions for the secondary prevention of Alzheimer' s dementia: the European Prevention of Alzheimer's Dementia (EPAD) project. The Lancet Psychiatry, 3(2), 179–186. doi:10.1016/S2215-0366(15)00454-X [PubMed: 26683239]

Ritchie K, Ropacki M, Albala B, Harrison J, Kaye J, Kramer J, . Ritchie CW. (2017). Recommended cognitive outcomes in preclinical Alzheimer's disease: Consensus statement from the European Prevention of Alzheimer' s Dementia project. Alzheimer's & Dementia, 13, 186–195. doi: 10.1016/j.jalz.2016.07.154

Salthouse TA (2010). Selective review of cognitive aging. Journal of the International Neuropsychological Society, 16(5), 754–760. doi: 10.1017/S1355617710000706 [PubMed: 20673381]

Salthouse TA (2013). Effects of age and ability on components of cognitive change. Intelligence, 41(5), 501–511. doi: 10.1016/j.intell.2013.07.005 [PubMed: 24159248]

Sheikh JI, & Yesavage JA (1986). Geriatric Depression Scale (GDS): recent evidence and development of a shorter version. Clinical Gerontologist, 5, 165–173.

Sliwinski MJ, Hofer SM, & Hall C (2003). Correlated and coupled cognitive change in older adults with and without preclinical dementia. Psychology of Aging, 18(4), 672–683. doi: 10.1037/0882-7974.18.4.672

Smith J, & Baltes PB (1997). Profiles of psychological functioning in the old and oldest old. Psychology of Aging, 12(3), 458–472.

Snowdon DA, Kemper SJ, Mortimer JA, Greiner LH, Wekstein DR, & Markesbery WR (1996). Linguistic ability in early life and cognitive function and Alzheimer's disease in late life. Findings from the Nun Study. JAMA, 275(7), 528–532. [PubMed: 8606473]

SPSS Inc. (Released 2016). IBM SPSS Statistics for Windows. Armonk, NY: IBM Corp.

Vermunt J, & Magidson J (2002). Latent class cluster analysis In Hagenaars J & McCutcheon A (Eds.), Advances in latent class analysis. Cambridge: Cambridge University Press.

Wechsler D (1987). Wechsler Memory Scale - Revised. San Antonio, TX: The Psychological Corporation.

Wechsler D (1997). Adult Intelligence Scale-III (3rd ed.), San Antonio, TX: Psychological Corporation.

Whalley LJ, Starr JM, Athawes R, Hunter D, Pattie A, & Deary IJ (2000). Childhood mental ability and dementia. Neurology, 55(10), 1455–1459. [PubMed: 11094097]

White LR, Edland SD, Hemmy LS, Montine KS, Zarow C, Sonnen JA, ... Montine TJ. (2016). Neuropathologic comorbidity and cognitive impairment in the Nun and Honolulu-Asia Aging Studies. Neurology, 86(11), 1000–1008. doi: 10.1212/wnl.0000000000002480 [PubMed: 26888993]

Wurpts IC, & Geiser C (2014). Is adding more indicators to a latent class analysis beneficial or detrimental? Results of a MonteCarlo study. Frontiers in Psychology, 5, 920. doi: 10.3389/fpsyg.2014.00920 [PubMed: 25191298]

Ylikoski R, Ylikoski A, Keskivaara P, Tilvis R, Sulkava R, & Erkinjuntti T (1999). Heterogeneity of cognitive profiles in aging: successful aging, normal aging, and individuals at risk for cognitive decline. European Journal of Neurology, 6(6), 645–652. [PubMed: 10529751]

