2016

# Sensitivity of composite scores to amyloid burden in preclinical Alzheimer's disease: Introducing the Z-scores of Attention, Verbal fluency, and Episodic memory for Nondemented older adults composite score

Yen Ying Lim

Peter J. Snyder

Robert H. Pietrzak

Albulene Ukiqi

Victor L. Villemagne

*See next page for additional authors*

## Authors

Yen Ying Lim, Peter J. Snyder, Robert H. Pietrzak, Albulene Ukiqi, Victor L. Villemagne, David Ames, Olivier Salvado, Pierrick Bourgeat, Ralph N. Martins, Colin L. Masters, Christopher C. Rowe, and Paul Maruff

Cognitive & Behavioral Assessment

# Sensitivity of composite scores to amyloid burden in preclinical Alzheimer's disease: Introducing the Z-scores of Attention, Verbal fluency, and Episodic memory for Nondemented older adults composite score

Yen Ying Lim[a,*], Peter J. Snyder[b,c], Robert H. Pietrzak[d], Albulene Ukiqi[a],
Victor L. Villemagne[a,e,f], David Ames[g,h], Olivier Salvado[i], Pierrick Bourgeat[i], Ralph N. Martins[j],
Colin L. Masters[a], Christopher C. Rowe[e,f], Paul Maruff[a,k], for the AIBL Research Group

[a]The Florey Institute of Neuroscience and Mental Health, The University of Melbourne, Parkville, Victoria, Australia
[b]Department of Neurology, Warren Alpert School of Medicine, Brown University, Providence, RI, USA
[c]Interdisciplinary Neuroscience Program, University of Rhode Island, Kingston, RI, USA
[d]Department of Psychiatry, Yale University School of Medicine, New Haven, CT, USA
[e]Department of Nuclear Medicine and Centre for PET, Austin Health, Heidelberg, Victoria, Australia
[f]Department of Medicine, Austin Health, The University of Melbourne, Heidelberg, Victoria, Australia
[g]Academic Unit for Psychiatry of Old Age, St. Vincent's Health, The University of Melbourne, Kew, Victoria, Australia
[h]National Ageing Research Institute, Parkville, Victoria, Australia
[i]Commonwealth Scientific Industrial Research Organization (CSIRO) Preventative Health National Research Flagship, Australian e-Health Research Centre-BioMedIA, Brisbane, Queensland, Australia
[j]Centre of Excellence for Alzheimer's Disease Research and Care, Edith Cowan University, Joondalup, Western Australia, Australia
[k]Cogstate Ltd., Melbourne, Victoria, Australia

**Abstract**

**Introduction:** Cognitive composite scores developed for preclinical Alzheimer's disease (AD) often consist of multiple cognitive domains as they may provide greater sensitivity to detect β-amyloid (Aβ)–related cognitive decline than episodic memory (EM) composite scores alone. However, this has never been empirically tested. We compared the rate of cognitive decline associated with high Aβ (Aβ+) and very high Aβ (Aβ++) in cognitively normal (CN) older adults on three multidomain cognitive composite scores and one single-domain (EM) composite score.

**Methods:** CN older adults (n = 423) underwent Aβ neuroimaging and completed neuropsychological assessments at baseline, and at 18-, 36-, 54-, and 72-month follow-ups. Four cognitive composite scores were computed: the ADCS-PACC (ADCS-Preclinical Alzheimer Cognitive Composite), ADCS-PACC without the inclusion of the mini-mental state examination (MMSE), an EM composite, and the Z-scores of Attention, Verbal fluency, and Episodic memory for Nondemented older adults (ZAVEN) composite.

**Results:** Compared with Aβ+ CN older adults, Aβ++ CN older adults showed faster rates of decline across all cognitive composites, with the largest decline observed for ZAVEN composite (d = 1.07). Similarly, compared with Aβ− CN older adults, Aβ+ CN older adults also showed faster rates of cognitive decline, but only for the ADCS-PACC no MMSE (d = 0.43), EM (d = 0.53), and ZAVEN (d = 0.50) composites.

**Discussion:** Aβ-related cognitive decline is best detected using validated neuropsychological instruments. Removal of the MMSE from the ADCS-PACC and replacing it with a test of executive

*Corresponding author. Tel.: +61-3-9389-2909; Fax: +61-3-9387-5061.

E-mail address: yen.lim@florey.edu.au

function (verbal fluency; i.e., the ZAVEN) rendered this composite more sensitive even in detecting Aβ-related cognitive decline between Aβ+ and Aβ++ CN older adults.

