An empirically derived composite cognitive test score with improved power to track and evaluate treatments for preclinical Alzheimer’s disease

Jessica B.S. Langbaum¹,¹¹,* Suzanne B. Hendrix², Napatkamon Ayutyanont¹,¹¹, Kewei Chen¹,⁴,¹¹, Adam S. Fleisher¹,¹¹, Raj C. Shah⁶,⁷, Lisa L. Barnes⁶,⁸,⁹, David A. Bennett⁶,⁸, Pierre N. Tariot¹,³,¹¹, and Eric M. Reiman¹,³,¹⁰,¹¹

¹Banner Alzheimer’s Institute, Phoenix, AZ, USA
²Pentara Corporation, Salt Lake City, UT, USA
³Department of Psychiatry, University of Arizona, Tucson, AZ, USA
⁴Department of Mathematics and Statistics, Arizona State University, Tempe, AZ, USA
⁵Rush Alzheimer’s Disease Center, Rush University Medical Center, Chicago, IL, USA
⁶Rush Alzheimer’s Disease Center, Rush University Medical Center, Chicago, IL, USA
⁷Department of Family Medicine, Rush University Medical Center, Chicago, IL, USA
⁸Department of Neurological Sciences, Rush University Medical Center, Chicago, IL, USA
⁹Department of Behavioral Sciences, Rush University Medical Center, Chicago, IL, USA
¹⁰Neurogenomics Division, Translational Genomics Research Institute, Phoenix, AZ, USA
¹¹Arizona Alzheimer’s Consortium, Phoenix, AZ, USA

Abstract

Background—There is growing interest in the evaluation of preclinical Alzheimer’s disease (AD) treatments. As a result, there is a need to identify a cognitive composite that is sensitive to tracking preclinical AD decline to be used as a primary endpoint in treatment trials.

Methods—Longitudinal data from initially cognitively normal, 70–85 year old participants in three cohort studies of aging and dementia from the Rush Alzheimer’s Disease Center were examined to empirically define a composite cognitive endpoint that is sensitive to detecting and tracking cognitive decline prior to the onset of cognitive impairment. The mean to standard deviation ratios (MSDR) of change over time were calculated in a search for the optimal combination of cognitive tests/sub-tests drawn from the neuropsychological battery in cognitively...
normal participants who subsequently progressed to clinical stages of AD during a two and five year period, using data from those who remained unimpaired during the same time period to correct for aging and practice effects. Combinations that performed well were then evaluated for representation of relevant cognitive domains, robustness across individual years prior to diagnosis, and occurrence of selected items within top performing combinations.

**Results**—The optimal composite cognitive test score is comprised of 7 cognitive tests/sub-tests with an MSDR=0.964. By comparison, the most sensitive individual test score, Logical Memory – Delayed Recall, MSDR= 0.64.

**Conclusions**—We have identified a composite cognitive test score representing multiple cognitive domains that has improved power compared to the most sensitive single test item to track preclinical AD decline and evaluate preclinical AD treatments. We are confirming the power of the composite in independent cohorts, and with other analytical approaches, which may result in refinements, and have designated it as the primary endpoint in the Alzheimer’s Prevention Initiative’s preclinical treatment trials for individuals at high imminent risk for developing symptoms due to late-onset AD.

---

**Introduction**

Without an effective treatment that postpones the onset or completely prevents the clinical consequences of Alzheimer’s disease (AD), the number of individuals afflicted by the disease will continue to rapidly increase (1;2). There is growing interest in the hypothesis that interventions may have their most profound effect if initiated in the preclinical AD phase (3), that is, in the absence of mild cognitive impairment (MCI) or AD dementia (4). Several such trials are underway or are in various planning stages, including those with the strategy of testing therapies in people who are at highest imminent risk of developing MCI or AD dementia due to factors such as age and genetic disposition or presence of biomarker evidence of AD (4–8). Traditional clinical outcomes, such as progression to clinical diagnosis, or cognitive outcomes developed for studies in MCI or AD dementia may not be well-suited for some preclinical treatment trials due to large sample size and long trial duration requirements or the psychometric properties of the tests themselves (9–12). Moreover, individually examining each cognitive assessment and treating as individual outcomes inflates Type 1 error if appropriate corrections are not made to guard against multiple comparisons. Use of an appropriate composite reduces the number of variables employed and thus risk of Type 1 error, it can be empirically derived and its sensitivity to detecting and tracking preclinical AD can be validated in multiple datasets. As a result, it affords a measure of multiple domains that can serve as a primary endpoint in preclinical treatment trials (13).