Zammit AR, Starr JM, Johnson W, & Deary IJ (2014). Patterns and associates of cognitive function, psychosocial wellbeing and health in the Lothian Birth Cohort 1936. BMC Geriatrics, 14, 53–53. doi: 10.1186/1471-2318-14-53 [PubMed: 24754844]
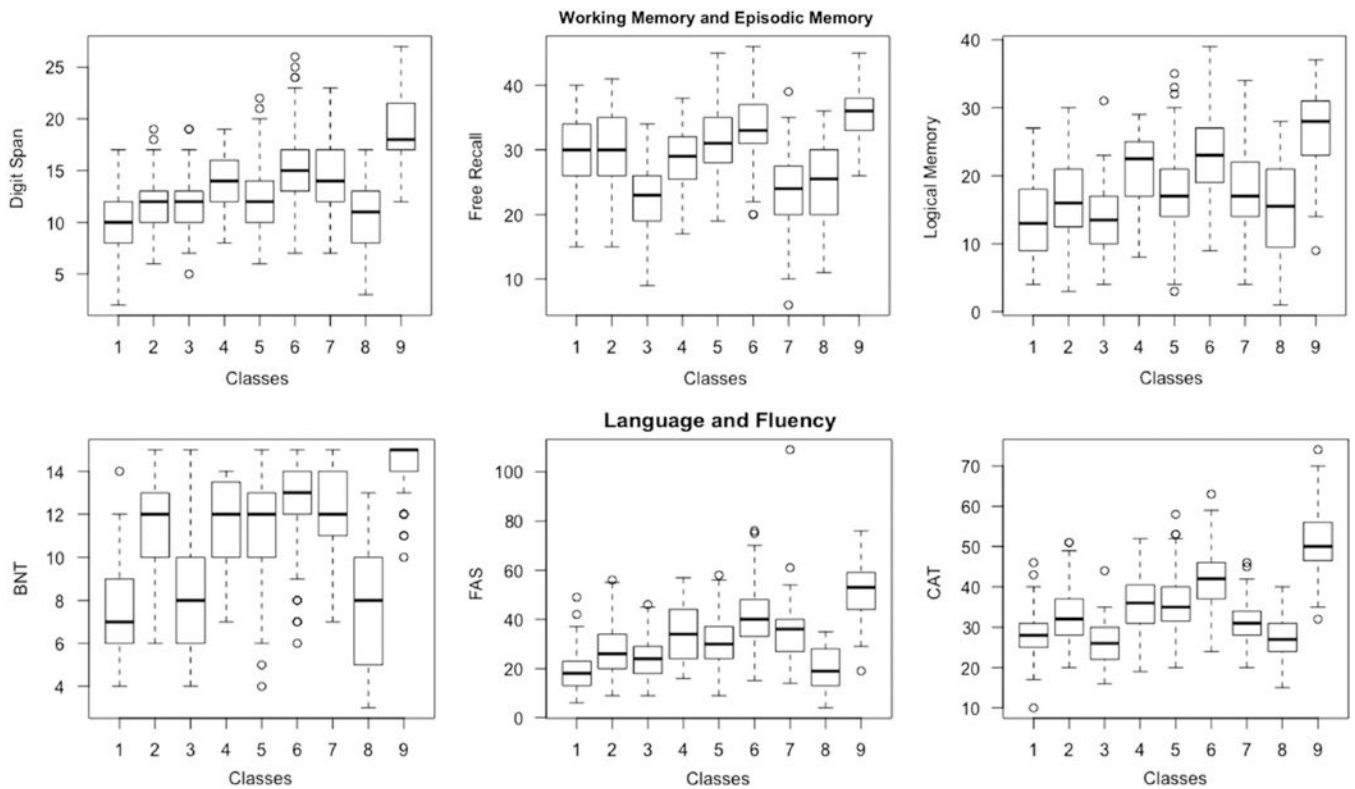
Fig. 1.
Boxplots illustrating how each of the classes performed on cognitive measures reflecting domains of Working and Episodic Memory, and Language and Fluency. *Note.* FAS =word fluency; CAT= Categories; BNT=Boston Naming Task; Class 1=the disadvantaged class; Class 2=poor language; Class 3=poor episodic memory and fluency; Class 4= poor processing speed and executive function; Class 5 =low average cognition; Class 6=high average cognition; Class 7 =average cognition; Class 8= poor executive and poor working memory; Class 9 =elite.