## 1. Introduction

There is now consensus that in cognitively normal (CN) older adults, high levels of β-amyloid (Aβ), assessed using Aβ imaging or cerebrospinal fluid sampling, represent the preclinical stage of Alzheimer's disease (AD) [1–3]. Multiple prospective studies have shown that substantial decline in cognitive function occurs in Aβ+ CN older adults over periods of 6–54 months, even in the absence of any progression to clinically recognizable mild cognitive impairment (MCI) or AD [4–7]. In Aβ+ CN older adults, this cognitive decline is associated with faster accumulation of Aβ [1,8] as well as greater loss of hippocampal volume and decreased levels of brain metabolism [9,10]. Although there is general agreement that Aβ levels should be classified as low (Aβ−) or high (Aβ+) [11,12], a recent analysis from our group using a two-graph receiver operator curve (ROC) analysis of Aβ levels in the Australian Imaging, Biomarkers, and Lifestyle (AIBL) cohort indicated that a standardized uptake value ratio (SUVR) of 1.9 provided the optimal cut point for distinction of Aβ levels in people with dementia from age-matched healthy controls [2]. Hence, when Aβ levels in CN older adults were classified additionally as being high (Aβ+: SUVR 1.50–1.90) or very high (Aβ++: SUVR >1.90), only Aβ++ CN individuals showed increased rates of cognitive decline relative to Aβ− CN older adults. In fact, CN older adults with Aβ+ did not show cognitive decline over a 36-month period [5]. Furthermore, in CN older adults, Aβ++ was associated with a higher risk of progression to MCI or AD compared with CN adults with Aβ+ [2]. Taken together, these data suggest that it may be prudent to consider Aβ burden beyond a single positive/negative category in the design of clinical trials for new anti-Aβ therapies [13,14].

An important consideration in measuring the effects of Aβ in clinical research studies and clinical trials of preclinical AD is the method used to operationalize a cognitive end point. Currently, there is consensus that the composite measures used commonly to characterize disease progression in patients with prodromal AD or dementia, such as the mini-mental status examination (MMSE), the Alzheimer's Disease Assessment Scale-Cognitive Subscale (ADAS-Cog), or the clinical dementia rating (CDR) scale, are not appropriate for use in CN older adults because data distributions from these scales are characterized by restricted range of possible scores, ceiling effects, and negative skew, thus rendering it insensitive to subtle changes [15–17]. As such, research into cognitive decline in preclinical AD uses composite outcome measures based on data from standardized neuropsychological tests, such as tests of episodic memory and executive function, on which performance is most affected in this early disease stage [5,18,19].

Multiple studies from different natural history cohorts indicate that in preclinical AD, episodic memory provides a highly reliable and sensitive index of Aβ-related cognitive decline [3–5,20]. Therefore, episodic memory (EM) composite scores provide a sound comparator for determining the extent to which newer composite measures based on tests that measure cognitive domains other than episodic memory can yield any improved sensitivity of Aβ-related cognitive decline. Some composite scores developed for preclinical AD include measures of additional cognitive domains on the basis that their inclusion may provide greater sensitivity to detect Aβ-related cognitive decline than EM composite scores alone [21]. For example, the recently validated cognitive composite for the Anti-Amyloid treatment in Asymptomatic Alzheimer's disease (A4) trial, the Alzheimer Disease Cooperative Study (ADCS) Preclinical Alzheimer Cognitive Composite (ADCS-PACC), emphasized measurement of specific cognitive domains, rather than specifying the tests used to operationally define those domains. The ADCS-PACC combines measures of episodic memory (e.g., measures of list learning such as the Free and Cued Selective Reminding Test [FCSRT] or the California Verbal Learning Test, Second Edition [CVLT-II] and measures of paragraph recall such as the Wechsler Memory Scale Logical Memory delayed recall [LM-DR] or New York University Paragraph Recall test), complex attention (e.g., the Wechsler Adult Intelligence Scale-Revised Digit Symbol Substitution Test [DSST] score), and a general cognitive screen (e.g., the MMSE total score) [18]. Although each of the tests used to define episodic memory and complex attention in the ADCS-PACC have demonstrated sensitivity to cognitive decline in early AD, the MMSE has not [16,17]. Therefore, its inclusion may reduce the sensitivity of the ADCS-PACC because of its suboptimal metric characteristics when its use is restricted to CN older adults (i.e., ceiling effects, negative skew, poor test-retest reliability) [15–17].

An additional limitation of the ADCS-PACC is that it does not include a measure of executive function when substantial Aβ-related decline in this domain is also observed reliably in preclinical AD, often to a greater extent than

that observed for attentional function [5,22]. Thus, cognitive composite scores used in studies of preclinical AD may have increased sensitivity to drug effects and Aβ-related cognitive decline if they also included a measure of executive function. Consequently, the sensitivity of the ADCS-PACC to detecting Aβ-related cognitive decline in preclinical AD might be improved if the MMSE was replaced with a measure of executive function (e.g., a measure of verbal fluency; [5]), so that it will reflect primarily attention, verbal fluency, and episodic memory. To address this possibility, we compared the rate of cognitive decline associated with Aβ+ and Aβ++ in CN adults using the ADCS-PACC, the ADCS-PACC without the MMSE, an EM composite, and a composite score derived from neuropsychological tests of attention, executive function, and episodic memory. The first hypothesis was that cognitive decline in Aβ++ CN older adults would be greater than that in Aβ+ CN older adults, which would in turn be greater than in Aβ− CN older adults. The second hypothesis was that composite scores that did not include the MMSE would be more sensitive in detecting Aβ-related cognitive decline in Aβ+ and Aβ++ CN older adults.