Small, but measurable cognitive decline occurs during preclinical AD. For instance, retrospective and prospective studies of cognitively healthy individuals who eventually progressed to AD dementia have shown episodic memory decline to be a defining feature of preclinical AD (14–18). In addition, decline in other cognitive domains, such as executive (19), visual spatial (16) and global cognitive functioning (16;20) occurs during the transition from normal aging to preclinical AD and into the clinical stages of AD. Studies of cognitively healthy individuals with significant fibrillar amyloid burden report decline

*Alzheimers Dement.* Author manuscript; available in PMC 2015 November 01.
primarily in episodic memory, executive function and language (21–25). Long-term recall memory performance has been found to begin to decline in relationship to apolipoprotein E (APOE) ε4 gene dose, reflecting three levels of genetic risk for late-onset AD, despite maintenance of normal clinical status (26).

There are multiple approaches for selecting an appropriate cognitive endpoint for use in preclinical AD studies and therapeutic trials. For instance, a theoretically driven approach reasons that a composite should be constructed a priori from cognitive assessments known to decline during preclinical AD. A related approach is to construct composites specific to an individual cognitive domain, such as memory (27) or executive functioning (28). Yet another is an empirically driven approach, in which the endpoint or composite is selected based on analyses demonstrating sensitivity (e.g., has the greatest power) to detect and track the outcome of interest, such as preclinical AD decline. These approaches are not necessarily mutually exclusive; for instance, theoretical knowledge of preclinical AD can be taken into account when empirically deriving a composite cognitive test score. Several different analyses methods are available to developing composites, including but not limited to latent variable analyses or partial least squares regression (29–31), principal components (32), item-response theory (33), Rasch Measurement Theory (34) or item-level analysis (35). While there have been some efforts focused on refining existing cognitive assessments, this may be best suited for MCI and early AD trials (36).

Here we propose a strategy to empirically determine the combination of cognitive assessments most sensitive to the tracking of preclinical AD in individuals who subsequently progress to MCI or probable AD dementia, while controlling for practice and normal aging effects using data from individuals who did not progress to the clinical stages of AD over the same duration. The goal of the present study was to develop a composite with optimal sensitivity to decline, not limited to a single cognitive domain, corresponding to a change from baseline analysis. This approach differs from optimizing an endpoint for discriminating those who progress from those who remain stable, which would result in a composite that could be used as a progression endpoint in preclinical treatment trials. We hypothesize that the composite will be more sensitive (i.e., have greater power) to detecting and tracking preclinical AD decline compared to the most sensitive individual cognitive test/sub-test score given that the approach allows for the addition of assessments that improve sensitivity overall, despite perhaps being less sensitive individually to preclinical AD decline. Longitudinal data from three cohort studies of aging and dementia at the Rush Alzheimer’s Disease Center in those who did and did not clinically progress over a two and five year period were used to develop a composite cognitive endpoint, employing the mean to standard deviation ratio (MSDR) of the change score as the measure of sensitivity to preclinical AD decline over time (31). The results from the present study are informing the design of trials for the Alzheimer’s Prevention Initiative (API) focused on individuals at high imminent risk for symptoms of late-onset AD based on their age and genetics.
**Materials and Methods**

**Participants**

Data from participants enrolled in the Rush Alzheimer’s Disease Center’s Religious Orders Study [ROS], Memory and Aging Project [MAP], or the Minority Aging Research Study [MARS] was downloaded on June 7, 2010. Enrollment criteria for the three studies are quite similar and have been previously reported (37–39). Briefly, participants from each study are older adults without dementia at the time of enrollment and who agree to annual clinical and neuropsychological evaluations, and for those in ROS and MAP, agree to brain donation at the time of death. For the present study, longitudinal data from participants who subsequently progressed to either MCI or AD dementia (possible or probable) and who had at least 1 follow-up visit (n = 1073), 2 years of data (n = 528) or 5 years of data (n = 213) and from participants who remained cognitively normal during the same time period and had at least 2 (n = 831) or 5 years (n = 413) of follow-up data were included in the analyses (Table 1). The studies were approved by the Institutional Review Board of Rush University Medical Center and each participant signed an informed consent.

