See Commentary on page 661.

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

Napatkamon Ayutyanont, PhD; Jessica B. S. Langbaum, PhD; Suzanne B. Hendrix, PhD; Kewei Chen, PhD; Adam S. Fleisher, MD; Michel Friesenhahn, MA; Michael Ward, PhD; Camilo Aguirre, MS; Natalia Acosta-Baena, MD; Lucìa Madrigal, PhD; Claudia Muñoz, MS; Victoria Tirado, MS; Sonia Moreno, MS; Pierre N. Tariot, MD; Francisco Lopera, MD; and Eric M. Reiman, MD

ABSTRACT

Objective: To identify a cognitive composite that is sensitive to tracking preclinical Alzheimer's disease decline to be used as a primary end point in treatment trials.

Method: We capitalized on longitudinal data collected from 1995 to 2010 from cognitively unimpaired presenilin 1 (*PSEN1*) E280A mutation carriers from the world's largest known early-onset autosomal dominant Alzheimer's disease kindred to identify a composite cognitive test with the greatest statistical power to track preclinical Alzheimer's disease decline and estimate the number of carriers age 30 years and older needed to detect a treatment effect in the Alzheimer's Prevention Initiative's (API) preclinical Alzheimer's disease treatment trial. The mean-to-standard-deviation ratios (MSDRs) of change over time were calculated in a search for the optimal combination of 1 to 7 cognitive tests/subtests drawn from the neuropsychological test battery in cognitively unimpaired mutation carriers during a 2- and 5-year follow-up period (n = 78 and 57), using data from noncarriers (n = 31 and 56) during the same time period to correct for aging and practice effects. Combinations that performed well were then evaluated for robustness across follow-up years, occurrence of selected items within top-performing combinations, and representation of relevant cognitive domains.

Results: The optimal test combination included Consortium to Establish a Registry for Alzheimer's Disease (CERAD) Word List Recall, CERAD Boston Naming Test (high frequency items), Mini-Mental State Examination (MMSE) Orientation to Time, CERAD Constructional Praxis, and Raven's Progressive Matrices (Set A), with an MSDR of 1.62. This composite is more sensitive than using either the CERAD Word List Recall (MSDR = 0.38) or the entire CERAD-Col battery (MSDR = 0.76). A sample size of 75 cognitively normal *PSEN1* E280A mutation carriers aged 30 years and older per treatment arm allows for a detectable treatment effect of 29% in a 60-month trial (80% power, $P = .05$).

Conclusions: We have identified a composite cognitive test score representing multiple cognitive domains that, compared to the most sensitive single test item, has improved power to track preclinical Alzheimer's disease decline in autosomal dominant Alzheimer's disease mutation carriers and to evaluate preclinical Alzheimer's disease treatments. This API composite cognitive test score will be used as the primary end point in the first API trial in cognitively unimpaired autosomal dominant Alzheimer's disease carriers within 15 years of their estimated age at clinical onset. We have independently confirmed our findings in a separate cohort of cognitively healthy older adults who progressed to the clinical stages of late-onset Alzheimer's disease, described in a separate report, and continue to refine the composite in independent cohorts and compared with other analytic approaches.

J Clin Psychiatry 2014;75(6):652–660

Submitted: *December 11, 2013; accepted February 13, 2014.*
Online ahead of print: *April 15, 2014 (doi:10.4088/JCP.13m08927).*
Corresponding author: *Napatkamon Ayutyanont, PhD, Banner Alzheimer's Institute, 901 E Willetta St, Phoenix, AZ 85006 (Napatkamon.Ayutyanont@bannerhealth.com).*

There is an urgent need to find effective preclinical Alzheimer's disease treatments, which we have previously defined as "interventions that are started in the absence of mild cognitive impairment or dementia and intended to postpone the onset, reduce the risk of, or completely prevent the clinical stages of Alzheimer's disease."[1(p2)] Several trials to address this need were recently launched or are being planned, including those with the strategy of testing therapies in people who are at the highest imminent risk of developing mild cognitive impairment or Alzheimer's disease dementia due to factors such as age and genetic backgrounds or presence of biomarker evidence of Alzheimer's disease.[2–5] Detecting a treatment effect in a preclinical Alzheimer's disease trial using clinical progression or cognitive outcome developed for studies in mild cognitive impairment or Alzheimer's disease dementia as the primary end point may not be desirable because of the large sample size and lengthy follow-up required[6] or the psychometric properties of the tests themselves.[7–9] Using multiple cognitive assessments that are sensitive to preclinical Alzheimer's disease as potentially successful outcomes inflates type I error. Using an appropriate composite minimizes the number of outcomes employed and thus the risk of type I error; additionally, a composite tool can be empirically derived, and its sensitivity to detecting and tracking preclinical Alzheimer's disease can be validated in independent datasets. As a result, it affords a measure of multiple cognitive domains that can serve as a primary end point in preclinical treatment trials.[10]