Prosessing Speed
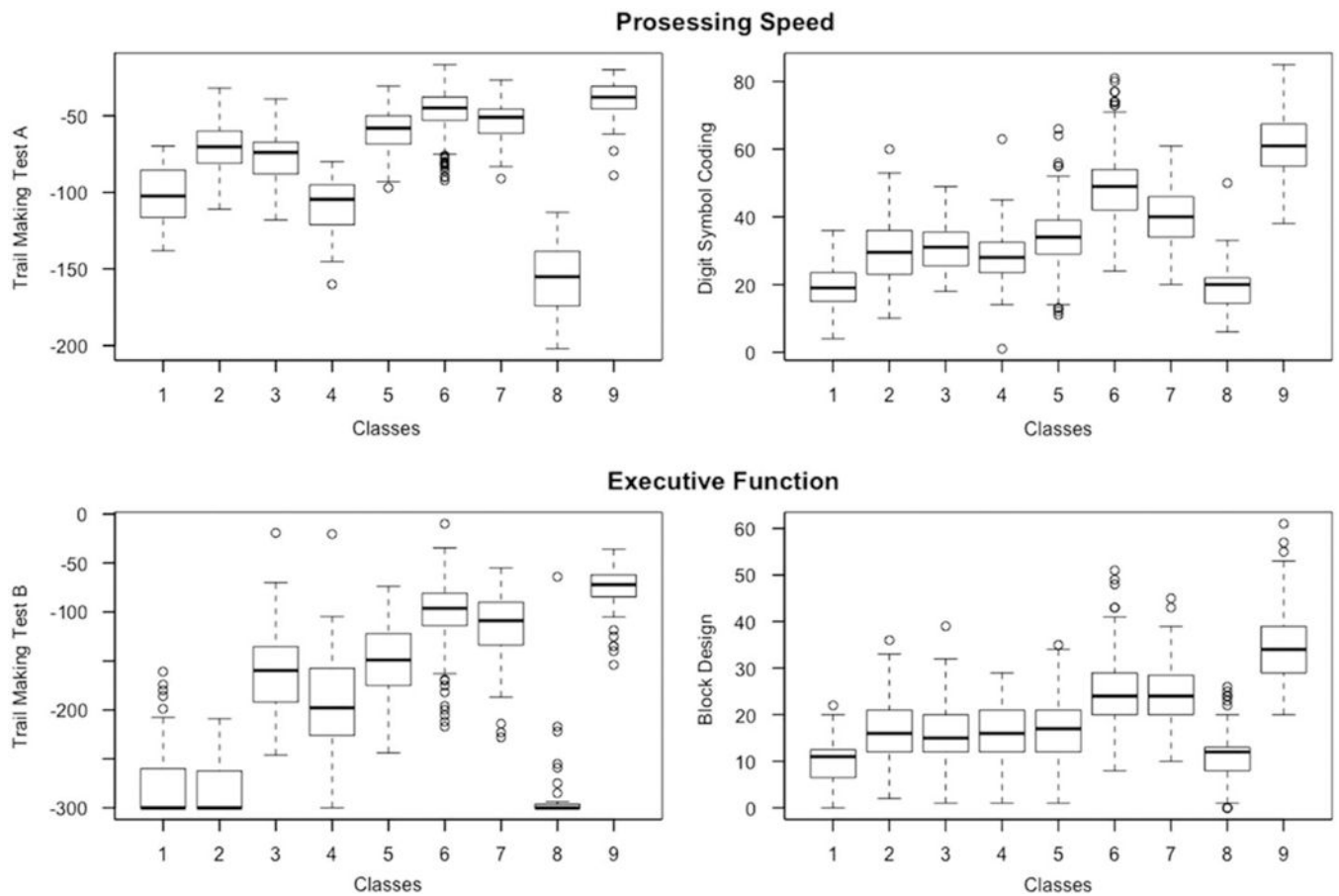
Executive Function

Fig. 2.
Boxplots illustrating how each of the classes performed on cognitive measures reflecting domains of Processing Speed and Executive Function. *Note.* Class 1= the disadvantaged class; Class 2 =poor language; Class 3= poor episodic memory and fluency; Class 4= poor processing speed and executive function; Class 5 =low average cognition; Class 6 =high average cognition; Class 7 =average cognition; Class 8=poor executive and poor working memory; Class 9= elite.

Table 1.

List of measures and how they were defined as external validators of the nine-class solutions