## 2. Methods

### 2.1. Participants

Participants were recruited from a group of CN older adults enrolled in the AIBL study, the recruitment of which has been described previously [12,23]. Briefly, exclusion criteria included schizophrenia; depression (15-item geriatric depression score ≥6); Parkinson's disease; symptomatic stroke; uncontrolled diabetes; sleep apnea; and alcohol use exceeding two standard drinks per day for women or four per day for men. Participants underwent medical, psychiatric, and neuropsychological assessments at baseline, and 18-, 36-, 54-, and 72-month follow-ups [23]. At each assessment, a clinical review panel considered all available

medical, psychiatric, and neuropsychological information to classify clinical status [23]. Clinical classification was blinded to neuroimaging results. Group demographic and clinical characteristics are shown in Table 1. The study was approved by and complied with the regulations of three institutional research and ethics committees [23]. All participants provided written informed consent.

### 2.2. Measures

#### 2.2.1. Neuroimaging and APOE genotyping

Fasted blood samples (80 mL) were collected from each participant, of which 0.5 mL was sent to a clinical pathology laboratory for *APOE* genotyping. Aβ imaging with positron emission tomography (PET) was conducted using either $^{11}$C-Pittsburgh Compound B (PiB), $^{18}$F-florbetapir, or $^{18}$F-flutemetamol. PET methodology has been described in detail previously [12,24,25]. A 30-minute acquisition was started 40 minutes after injection of PiB, a 20-minute acquisition was performed 50 minutes after injection of florbetapir and 90 minutes after injection of flutemetamol. For PiB-PET, standardized uptake value (SUV) data were summed and normalized to the cerebellar cortex SUV, resulting in a region-to-cerebellar ratio termed SUVR. The whole cerebellum was the reference region for florbetapir [24], whereas for flutemetamol, the reference region was the pons [25]. SUVR was classified as either negative (Aβ−), high positive (Aβ+), or very high positive (Aβ++). For PiB, the usual 1.5 SUVR was used to discriminate between Aβ− and Aβ+, whereas a 1.9 SUVR obtained from a ROC curve analysis was used as the optimal cut point to discriminate age-matched CN older adults from AD patients. Thus, this higher SUVR cut point was used to discriminate between Aβ++ (SUVR >1.9) and Aβ+ (SUVR 1.5–1.9) [2]. Similarly, for flutemetamol, SUVR thresholds of 0.61, 0.61–0.82, and ≥0.82 were used to discriminate between Aβ−, Aβ+,

Table 1
Demographic and clinical characteristics of the sample by Aβ status

| Characteristics | CN Aβ− (n = 326) n (%) or mean (SD) | CN Aβ+ (n = 33) n (%) or mean (SD) | CN Aβ++ (n = 64) n (%) or mean (SD) | P |
|---|---|---|---|---|
| Female sex | 179 (54.9) | 22 (66.7) | 30 (46.9) | .174 |
| *APOE* ε4 genotype | **64 (19.6)** | **19 (57.6)** | **32 (50.0)** | **.000** |
| Age | **68.27 (5.95)** | **73.06 (7.13)** | **73.19 (7.41)** | **.000** |
| Premorbid IQ | 108.27 (7.07) | 109.18 (7.75) | 110.37 (6.62) | .085 |
| GDS | 0.92 (1.40) | 0.76 (1.17) | 0.63 (1.03) | .247 |
| HADS depression | 2.61 (2.20) | 2.97 (2.87) | 2.49 (2.59) | .623 |
| HADS anxiety | 4.26 (2.85) | 4.76 (2.77) | 4.14 (2.95) | .583 |
| CDR | 0.03 (0.11) | 0.08 (0.22) | 0.02 (0.11) | .096 |
| CDR sum of boxes | 0.04 (0.16) | 0.05 (0.15) | 0.02 (0.11) | .746 |
| MMSE | 28.95 (1.14) | 28.58 (1.25) | 28.91 (1.18) | .203 |

Abbreviations: Aβ, β-amyloid; CN, cognitively normal; SD, standard deviation; *APOE*, apolipoprotein E; GDS, geriatric depression scale; HADS, hospital anxiety and depression scale; CDR, clinical dementia rating scale; MMSE, mini-mental state examination; premorbid IQ was assessed using the Wechsler Test of Adult Reading.

NOTE. Bold values indicate statistical significance at *P* <.001 level.

and Aβ++ CNs, whereas for florbetapir, the SUVR thresholds used were 1.10, 1.10–1.29, and ≥1.29, respectively [2].

### 2.2.2. Cognitive composite scores

Cognitive composite scores were computed by standardizing outcome measures for each neuropsychological test on each assessment using baseline mean and standard deviation (SD) for the entire group. This yielded four cognitive composite scores as follows: (1) the ADCS-PACC was computed as described previously [18], by averaging standardized scores on measures of episodic memory (CVLT-II total recall score and LM-DR score), attention (DSST score), and the MMSE; (2) the ADCS-PACC no MMSE was computed by averaging the standardized scores on the CVLT-II total recall score, LM-DR, and DSST only; (3) the AIBL EM composite was computed as described previously [5], by averaging standardized scores on the CVLT-II delayed recall, LM-DR, and Rey Complex Figure Test delayed recall tasks; and (4) the Z-scores of Attention, Verbal fluency, and Episodic memory for Nondemented older adults (ZAVEN) composite was computed by averaging standardized scores on measures of attention (DSST), executive function (FAS), and episodic memory (CVLT-II total recall and LM-DR).