**Cognitive and Clinical Evaluations**

The three studies included uniform and structured annual clinical evaluations with medical history questions, neurological examination, and detailed cognitive testing using a battery of 19–21 neuropsychological tests (40;41). Diagnostic classification followed a multi-step procedure as previously described (42;43). Briefly, neuropsychological tests encompassing a wide range of cognitive functions were administered by trained technicians and scores were adjusted for education using an automated scoring algorithm computed in SAS (43;44). Participants were examined or records were reviewed by a clinician (primarily neurologists or geriatricians, supplemented by advanced practice geriatric nurses and neuropsychologists) and diagnostically classified using the recommendations of the joint working group of the National Institute of Neurologic and Communicative Disorders and Stroke and the Alzheimer’s Disease and Related Disorders Association (NINCDS/ADRDA) (45). The diagnosis of dementia and probable AD dementia followed a three step procedure and was validated pathologically (43;44;46). History of cognitive decline was determined by structured interview, evidence of impairment in memory and other cognitive abilities was based on neuropsychological performance tests summarized by an experienced neuropsychologist. Dementia due to AD and other causes was based on review of records and an interview and examination by a clinician with expertise in dementia evaluations. MCI was based on the presence of cognitive impairment as determined by the neuropsychologist and the absence of dementia as determined by the clinician (43;47;48).

**Data Analysis**

Data from tests were included in the following analyses only if all cohorts (ROS, MAP and MARS) received the particular assessment. For example, only MAP and MARS participants had Stroop Color Naming test data; as a result, this assessment was not included in the analyses. The scores for each test, T, were standardized to a range of 0.0 to 1.0 as shown in the following equation:
where $S_{j,i}^T$ is the score at time $t_i^j$ ($i=1, 2$) for subject $j$, and $T_{scr}$ symbolizes all the possible scores this test $T$ can take.

The annualized difference score $y_j^T$ for subject $j$, test $T$ is defined as

$$y_j^T = \frac{x_j^T - x_j^{T,1}}{t_2^j - t_1^j}$$

For assessments without a defined maximum score (such as category fluency), a maximum threshold was established that was 2 standard deviations above the mean. All analyses were conducted using SAS v. 9.2 (SAS Institute Inc, Cary, NC, USA).

Cognitive domains that were expected to change early in the disease process were identified prior to analyses. Sensitivity of each of the cognitive tests/sub-tests and correlations between them were examined, with the intention of identifying items to represent the relevant cognitive domains in a composite. Due to the complexity of constructing a multivariate composite based on univariate or bivariate summary statistics, an integrated approach was used to evaluate all possible combinations of items to optimize the sensitivity resulting in an analysis that is mathematically closely related to principal component analysis (43) as follows:

$$X = \sum_{T=1}^{K} w_T y_T^T$$

where $y_T^T = (y_1^T, y_2^T, \ldots, y_N^T)$, $N$ is the number of subjects, $K$ is the number of tests, $w_T$ is the weight for test $T$, and we have $w_T \geq 0$ ($T=1, \ldots, K$) and $\Sigma w_T = 1$.

The optimization criterion is the maximization of MSDR of $X$ with regards to weights of 0 or 1 for each item for the exclusion or inclusion of that particular item:

$$MSDR = \frac{\bar{X}}{\text{std}(X)}$$

where $\bar{X}$ is the mean of $X$ and std($X$) is its standard deviation.