External validators	Definition
Demographics	
Ethnicity	We categorized race into Non-Hispanic White, African American, and Other, which included Hispanics, Asians, and others.
WRAT[a]	Administered during the clinical interview as a measure for premorbid IQ. Assessed on a continuous scale.
BIMC[b]	Administered during the clinical interview as a measure of global cognitive function. Assessed on a continuous scale.
Physical function	
Gait speed	Continuous measure assessing the time taken (in seconds) to walk 20 feet, turn, and return (40 feet total) at the subject's normal walking pace on a 12 foot instrumented walkway.
Grip strength	Continuous measure of the maximum strength on 3 trials on the dominant hand using a grip dynamometer.
Peak Flow	Continuous measure of mean expiration of 3 trials of blowing into a peak flow meter.
Psychological well-being	
Stress	Perceived Stress Scale (PSS-14, range 0–52) (Cohen, Kamarck, & Mermelstein, 1983); continuous scale with higher scores indicating more stress. We assessed the Distress and Coping components separately on a continuous scale with higher scores indicating more stress. The PSS-14 has been validated in our cohort (Ezzati et al., 2014) and applied using this approach in prior studies (Hewitt, Flett, & Mosher, 1992), including on our own dataset (Ezzati et al., 2014).
Depression	The Geriatrics Depression Scale - short form (Sheikh & Yesavage, 1986); continuous scale, with higher scores indicating more depression symptomatology.
Neuroticism	The International Personality Item Pool (IPIP) (Goldberg et al., 2006) was used to measure neuroticism.
Health and disease	
Smoking	Ordinal: 1) Never; 2) past; 3) current
Diabetes (DM)	DM was defined according to the NCEP Adult Treatment Panel III ("Executive Summary of The Third Report of The National Cholesterol Education Program (NCEP) Expert Panel on Detection, Evaluation, And Treatment of High Blood Cholesterol In Adults (Adult Treatment Panel III)," 2001) : Elevated fasting glucose at 100mg/dL. All participants with DM in the EAS were treated.
Hypertension (HTN)	HTN was defined according to the NCEP Adult Treatment Panel III ("Executive Summary of The Third Report of The National Cholesterol Education Program (NCEP) Expert Panel on Detection, Evaluation, And Treatment of High Blood Cholesterol In Adults (Adult Treatment Panel III)," 2001):Systolic blood pressure 130mmHg diastolic BP 85 mmHg.We analyzed both treated and untreated HTN.
History of stroke	Collected from the medical history during the clinical interview.

[a] WRAT = Wide Range Achievement Test - we used this as a marker of premorbid IQ (Franzen, Burgess, & Smith-Seemiller, 1997).

[b] BIMC= Blessed Information Memory Concentration Test (Katzman et al., 1983).

Table 2.

Characteristics of whole and subsamples 1 and 2

	Whole sample N = 1,345	Sample 1 N = 672	Sample 2 N = 673
Females (%)	828 (61.6)	442 (51.0)	406 (49.0)
Age, years (*SD*)	78.0 (5.4)	77.9 (5.2)	78.1 (5.5)
Education, years (*SD*)	13.6 (3.5)	13.5 (3.5)	13.7 (3.6)
Non-Hispanic white (%)	914 (68)	458 (68.2)	456 (67.8)
BIMC (*SD*)	2.4 (2.3)	2.4 (2.3)	2.4 (2.3)
WRAT (*SD*)	67.5 (15.2)	67.0 (16.2)	67.9 (14.3)

Note. BIMC=Blessed Information Memory Concentration Test; WRAT=Wide Range Achievement Test.

Table 3.

Demographic characteristics that were and were not included in the latent 9-class model

Characteristics	Class 1	Class 2	Class 3	Class 4	Class 5	Class 6	Class 7	Class 8	Class 9	F/X²	p-Value
N(%)	77 (5.70%)	90 (6.70%)	81 (6.00%)	37 (2.80%)	346 (25.7%)	440 (32.7%)	136 (10.10%)	34 (2.5%)	104 (7.7%)		
*Age, years (SD)	78.0 (5.9)	81.3 (6.3)	79.0 (5.0)	78.7 (5.1)	78.3 (4.9)	76.6 (4.7)	80.4 (5.8)	81.4 (5.9)	75.1 (3.9)	19.7 (8, 1336)	<.001
*Females (%)	58 (75.3)	59 (65.6)	39 (48.1)	21 (2.5)	238 (68.6)	284 (65.4)	39 (28.7)	20 (58.8)	70 (67.3)	86.3	<.001
*Education, years (SD)	8.4 (2.8)	11.6 (2.7)	12.7 (2.9)	16.5 (3.4)	12.2 (2.6)	15.0 (2.9)	15.3 (3.2)	11.4 (2.8)	16.8 (2.7)	91.3 (8, 1336)	<.001
WRAT (SD)	25.1 (4.4)	29.7 (5.7)	31.6 (5.8)	34.5 (5.1)	31.2 (5.7)	35.9 (4.9)	35.5 (5.1)	29.6 (5.4)	39.0 (3.4)	29.5 (8, 568)	<.001
BIMC (SD)	4.9 (2.5)	3.4 (2.4)	3.7 (2.3)	2.4 (2.3)	2.8 (2.1)	1.4 (1.4)	2.8 (2.4)	4.4 (2.5)	0.8 (0.9)	52.8 (8, 1335)	<.001
Non-Hispanic White	34 (44.2)	42 (46.7)	54 (66.7)	17 (45.9)	220 (63.6)	331 (75.2)	110(80.9)	14 (41.2)	92 (88.5)	102.5	<.001
African American	34 (44.2)	44 (48.9)	23. (28.4)	19 (51.4)	103 (29.8)	91 (20.7)	19 (14.0)	19 (55.9)	7 (7.6)		
Other	9(11.7)	4 (4.4)	4 (4.9)	1 (2.7)	23 (6.6)	18(4.1)	7(5.1)	1 (2.9)	5 (4.8)		

Note. Asterisks indicate variables that were used as covariates in the latent class model.