### 2.3. Data analysis

Data analyses proceeded in two steps. First, average measure intraclass correlation coefficients were used to compute the test-retest reliability of each cognitive composite over the five time points, and skew statistics were reported for each composite at baseline. Second, for each cognitive composite score, three planned comparisons were conducted using repeated-measures linear mixed-effects models with maximum likelihood estimation and an unstructured covariance matrix. Linear mixed modeling was used because of its ability to model both fixed and random effects, which accounts for multiple sources of variability in longitudinal studies. Additionally, both empirical and theoretical models of AD show that once the threshold for Aβ positivity is reached, there is a linear trend in cognitive decline, neuro-

degeneration, and Aβ accumulation until a clinical diagnosis of AD is reached [1,8,26]. In these analyses, group (Aβ−, Aβ+, and Aβ++), time, and group × time interaction were entered as fixed factors; participant as a random factor; age and APOE ε4 status as covariates; and cognitive composite score as the dependent variable. Within the model, the magnitude of difference from the Aβ− group was expressed using Cohen's d.

## 3. Results

### 3.1. Demographic and clinical characteristics

At baseline, there were statistically significant group differences in age and APOE ε4 carriage (Table 1). As such, these variables were entered as covariates in all subsequent analyses. None of the other demographic or clinical characteristics differed between groups. Participants across all groups showed high levels of premorbid intelligence, low levels of depressive and anxiety symptoms, and their cognitive health was reflected in the low scores on the CDR and high scores on the MMSE (Table 1).

### 3.2. Test-retest reliability of cognitive composite scores

High test-retest reliability across the three assessment time points was observed for the EM composite ($r = 0.93$, $P < .001$), the ADCS-PACC ($r = 0.92$, $P < .001$), the ADCS-PACC no MMSE ($r = 0.94$, $P < .001$), and the ZAVEN ($r = 0.96$, $P < .001$). Data for each composite were normally distributed, with skewness of −0.28 (standard error [SE] = 0.12) for EM, −0.36 (SE = 0.12) for the ADCS-PACC, −0.09 (SE = 0.13) for ADCS-PACC no MMSE, and −0.16 for the ZAVEN (SE = 0.12).

### 3.3. Effect of Aβ levels on cognitive change in cognitive composite scores

There were significant interactions between group and time for all cognitive composite scores (Table 2). Group mean slopes for each Aβ/ε4 group for each composite

Table 2
Effect of Aβ group on cognitive decline over 72 months in CN older adults and group mean (SD) of slope (rate of change) in each cognitive composite for each Aβ group

| Cognitive composite | Group (df) F | P | Time (df) F | P | Group × time (df) F | P | CN Aβ− (n = 326) Mean slope (SD) | CN Aβ+ (n = 33) Mean slope (SD) | CN Aβ++ (n = 64) Mean slope (SD) |
|---|---|---|---|---|---|---|---|---|---|
| EM | (2437) 3.69 | .03 | (1365) 5.48 | .00 | (2364) 3.94 | .00 | 0.027 (0.163) | −0.057 (0.125) | −0.132 (0.131) |
| ADCS-PACC | (2442) 5.95 | .00 | (1396) 21.80 | .00 | (2392) 20.39 | .00 | 0.002 (0.335) | −0.082 (0.245) | −0.205 (0.263) |
| ADCS-PACC (no MMSE) | (2441) 0.904 | .02 | (1380) 12.93 | .00 | (2378) 23.68 | .00 | 0.028 (0.165) | −0.042 (0.126) | −0.134 (0.133) |
| ZAVEN | (2435) 4.80 | .01 | (1377) 24.28 | .00 | (2375) 26.88 | .00 | 0.014 (0.141) | −0.055 (0.108) | −0.133 (0.113) |

Abbreviations: Aβ, β-amyloid; CN, cognitively normal; SD, standard deviation; EM, episodic memory composite; ADCS-PACC; ADCS Preclinical Alzheimer Cognitive Composite; ADCS-PACC (no MMSE), ADCS Preclinical Alzheimer Cognitive Composite without MMSE; MMSE, mini-mental state examination; ZAVEN, Z-scores of Attention, Verbal fluency, and Episodic memory in Nondemented older adults.
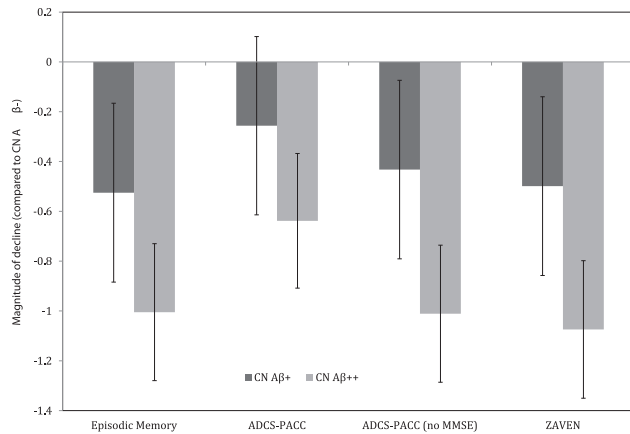NOTE. All models have been adjusted for age and APOE.