Maximizing the MSDR across all combinations of items is not a statistical inference procedure; rather, it is a method that produces a metric that can guide our search for combinations that are sensitive. Those participants who eventually progressed to cognitive impairment (MCI or probable dementia AD), were assessed with this methodology, looking backwards zero to two ($n = 528$) or five years ($n = 213$) prior to diagnosis. To most accurately reflect an API preclinical treatment trial, a person was considered diagnosed when their clinical diagnosis first progressed from “no cognitive impairment” to either MCI or AD. In order to account for longitudinal aging and practice effects, which impact the sensitivity of outcome measures in this preclinical population (49;50), annualized MSDRs
were also calculated for those who remained cognitively normal during the two (n = 831) or five-year (n = 413) follow-up period. Adjusted MSDRs were calculated by subtracting the mean for the controls from that obtained from individuals who progressed to MCI or dementia prior to dividing by the standard deviation.

Results from these analyses were used as one way to assess the combinations and determine a “best” composite. Items that were consistently represented in the combinations with the highest sensitivity and that also demonstrated consistency within separate years of the 1–5 year time period were identified as robust items for measuring change. Construct validity was assessed by giving preference to combinations that represented multiple cognitive domains known to be important that also had consistent sensitivity across the two and five years of decline. Corresponding sample size estimates were calculated using the adjusted MSDRs, though it should be noted that these estimates are only to aid in gauging the comparable sensitivity of the tests and should not be directly used for powering a trial given that the estimates were calculated based on data from individuals who all subsequently progressed to MCI or AD dementia, and it would be impossible to enroll a similar population into a trial.

Following selection of the composite, we sought to determine whether the corresponding MSDR could be substantially increased (i.e., improved, resulting in smaller sample sizes) by weighting the individual assessments included in the composite (as opposed to the equal weights used in the initial analyses). A search of potential weighting combination, such that the sum of the weighting of the individual assessments equaled one, was conducted in those who progressed to cognitive impairment (MCI or AD) at two and five years prior to diagnosis.

Results

Participant Characteristics

There were some demographic differences between participants who progressed to clinical stages of AD and those who remained cognitively healthy during the same time period (Table 1). For instance, in the MSDR analyses of the two years and five years prior to diagnosis, those who progressed to clinical stages of AD were older (p < 0.0001) and had different percent distribution between the racial categories (two-year analysis p < 0.0001; five-year analysis p = 0.01).

Individual Cognitive Assessment Properties

The sensitivity of the individual cognitive assessments over the five years prior to a diagnosis of MCI or probable AD dementia, unadjusted for normal aging effects, is shown in Table 2. Among individuals who progressed to clinical stages of AD, Category Fluency (fruits and vegetables) had the highest sensitivity of all tests (MSDR 0.825), followed by Symbol Digit Modalities (0.71), and MMSE Total (0.665). Adjusting for longitudinal aging and practice effects using data from those participants who remained cognitively normal during the same time period resulted in an increased MSDR for some cognitive assessments (e.g., Logical Memory Delayed Recall adjusted MSDR = 0.64) due to observed increases in
normals, while others decreased (e.g., Symbol Digit Modalities adjusted MSDR = 0.385) due to observed worsening in normals (Table 2).

**Deriving the Alzheimer’s Prevention Initiative (API) Composite Cognitive Test Score**

Results from the MSDR calculation for every possible combination of neuropsychological assessments indicate that the composite most sensitive to detecting preclinical cognitive decline related to AD, adjusting for longitudinal aging and practice effects, that has construct validity and also includes items that are robust across individual years consisted of: Category Fluency (fruits and vegetables), Boston Naming Test (15 item), Logical Memory – Delayed Recall, East Boston Naming Test – Immediate Recall, Ravens Progressive Matrices Subset (9 items), Symbol Digit Modalities, and MMSE Orientation to Time (Table 2). Based on the data five years prior to diagnosis, the total five-year MSDR of the composite is 0.9639. The best 50 combinations had annual MSDRs that ranged from 0.1928–0.1862 and were comprised of 6–7 test/sub-tests of episodic memory, working memory, language, global functioning, and visual spatial ability.