WRAT = Wide Range Achievement Test; BIMC = Blessed Information Memory Concentration test; Class 1 =the disadvantaged class; Class 2 = poor language; Class 3 =poor episodic memory and fluency; Class 4 = poor processing speed and executive function; Class 5 = low average cognition; Class 6 = high average cognition; Class 7 = average cognition; Class 8 = poor executive and poor working memory; Class 9 = elite.

Table 4.

Test performance by class membership on the neuropsychological battery that was used to form the latent classes (*SDs* in parentheses)

Tests	Class 1	Class 2	Class 3	Class 4	Class 5	Class 6	Class 7	Class 8	Class 9	F	p-Value
N	77 (5.70%)	90 (6.70%)	81 (6.00%)	37 (2.80%)	346 (25.7%)	440 (32.7%)	136 (10.10%)	34 (2.5%)	104 (7.7%)		
FCSRT - Free Recall	29.3 (6.1)	30.3 (5.4)	23.4 (5.8)	29.1 (5.4)	31.4(4.7)	33.3 (4.3)	23.6 (5.4)	24.8 (6.8)	35.5 (4.0)	98.9 (8, 1321)	<.001
Boston Naming	7.7 (2.0)	11.2 (2.1)	8.1 (2.4)	11.7 (2.1)	11.5 (2.1)	13.0(1.7)	12.0 (2.1)	7.8 (2.7)	14.2 (1.1)	140.6 (8, 1316)	<.001
Digit Span	10.0 (2.9)	12.0 (2.6)	11.8 (2.7)	14.0 (2.9)	12.3 (2.7)	15.0 (3.2)	14.6 (3.2)	10.7 (3.1)	18.9 (3.7)	84.6 (8, 1331)	<.001
Digit Symbol Coding	19.2 (6.6)	30.4 (9.0)	31.0 (7.2)	28.3 (10.1)	34.1 (8.0)	48.9 (9.5)	40.9 (8.4)	19.6 (8.2)	61.1 (10.1)	258.6 (8, 1323)	<.001
Block Design	9.8 (5.2)	16.7 (7.1)	15.7 (6.6)	15.8 (6.2)	17.0 (6.3)	24.6 (7.0)	24.4 (6.5)	11.7 (7.6)	34.9 (8.1)	121.6 (8, 1175)	<.001
Word Fluency	19.1 (8.2)	27.0 (9.7)	24.0 (8.5)	34.2 (10.9)	30.9 (9.5)	40.5 (10.6)	35.5 (11.1)	19.8 (9.6)	51.3 (10.8)	100.8 (8, 1230)	<.001
Categories	28.0 (6.3)	33.1 (7.0)	25.9 (5.8)	35.9 (7.4)	35.7 (6.2)	41.7 (6.7)	31.3 (5.2)	27.0 (5.7)	51.3 (8.3)	167.0 (8, 1326)	<.001
Logical Memory	13.5 (5.7)	16.7 (5.9)	13.6 (5.3)	20.8 (5.6)	17.5 (5.7)	23.2 (5.9)	17.5 (5.7)	15.2 (7.2)	27.5 (5.7)	76.4 (8, 1285)	<.001
Trail Making Test A	100.3 (18.7)	69.6 (16.1)	76.1 (16.1)	109.1 (18.6)	59.2 (13.0)	46.2 (12.4)	53.6 (12.8)	157.2 (23.3)	39.1 (11.3)	410.1 (8, 1244)	<.001
Trail Making Test B	277.1 (37.8)	280.8 (26.4)	160.7 (43.7)	193.4 (56.0)	149.8 (33.1)	100.1 (27.6)	113.8 (33.4)	282.9 (46.3)	75.0 (20.9)	585.5 (8, 1244)	<.001

Note. Class 1 = the disadvantaged class; Class 2 = poor language; Class 3 = poor episodic memory and fluency; Class 4 = poor processing speed and executive function; Class 5 = low average cognition; Class 6 = high average cognition; Class 7 = average cognition; Class 8 = poor executive and poor working memory; Class 9 = elite.