Fig. 1. Magnitude of decline on each cognitive composite over 72 months relative to the CN Aβ− group (error bars represent 95% CIs). Abbreviations: CN, cognitively normal; Aβ, β-amyloid; CI, confidence interval; ADCS-PACC, ADCS-Preclinical Alzheimer Cognitive Composite; MMSE, mini-mental state examination; ZAVEN, Z-scores of Attention, Verbal fluency, and Episodic memory in Nondemented older adults.

cognitive score are summarized in Table 2. Compared with Aβ− CN older adults, Aβ+ CN older adults showed a faster rate of decline, but only on the ADCS-PACC no MMSE, ZAVEN, and EM composites (Table 2), with these differences moderate in magnitude (Fig. 1). Relative to Aβ+ CN older adults, Aβ++ CN older adults showed a significantly faster decline on all cognitive composites (Table 2), with these differences also moderate in magnitude (Fig. 1).

## 4. Discussion

The results of this study supported the first hypothesis that cognitive decline in Aβ++ CN adults would be greater than that in Aβ+ CN adults, which would in turn be greater than in Aβ− CN adults. Specifically, compared with Aβ+ CN older adults, Aβ++ CN older adults showed faster rates of decline across all cognitive composite scores, with the largest decline observed for the ZAVEN composite ($d = 0.70$; Table 2). Similarly, compared with Aβ− CN older adults, Aβ+ CN older adults also showed faster rates of cognitive decline, but only for the EM and ZAVEN composite scores (Table 2). We and others have typically grouped Aβ levels into two categories (positive and negative) to determine associations between Aβ levels and cognitive function [3,4,7,27]. However, we recently showed that when SUVR distributions are further divided into Aβ+ and Aβ++, additional prognostic information can be derived. Specifically, compared with Aβ+ CN older adults, Aβ++ CN older adults have an increased risk of progression to clinically classified MCI or AD, as well as an increased rate of memory decline over a 36-month period [2,5]. In accordance with our previous observations, results of the present study suggest that the additional grouping of

SUVR distributions into two categories of abnormality can improve understanding of the rate of Aβ-related cognitive decline in CN older adults.

The second hypothesis that composite scores that did not include the MMSE would be more sensitive in detecting Aβ+-related cognitive decline in Aβ+ older adults was also supported. First for Aβ++ CN older adults, the ADCS-PACC provided the lowest sensitivity to Aβ-related change ($d = -0.64$ in CN Aβ++). Removing the MMSE from the ADCS-PACC (i.e., the ADCS-PACC no MMSE) increased this estimate of sensitivity substantially ($d = -1.01$ in CN Aβ++). Adding a test of verbal fluency to that for the CVLT-II, LM-DR, and DSST (i.e., ZAVEN) resulted in similar sensitivity to that observed for the ADCS-PACC no MMSE ($d = -1.07$ in CN Aβ++). Furthermore, statistically significant differences between Aβ− CN older adults and Aβ+ CN older adults were observed for all composites that did not include the MMSE, that is the ADCS-PACC no MMSE ($d = 0.43$), ZAVEN ($d = 0.50$), and EM composites ($d = 0.53$), but not the ADCS-PACC ($d = 0.26$). These results are consistent with our previous reports that Aβ++ in CN older adults is associated with increased rates of decline in cognitive function [5]. However, although we have reported that Aβ+ was not associated with decline in cognitive function over a 36-month period in CN older adults, the current results indicate that over a 72-month period, Aβ+ is associated with significant cognitive decline of a moderate magnitude in this population, albeit detectable only with composites that do not include the MMSE.

The ADCS-PACC is an empirically derived cognitive outcome measure developed for use in clinical trials of novel anti-Aβ therapies in preclinical AD [18]. One notable characteristic of the ADCS-PACC was that it emphasized the cognitive domains that should be included as opposed to specifying which cognitive tests should operationalize each domain. For example, in the validation study, ADCS-PACC composites from different studies were based on three different word list learning tasks (i.e., FCSRT, CVLT-II, and the delayed word recall task from the ADAS-Cog) and two different paragraph recall tasks (i.e., LM-DR and New York University paragraphs). Despite this flexibility, we believe that there are two limitations to the ADCS-PACC. The first is the inclusion of the MMSE, and the second is the absence of a measure of executive function. Hence, in the present study, we determined whether replacing the MMSE with a measure of executive function (e.g., the verbal fluency FAS task used in this study) would increase the sensitivity of the cognitive composite to Aβ-related cognitive decline in CN older adults. When we replaced the MMSE with this measure of executive function, the sensitivity of this composite score to detecting Aβ-related cognitive decline in both the Aβ+ and Aβ++ CN groups increased beyond that observed for the ADCS-PACC.