A composite with an MSDR of 0.9639 requires an estimated 264 completers per treatment arm to detect a 25% treatment effect in a five-year trial, noting that caveats mentioned previously of applying this sample size estimate when designing a prevention trial still apply. In comparison, the most sensitive individual cognitive assessment is Logical Memory Delayed Recall, with a five-year MSDR of 0.640 and an estimated sample size of 611 completers per treatment arm to detect a 25% treatment effect, making the API composite cognitive test score considerably more sensitive to tracking preclinical AD decline (Figure 1). Based on the data two years prior to diagnosis, a shorter study with the same composite cognitive test score would result in a lower total two-year MSDR of 0.3398. Results from analyses based on the search of potential weighting combinations revealed that weighting provided minimal improvement over the unweighted MSDRs (increase in MSDR < 5%).

**Discussion**

We empirically identified an API composite cognitive test score sensitive to preclinical AD decline, and suggest that it can be used in preclinical trials to evaluate treatment effects with smaller sample sizes and improved statistical power compared to the most sensitive individual cognitive assessments and larger test batteries, and in a manner that is reasonably likely to predict a treatment’s effect on clinical progression to MCI or AD. This API composite cognitive test score and the approach taken to develop it appears to fit into the framework provided by the Food and Drug Administration’s recent draft guidance concerning a cognitive assessment serving as a primary efficacy measure in preclinical AD trials (51). Moreover, the composite identified in the present study was comprised of the same cognitive domains/assessments with the exception of one (present study included a test of visual spatial ability – Symbol Digits Modalities) as a composite cognitive endpoint identified in cognitively healthy ADAD mutation carriers, despite substantial differences in the cohorts’ neuropsychological test batteries (52). This consistency is noteworthy as confirmation of the composite’s performance in an independent population, as well as suggesting that there is extensive overlap in the pattern of cognitive decline between the two...
forms of the disease even though they strike at different ages, and may have different preclinical and clinical time courses, underlying etiologies, and biological processes.

We employed an empirical strategy refined by theoretical understanding of preclinical AD to develop a sensitive composite cognitive endpoint, controlling for aging and practice effects, by examining longitudinal data in the two and five years prior to clinical progression. With this approach, we focused primarily on the aspects of the disease that decline consistently across individuals in order to assess effectiveness of a treatment in slowing decline in a preclinical trial, rather than discrimination between those who progress and those that do not, or the neuropathological underpinnings of AD that result in a change in cognitive functioning. This approach has the added advantage in that it incorporates data from participants at various points along the preclinical AD continuum and does not presuppose the cognitive assessments sensitive to detecting and tracking this decline. Just as in a clinical trial, some participants may progress to cognitive impairment within months while others are several years away.

The optimal composite cognitive test score identified in the current study incorporates cognitive assessments from several different domains, complementing recent studies suggesting that multiple cognitive domains decline in preclinical AD (16;53), not just decline in episodic memory, though that remains a defining feature of preclinical AD (14–18). With the exception of the Ravens Progressive Matrices, all of the assessments that comprise the composite do have a language component. Although on an individual basis, the neuropsychological assessments have varying levels of sensitivity to tracking preclinical decline, as measured by their MSDR, this empirical strategy, focused on the years prior to diagnosis (while controlling for longitudinal aging and practice effects), identified these items as providing a sensitive combination of assessments across multiple cognitive domains. More sensitive cognitive tests/sub-tests may not be included in the composite endpoint, since these items may correlate with another assessment that captures the same information and has a higher MSDR (32). The cognitive tests/sub-tests that are included and have a smaller MSDR may measure variability not captured by other assessments. As a result, the optimal composite was not comprised solely of the cognitive tests with the largest MSDRs.