Table 5.

Means of each of the 9 classes on the external variables used for external validity divided into sections of physical fitness, psychological wellbeing, and health and disease (*SDs* in parentheses)

	Class 1	Class 2	Class 3	Class 4	Class 5	Class 6	Class 7	Class 8	Class 9	F/χ^2	p-Value
N	77 (5.70%)	90 (6.70%)	81 (6.00%)	37 (2.80%)	346 (25.7%)	440 (32.7%)	136 (10.10%)	34 (2.5%)	104 (7.7%)		
Physical fitness											
Peak flow average, L/min	226.0 (101.3)	252.1 (102.1)	246.5 (103.0)	289.0 (118.9)	271.8 (102.4)	300.1 (109.4)	318.8 (124.1)	300.0 (115.3)	313.4 (101.3)	3.1 (8, 533)	.002
Grip strength, kg	18.5 (7.1)	19.4 (8.2)	17.7 (7.6)	20.9 (6.9)	20.1 (7.5)	21.3 (7.9)	22.7 (7.3)	17.5 (11.7)	20.9 (8.0)	1.5 (8, 546)	.166
Gait speed, cm/sec	76.2 (25.3)	83.4 (23.9)	90.2 (26.1)	89.6 (28.7)	90.2 (22.5)	99.6 (19.6)	94.0 (20.0)	79.3 (25.9)	108.6 (20.6)	9.1 (8, 558)	<.001
Psychological wellbeing GDS	3.4 (3.2)	2.6 (2.7)	3.2 (2.7)	2.5 (2.5)	2.7 (2.4)	2.0 (2.2)	2.1 (1.8)	3.0 (1.9)	1.4 (1.8)	7.1 (8, 1115)	<.001
PSS - Distress	4.8 (3.8)	5.4 (3.9)	6.7 (5.4)	6.4 (5.8)	8.2 (5.0)	6.9 (4.3)	6.5 (4.6)	6.9 (5.4)	6.5 (3.8)	1.9 (8, 511)	.054
PSS - Coping	12.5 (6.4)	9.8 (6.1)	11.1 (6.3)	10.0 (6.7)	9.3 (5.4)	8.5 (4.8)	8.6 (5.4)	12.6 (2.3)	7.4 (4.8)	2.8 (8, 508)	.005
Neuroticism	24.7 (8.1)	17.5 (4.6)	28.0 (8.7)	21.0 (8.0)	22.2 (5.8)	21.1 (6.2)	20.6 (6.8)	21.8 (6.2)	17.8 (5.3)	3.6 (8, 268)	<.001
Health and disease											
Current smokers	7 (9.5%)	10 (11.5%)	4 (5.2%)	3 (8.8%)	30 (8.9%)	29 (7.0%)	4 (3.0%)	1 (3.0%)	5 (5.3%)	10.0	.268
Past smokers	31 (40.2%)	42 (46.6%)	38 (46.9%)	16 (43.2%)	171 (49.4%)	263 (59.7%)	75 (55.1%)	13 (38.2%)	60 (57.7%)	30.3	.001
Diabetes	26 (33.8%)	11 (12.5%)	19 (23.5%)	5 (14.3%)	49 (14.4%)	60 (14.3%)	18 (13.3%)	8 (23.5%)	8 (8.4%)	49.1	<.001
Hypertension (130/85)	48 (62.3%)	62 (68.8%)	53 (65.4%)	20 (54.0%)	233 (67.3%)	298 (67.7%)	91 (66.9%)	27 (79.4%)	73 (70.1%)	14.4	.072
Treated hypertensives	45 (58.4%)	51 (58.0%)	49 (60.5%)	19 (54.3%)	214 (62.9%)	247 (58.7%)	74 (54.8%)	27 (79.4%)	52 (54.7%)	10.1	.261
Stroke	6 (7.8%)	11 (12.5%)	6 (7.4%)	6 (17.1%)	31 (9.1%)	30(7.1%)	15 (11.1%)	8 (23.5%)	0 (0%)	27.7	.001

Note. GDS = Geriatric Depression Scale; PSS = Perceived Stress Scale; Class 1 = the disadvantaged class; Class 2 = poor language; Class 3 = poor episodic memory and fluency; Class 4 = poor processing speed and executive function; Class 5 = low average cognition; Class 6 = high average cognition; Class 7 = average cognition; Class 8 = poor executive and poor working memory; Class 9 = elite.