There is now sufficient evidence of the deleterious effects of high Aβ burden in CN older adults such that clinical trials of anti-Aβ therapies designed to halt or slow the progression of AD are now being conducted in Aβ+ CN older adults [13,14,28]. These clinical trials have emphasized the utility of a composite measure of cognitive function as primary cognitive end points [18,19,28]. Although experimental models have consistently demonstrated that measures of EM or EM composites provide the most sensitive indicator of Aβ-related cognitive decline [3–5], others have suggested that a multidomain cognitive composite may provide even greater sensitivity to detecting Aβ-related cognitive decline [21]. The results of our present study show that the EM composite was equally sensitive as our multidomain cognitive composite (i.e., ZAVEN) in detecting cognitive decline associated with both Aβ+ and Aβ++ in CN adults, and performed better than the ADCS-PACC. However, given this equivalence, the ZAVEN or ADCS-PACC no MMSE might be preferable given that their coverage of cognitive domains other than memory might make any change detected more clinically relevant.

One assumption in the generation of the ZAVEN composite measure was that inclusion of the MMSE in the ADCS-PACC would limit its sensitivity to detecting Aβ-related cognitive decline. Interestingly, when the MMSE was combined with scores from the neuropsychological tests of EM and attention, no substantial skewness was observed and estimates of test-retest reliability remained high. However, it is notable that estimates of the SD associated with the mean rate of change on the ADCS-PACC for each Aβ subgroup was approximately twice that observed for estimates of the SDs for Aβ-related cognitive change on the EM and ZAVEN composite scores (Table 2). Similarly, the ADCS-PACC no MMSE yielded change scores with much lower SDs across all groups and this resulted in its improved sensitivity to Aβ-related cognitive decline (Table 2, Fig. 1).

The relatively lower sensitivity of the ADCS-PACC due to the inclusion of the MMSE demonstrates the importance of test selection in constructing composite scores as cognitive end points in clinical research studies and clinical trials in Aβ+ CN older adults, that is, composite scores should consist of measures of relevant cognitive domains that also have good metric properties. Hence, if a measure of general cognitive function was required in the ADCS-PACC, we would recommend the inclusion of an instrument with suitable metric properties in CN older adults, such as, the dementia rating scale-2 [29]. It is noteworthy that the cognitive composite score developed by the Alzheimer's Prevention Initiative (API) trial also includes a subtest of the MMSE (i.e., MMSE orientation to time subtest) [19]. On the basis of the results of the present study, this composite may also have some decreased sensitivity to detecting Aβ-related cognitive decline in CN older adults, compared with the

same composite without this MMSE item. Unfortunately, some of the tests nominated for inclusion in the API composite measure were not included in the AIBL study (e.g., East Boston Naming Test, Raven's Progressive Matrices; [19]), and so we were unable to construct an analogue in this data set. It is also important to emphasize that in the present study, we validated a composite where we removed the MMSE and replaced it with a test of executive function. We believe that it is important to challenge this composite with other measures of executive function and in other cohorts to determine whether the sensitivity of these measures can be replicated and even improved further.

In the present study, composite scores were computed by averaging standardized scores of each of the tests on which they were based. Thus, all tests were weighted equally in their contribution to the composite score. The method of averaging standardized scores was chosen because it had been shown to be informative by the statisticians responsible for the design of the A4 trial [18]. However, given the current finding that it is possible to improve the sensitivity of the ADCS-PACC to Aβ-related cognitive change by changing the tests on which it is based, future studies could advance the field further by also investigating whether improvements in the statistical methods for combining data from different tests could improve this sensitivity further. For example, statistical approaches that optimize the contribution or weight of data from individual neuropsychological tests into composite scores, such as latent variable modeling or item response theory, may increase further their sensitivity to Aβ-related change cognitive change [19,30]. Alternatively, such approaches may find a method for retaining clinically important tests such as the MMSE in composite scores despite their poor metric characteristics in the sample of interest. However, we note that Donohue et al. [18] reported that the use of optimized reweighting strategies did not improve the sensitivity of their ADCS-PACC. Thus, although other approaches for the identification of optimal composite scores should be investigated, the current findings do provide a good starting point for such challenges.

Notwithstanding these limitations, results of the present study along with previous studies suggest that Aβ-related cognitive decline is best detected using validated neuropsychological instruments with a particular emphasis on the measurement of episodic memory. The approach taken by the A4 trial in developing the ADCS-PACC is important, as it emphasizes the characterization of cognitive domains affected early in the disease stage rather than specifying particular neuropsychological tests that should be included [18]. This test-agnostic approach allows for the construction of similar cognitive composites and comparison and integration of data across various research groups despite the variation in neuropsychological tests used. Our approach to remove

the MMSE from the ADCS-PACC and replace it with a test of executive function (i.e., verbal fluency) rendered this composite measure more sensitive in detecting Aβ-related cognitive decline, including in CN older adults with different levels of high Aβ burden (i.e., Aβ+ CN vs. Aβ++ CN). Importantly, the current data should not be taken to diminish the importance of the MMSE for the identification of cognitive impairment in AD. Rather, they indicate only that the metric characteristics of this instrument limit its utility in understanding cognitive change in the preclinical stages of AD. On a related point, the data from the present study do not provide information about the diagnostic utility of any of the composite scores for cognitive impairment. Instead, the data indicate the extent to which they are sensitive to Aβ-related cognitive change in preclinical AD. Although the ZAVEN composite score includes measures of the main domains of cognitive function that are affected in preclinical AD (i.e., episodic memory, executive function and attention), it also pays tribute to a preeminent figure of AD research, Dr Zaven Khachaturian. We are hopeful that just as Dr Khachaturian has contributed substantially to the understanding of AD, so will researchers find the composite score we have named after him useful in understanding preclinical AD.