There are some limitations to the present study. Development of the optimal composite cognitive endpoint was constrained by the neuropsychological test battery used in the Rush cohort studies. That said, we achieved remarkably similar results with independent efforts with an empirically driven approach to derive a composite cognitive endpoint for ADAD mutation carriers despite differences in the cohorts’ neuropsychological assessment battery (52). Likewise, scientists preparing for the Alzheimer’s Disease Cooperative Study (ADCS) “A4” trial for cognitively healthy individuals with presence of fibrillar amyloid burden have undertaken a similar effort using datasets from other cohorts and have produced results comparable to those reported here (6). The sample size and power estimates reported herein serve as a guide in comparing sensitivity and power of the various measures and are not directly applicable to future trials given that they were calculated in a sample of individuals all of whom progressed to MCI or dementia. We acknowledge that there are multiple analytical methods for deriving composites and there are limitations to the approach used,
including but not limited to the assumption of a simple model structure and that all variables are measured without error). Additional efforts under the auspices of the API are underway using other analysis methods, such as partial least squares, and will be reported separately. It should be noted that this study capitalized on longitudinal data from a cohort of cognitively unimpaired research participants who subsequently progressed to the clinical stages of AD in an effort to provide a more sensitive measure of the cognitive decline associated with AD and a cognitive endpoint could possibly qualify for use in preclinical AD trials. Since antemortem brain imaging or CSF biomarker measurements were not available in the Rush cohorts, additional studies may be helpful in clarifying the composite cognitive test score’s power to track declines and evaluate preclinical AD treatments in cognitively unimpaired subjects who meet the recently proposed research criteria for preclinical AD. We are also confirming the generalizability and power of the composite cognitive endpoint in other cohorts followed to progression of cognitive impairment, and to estimate the statistical power in different at-risk groups (e.g., APOE ε4 homozygotes or heterozygotes, different age groups, older adults with or without biomarker evidence of AD). Indeed, it may be that certain cognitive assessments or combinations of assessments are more sensitive to cognitive decline depending on the time frame prior to (or following) diagnosis. Similarly, there may be issues of generalizability to other populations. For instance, the Rush cohorts are not population-based samples, and all participants in these cohort studies are expected to undergo annual assessments and some are required consent to brain donation upon death. That said, given expectations of participants in clinical trials (e.g., frequent in-person visits with cognitive testing, biomarker assessments), the composite developed using data from the Rush cohorts is likely to be sensitive to tracking preclinical decline in individuals who would enroll in a preclinical AD trial. However, it remains unknown whether the composite is sensitive to detecting a treatment effect.

In summary, we examined longitudinal cognitive data from three well-characterized cohorts and conducted a search of every combination of cognitive assessments to identify the optimal combination that is sensitive to tracking preclinical AD decline in the two and five years prior to diagnosis of MCI or dementia, while controlling for aging and practice effects in individuals who remained cognitively normal during the same time interval. This combination was also selected to include robust items and to represent domains important in early disease. The approach allowed us to empirically identify an API composite cognitive endpoint that consists of multiple cognitive domains and is sensitive to preclinical cognitive decline associated with AD. A similar composite cognitive endpoint with even greater statistical power was independently derived in cognitively normal, ADAD mutation carriers and is being used as the primary endpoint in the first API preclinical treatment trial set to begin in 2013. As a result of these efforts, other preclinical trial investigators have extended the API composite cognitive test score development strategy for use in their planned trials and studies.

Acknowledgments

This work was supported by the National Institute on Aging (R01AG031581 and P30AG19610 to EMR; P30AG10161, R01AG15819, and R01AG17917 to DAB; and R01AG22018 to LLB), the state of Arizona, and contributions from the Banner Alzheimer’s Foundation and the Nomis Foundation. We thank Dr. Richard Caselli, Laura Jakimovich, Carolyn Langlois, and Stephanie Stanworth for their assistance.
Reference List


52. Ayutyanont, N.; Langbaum, JBS.; Hendrix, SB.; Chen, K.; Fleisher, AS.; Frieseahn, M.; Ward, M.; Aguirre, C.; Acosta-Baena, N.; Madrigal, L.; Munoz, C.; Tirado, V.; Moreno, S.; Tariot, PN.; Lopera, F.; Reiman, EM. The Alzheimer’s Prevention Initiative composite cognitive test score: sample size estimates for the evaluation of preclinical Alzheimer’s disease treatments in presenilin 1 E280A mutation carriers. submitted


Research In Context

Systematic Review

A PubMed search was conducted to identify relevant studies from 1980-present examining cognitive decline in preclinical AD, as well as methodological approaches for development of composite endpoints.

Interpretation

The analyses identified a composite cognitive test score with optimal statistical power to track preclinical AD decline compared to the most sensitive individual neuropsychological test score. The composite is similar to one independently identified for tracking preclinical decline in autosomal dominant AD and is well-suited for use as a primary endpoint in AD prevention treatment trials.

Future Directions

From a regulatory perspective it may be important for the field to reach consensus regarding the optimal approaches for identification and use of composite endpoints, particularly given the growing interest in preclinical AD treatment trials.
Figure 1.
Estimated sample size per group required to detect a 25% or 50% treatment effect with two-tailed $p = 0.05$ and 80% power in a randomized, placebo-controlled trial.
## Table 1

Demographic information for those who progressed to MCI or dementia (converters) compared to those who remained cognitively healthy (non-converters) in the two and five year MSDR analyses

<table>
<thead>
<tr>
<th></th>
<th>2-year MSDR analysis</th>
<th></th>
<th>5-year MSDR analysis</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Non-Converter (n = 831)</td>
<td>Converter (n = 528)</td>
<td>Test statistic, p-value</td>
<td>Non-Converter (n = 413)</td>
</tr>
<tr>
<td>Age, years</td>
<td>74.1 (7.1)</td>
<td>77.6 (6.9)</td>
<td>t (1357) = 8.9, p &lt;0.0001</td>
<td>73.3 (6.4)</td>
</tr>
<tr>
<td>Sex, %</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>72.8</td>
<td>71.0</td>
<td>χ² (1) = 0.5, p = 0.48</td>
<td>72.6</td>
</tr>
<tr>
<td>Male</td>
<td>27.2</td>
<td>29.0</td>
<td>27.4</td>
<td>33.3</td>
</tr>
<tr>
<td>Educational level, years</td>
<td>16.2 (3.8)</td>
<td>16.5 (3.7)</td>
<td>t (1357) = 1.8, p = 0.07</td>
<td>16.8 (3.7)</td>
</tr>
<tr>
<td>APOE ε4 carrier status, %</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ε4 carriers</td>
<td>17.0</td>
<td>19.5</td>
<td>χ² (1) = 0.4, p =0.53</td>
<td>16.9</td>
</tr>
<tr>
<td>ε4 n on-carriers</td>
<td>69.7</td>
<td>73.1</td>
<td>72.9</td>
<td>74.2</td>
</tr>
<tr>
<td>Missing</td>
<td>13.3</td>
<td>7.4</td>
<td>10.2</td>
<td>4.2</td>
</tr>
<tr>
<td>Race, %</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Caucasian</td>
<td>78.3</td>
<td>89.4</td>
<td>χ² (3) = 28.8, p &lt;0.0001</td>
<td>93.2</td>
</tr>
<tr>
<td>Black, African American</td>
<td>21.2</td>
<td>10</td>
<td>6.3</td>
<td>1.4</td>
</tr>
<tr>
<td>Native American, Indian</td>
<td>0.2</td>
<td>0.2</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Asian or Pacific Islander</td>
<td>0.2</td>
<td>0.4</td>
<td>0.5</td>
<td>0</td>
</tr>
<tr>
<td>Missing</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Values are mean(standard deviation) or percentage. p-values were calculated using t-test to compare participant groups for age and educational level, and with chi-square tests to compare the groups for sex, APOE ε4 carrier status and race.
### Table 2

Five-Year MSDRs for Individual Cognitive Assessment Test Items Considered for the API Composite Cognitive Test Score