## RESEARCH IN CONTEXT

1. Systematic review: We conducted a review of recent efforts to derive new cognitive composite scores that are increasingly being relied on for large prospective clinical trials in preclinical and early Alzheimer's disease (AD). This study sought to compare one new composite score (ADCS-Preclinical Alzheimer Cognitive Composite [ADCS-PACC]) against both a single-domain composite score (episodic memory), and a more theoretically derived composite score that the authors propose (Z-scores of Attention, Verbal fluency and Episodic memory for Non-demented older adults [ZAVEN]), based on published literature that consistently supports the findings of early disease-related changes in the cognitive domains of episodic memory, attention, and executive function. All three composite scores were compared within the same larger prospective natural history study (Australian Imaging, Biomarkers, and Lifestyle).

2. Interpretation: The results of our study suggest that in cognitively normal (CN) older adults, β-amyloid (Aβ)–related decline in cognitive function is best detected using validated neuropsychological instruments, particularly of episodic memory. Removal of the mini-mental state examination (MMSE) from the ADCS-PACC and replacing it with a test of executive function (i.e., verbal fluency) rendered it more sensitive even to Aβ-related cognitive decline in CN older adults with only high Aβ burden (i.e., CN Aβ+).

3. Future directions: Additional prospective research should be aimed at further validating the ZAVEN composite score for use in both clinical research studies and clinical trials of preclinical AD. In particular, research groups that have developed cognitive composite scores that include the MMSE or subtests of the MMSE (e.g., the Alzheimer's Prevention Initiative trial) should challenge their composites by removing the MMSE and replacing it with a test of executive function.

## References

[1] Jack CR, Wiste HJ, Lesnick TG, Weigand SD, Knopman DS, Vemuri P, et al. Brain β-amyloid load approaches a plateau. Neurology 2013;80:1–7.

[2] Rowe CC, Bourgeat P, Ellis KA, Brown B, Lim YY, Mulligan R, et al., for the AIBL Research Group. Predicting Alzheimer disease with β-amyloid imaging: Results from the Australian imaging, biomarkers, and lifestyle study of ageing. Ann Neurol 2013;74:905–13.

[3] Sperling RA, Johnson KA, Doraiswamy PM, Reiman EM, Fleisher AS, Sabbagh MN, et al. Amyloid deposition detected with florbetapir F 18 ((18)F-AV-45) is related to lower episodic memory performance in clinically normal older individuals. Neurobiol Aging 2013;34:822–31.

[4] Doraiswamy PM, Sperling RA, Johnson K, Reiman EM, Wong TZ, Sabbagh MN, et al., for the AV45-A11 Study Group. Florbetapir F 18 amyloid PET and 36-month cognitive decline: A prospective multi-center study. Mol Psychiatry 2014;19:1044–51.

[5] Lim YY, Maruff P, Pietrzak RH, Ames D, Ellis KA, Harrington K, et al., for the AIBL Research Group. Effect of amyloid on memory and non-memory decline from preclinical to clinical Alzheimer's disease. Brain 2014;137:221–31.

[6] Lim YY, Villemagne VL, Laws SM, Pietrzak RH, Snyder PJ, Ames D, et al. APOE and BDNF polymorphisms moderate amyloid β-related cognitive decline in preclinical Alzheimer's disease. Mol Psychiatry 2015;20:1322–8.

[7] Mormino EC, Betensky RA, Hedden T, Schultz AP, Ward A, Huijbers W, et al. Amyloid and APOE E4 interact to influence short-term decline in preclinical Alzheimer's disease. Neurology 2014; 82:1760–7.

[8] Villemagne VL, Burnham S, Bourgeat P, Brown B, Ellis KA, Salvado O, et al., for the AIBL Research Group. Amyloid β deposition, neurodegeneration and cognitive decline in sporadic Alzheimer's disease: A prospective cohort study. Lancet Neurol 2013;12:357–67.

[9] Dore V, Villemagne VL, Bourgeat P, Fripp J, Acosta O, Chetelat G, et al. Cross-sectional and longitudinal analysis of the relationship between Aβ deposition, cortical thickness, and memory in cognitively unimpaired individuals and in Alzheimer's disease. JAMA Neurol 2013;70:903–11.

[10] Landau SM, Mintun MA, Joshi AD, Koeppe RA, Petersen RC, Aisen PS, et al., for the Alzheimer's Disease Neuroimaging Initiative. Amyloid deposition, hypometabolism, and longitudinal cognitive decline. Ann Neurol 2012;72:578–86.