<table>
<thead>
<tr>
<th>Cognitive assessment</th>
<th>Domain</th>
<th>Unadjusted MSDR</th>
<th>Adjusted MSDR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Boston Naming Test (15 item) *</td>
<td>Language / Semantic memory</td>
<td>0.36</td>
<td>0.305</td>
</tr>
<tr>
<td>Category fluency – Animals</td>
<td>Language / Semantic memory</td>
<td>0.645</td>
<td>0.44</td>
</tr>
<tr>
<td>Category fluency – Fruits/Vegetables *</td>
<td>Language / Semantic memory</td>
<td>0.825</td>
<td>0.61</td>
</tr>
<tr>
<td>CERAD Word list recall (Immediate)</td>
<td>Episodic memory</td>
<td>0.545</td>
<td>0.51</td>
</tr>
<tr>
<td>CERAD Word list memory (Delayed recall)</td>
<td>Episodic memory</td>
<td>0.635</td>
<td>0.52</td>
</tr>
<tr>
<td>CERAD Word List Recognition</td>
<td>Auditory memory</td>
<td>0.415</td>
<td>0.4</td>
</tr>
<tr>
<td>Complex Ideational Material</td>
<td>Auditory comprehension</td>
<td>0.3</td>
<td>0.285</td>
</tr>
<tr>
<td>Digit Ordering</td>
<td>Working memory</td>
<td>0.39</td>
<td>0.24</td>
</tr>
<tr>
<td>Digit Span - forward</td>
<td>Working memory</td>
<td>0.35</td>
<td>0.175</td>
</tr>
<tr>
<td>Digit Span - backward</td>
<td>Working memory</td>
<td>0.42</td>
<td>0.175</td>
</tr>
<tr>
<td>East Boston Naming Test, Immediate recall (Memory I) *</td>
<td>Episodic memory</td>
<td>0.44</td>
<td>0.485</td>
</tr>
<tr>
<td>East Boston Naming Test, Delayed recall (Memory II)</td>
<td>Episodic memory</td>
<td>0.54</td>
<td>0.435</td>
</tr>
<tr>
<td>Judgment of Line Orientation</td>
<td>Visuospatial</td>
<td>0.375</td>
<td>0.295</td>
</tr>
<tr>
<td>Logical Memory Ia (Immediate)</td>
<td>Episodic memory</td>
<td>0.405</td>
<td>0.55</td>
</tr>
<tr>
<td>Logical Memory IIa (Delayed) *</td>
<td>Episodic memory</td>
<td>0.455</td>
<td>0.64</td>
</tr>
<tr>
<td>MMSE – Orientation to Time *</td>
<td>Orientation</td>
<td>0.555</td>
<td>0.465</td>
</tr>
<tr>
<td>MMSE – Orientation to Place</td>
<td>Orientation</td>
<td>0.465</td>
<td>0.44</td>
</tr>
<tr>
<td>MMSE – Registration</td>
<td>Working memory</td>
<td>0.2</td>
<td>0.145</td>
</tr>
<tr>
<td>MMSE – Attention and Concentration</td>
<td>Attention and Concentration</td>
<td>0.225</td>
<td>0.185</td>
</tr>
<tr>
<td>MMSE – Recall</td>
<td>Episodic memory</td>
<td>0.145</td>
<td>0.11</td>
</tr>
<tr>
<td>MMSE – Language</td>
<td>Language</td>
<td>0.17</td>
<td>0.04</td>
</tr>
<tr>
<td>National Adult Reading Test (NART) – 10 items</td>
<td>General / Global Cognition</td>
<td>0.155</td>
<td>0.17</td>
</tr>
<tr>
<td>Number Comparison Test</td>
<td>Perceptual speed</td>
<td>0.56</td>
<td>0.355</td>
</tr>
<tr>
<td>Ravens Progressive Matrices – 16 items</td>
<td>Visuospatial / working memory</td>
<td>0.54</td>
<td>0.385</td>
</tr>
<tr>
<td>Ravens Progressive Matrices Subset – 9 items *</td>
<td>Visuospatial / working memory</td>
<td>0.5</td>
<td>0.43</td>
</tr>
<tr>
<td>Symbol Digit Modalities *</td>
<td>Perceptual speed</td>
<td>0.71</td>
<td>0.385</td>
</tr>
<tr>
<td>Wide Range Achievement Test (WRAT) – 15 items</td>
<td>General / Global Cognition</td>
<td>0.17</td>
<td>0.12</td>
</tr>
</tbody>
</table>

CERAD = Consortium to Establish a Registry for Alzheimer’s Disease

* Item included in the composite cognitive test score