[11] Jack CR, Lowe VJ, Senjem ML, Weigand SD, Kemp BJ, Shiung MM, et al. 11C PiB and structural MRI provide complementary information in imaging of Alzheimer's disease and amnestic mild cognitive impairment. Brain 2008;131:665–80.

[12] Rowe CC, Ellis KA, Rimajova M, Bourgeat P, Pike KE, Jones G, et al. Amyloid imaging results from the Australian Imaging, Biomarkers and Lifestyle (AIBL) study of aging. Neurobiol Aging 2010; 31:1275–83.

[13] Reiman EM, Langbaum JB, Fleisher AS, Caselli RJ, Chen K, Ayutyanont N, et al. Alzheimer's Prevention Initiative: A plan to accelerate the evaluation of presymptomatic treatments. J Alzheimers Dis 2011;26(Supp 3):321–9.

[14] Sperling RA, Jack CR, Aisen PS. Testing the right target and right drug at the right stage. Sci Transl Med 2011;3:111cm133.

[15] Lacy M, Kaemmerer T, Czipri S. Standardized mini-mental state examination scores and verbal memory performance at a memory center: Implications for cognitive screening. Am J Alzheimers Dis Other Demen 2015;30:145–52.

[16] Lim YY, Ellis KA, Harrington K, Pietrzak RH, Gale J, Ames D, et al., for the AIBL Research Group. Cognitive decline in adults with mild cognitive impairment and high Aβ amyloid: Prodromal Alzheimer's disease? J Alzheimers Dis 2013;33:1167–76.

[17] Spencer RJ, Wendell CR, Giggey PP, Katzel LI, Lefkowitz DM, Siegel EL, et al. Psychometric limitations of the mini-mental state examination among nondemented older adults: An evaluation of neurocognitive and magnetic resonance imaging correlates. Exp Aging Res 2013;39:382–97.

[18] Donohue MC, Sperling RA, Salmon DP, Rentz DM, Raman R, Thomas RG, et al. The preclinical Alzheimer cognitive composite: Measuring amyloid-related decline. JAMA Neurol 2014;71:961–70.

[19] Langbaum JB, Hendrix SB, Ayutyanont N, Chen K, Fleisher AS, Shah RC, et al. An empirically derived composite cognitive test score with improved power to track and evaluate treatments for preclinical Alzheimer's disease. Alzheimers Dement 2014;10:666–74.

[20] Kantarci K, Lowe VJ, Przybelski SA, Senjem ML, Weigand SD, Ivnik RJ, et al. Magnetic resonance spectroscopy, β-amyloid load, and cognition in a population-based sample of cognitively normal older adults. Neurology 2011;77:951–8.

[21] Kryscio RJ. Secondary prevention trials in Alzheimer disease: The challenge of identifying a meaningful end point. JAMA Neurol 2014;71:947–9.

[22] Insel PS, Mattsson N, Mackin RS, Kornak J, Nosheny R, Tosun-Turgut D, et al. Biomarkers and cognitive endpoints to optimize trials in Alzheimer's disease. Ann Clin Transl Neurol 2015;2:534–47.

[23] Ellis KA, Bush AI, Darby D, De Fazio D, Foster J, Hudson P, et al., for the AIBL Research Group. The Australian Imaging, Biomarkers and Lifestyle (AIBL) study of aging: Methodology and baseline characteristics of 1112 individuals recruited for a longitudinal study of Alzheimer's disease. Int Psychogeriatr 2009;21:672–87.

[24] Clark CM, Schneider JA, Bedell BJ, Beach TG, Bilker WB, Mintun MA, et al., for the AV45-A07 Study Group. Use of florbetapir-PET for imaging beta-amyloid pathology. JAMA 2011;305:275–83.

[25] Vandenberghe R, Van Laere K, Ivanoiu A, Salmon E, Bastin C, Triau E, et al. 18F-flutemetamol amyloid imaging in Alzheimer disease and mild cognitive impairment: A phase 2 trial. Ann Neurol 2010;68:319–29.

[26] Jedynak BM, Lang A, Liu B, Katz E, Zhang Y, Wyman BT, et al., for the Alzheimer's Disease Neuroimaging Initiative. A computational neurodegenerative disease progression score: Method and results with the Alzheimer's Disease Neuroimaging Initiative cohort. Neuroimage 2012;63:1478–86.

[27] Lim YY, Ellis KA, Pietrzak RH, Ames D, Darby D, Harrington K, et al., for the AIBL Research Group. Stronger effect of amyloid load than APOE genotype on cognitive decline in healthy older adults. Neurology 2012;79:1645–52.

[28] Mills SM, Mallmann J, Santacruz AM, Fuqua A, Carril M, Aisen PS, et al. Preclinical trials in autosomal dominant AD: Implementation of the DIAN-TU trial. Rev Neurol 2013;169:737–43.

[29] Jurica PJ, Leitten CL, Mattis S. DRS-2 dementia rating scale-2: professional manual. Odessa, FL: Psychological Assessment Resources; 2004.

[30] Crane PK, Narasimhalu K, Gibbons LE, Mungas DM, Haneuse S, Larson EB, et al. Item response theory facilitated cocalibrating cognitive tests and reduced bias in estimated rates of decline. J Clin Epidemiol 2008;61:1018–27.