Clinical Points

- We have identified a composite cognitive test score representing multiple cognitive domains that has improved power to track preclinical Alzheimer's disease decline in autosomal dominant Alzheimer's disease mutation carriers and to evaluate preclinical Alzheimer's disease treatments.
- This Alzheimer's Prevention Initiative (API) composite cognitive test score will be used as the primary end point in the first API trial in cognitively unimpaired autosomal dominant Alzheimer's disease carriers within 15 years of their estimated age at clinical onset.
- We have independently confirmed our findings in a separate cohort of cognitively healthy older adults who progressed to the clinical stages of late-onset Alzheimer's disease and continue to refine the composite.

Slight, but measurable cognitive decline has been reported during preclinical Alzheimer's disease. Retrospective and prospective studies of cognitively normal individuals who subsequently progressed to Alzheimer's disease dementia have found episodic memory decline to be a defining feature of preclinical Alzheimer's disease.[11–15] Decline in other cognitive domains, such as executive,[16] visuospatial,[13] and global cognitive functioning[13,17] also occurs during the transition from normal aging to preclinical Alzheimer's disease and into the clinical stages of Alzheimer's disease. In cognitively unimpaired individuals with significant fibrillar amyloid burden, decline has been observed primarily in episodic memory, executive function, and language.[18–22] Studies of cognitively unimpaired autosomal dominant Alzheimer's disease mutation carriers have reported subtle decline in memory, language, praxis, abstract reasoning, and attention.[23]

Recent research has focused on developing a measure of Alzheimer's disease–related cognitive decline to track the progression of preclinical Alzheimer's disease in order to evaluate investigational preclinical Alzheimer's disease treatments with increased statistical power.[24] A theoretically driven approach reasons that a composite (ie, a test score derived from 2 or more different cognitive tests) should be constructed a priori from cognitive assessments known to decline relatively early in the disease progression. A related approach is to construct a composite score that summarizes the performance in a specific domain, such as memory[25] or executive functioning,[26] believed to be preferentially affected by Alzheimer's disease. An empirically driven approach employs computational modeling techniques to identify an end point or composite[24] based on its sensitivity to detect and track the outcome of interest, such as preclinical Alzheimer's disease. Analysis methods that can be used for developing cognitive composites include, among others, latent variable analyses or partial least squares regression,[24,27,28] item response theory[29] and principal components regression analysis,[30] Rasch measurement theory, or item-level analysis. Note that these approaches are not mutually exclusive (eg, theoretical knowledge of preclinical Alzheimer's disease can be taken into account when empirically deriving a composite cognitive test score).

In the present study, we aimed to develop a composite cognitive test score that is highly sensitive to detecting and tracking preclinical cognitive decline, corresponding to an analysis of a change from baseline, rather than to optimize the discrimination between those who progress to clinical Alzheimer's disease versus those who remain cognitively unimpaired. We examined longitudinal data from cognitively unimpaired presenilin 1 (*PSEN1*) E280A mutation carriers and noncarriers from the world's largest known early-onset autosomal dominant Alzheimer's disease kindred to develop a composite cognitive test score most sensitive to detecting and tracking preclinical cognitive decline and calculate the number of cognitively unimpaired *PSEN1* E280A mutation carriers within 15 years of their estimated age at clinical onset needed to evaluate an amyloid-modifying treatment in the first preclinical Alzheimer's disease trial of the Alzheimer's Prevention Initiative (API).[1,31–33] We hypothesize that the identified composite will have higher sensitivity and have greater statistical power to detect and track cognitive decline associated with preclinical Alzheimer's disease compared to the most sensitive individual cognitive test/subtest score or to the entire neuropsychological assessment battery, given that the empirically driven approach allows for the addition of assessments that improve overall sensitivity despite, perhaps, being less sensitive individually to preclinical Alzheimer's disease decline.

METHOD

Participants

Descendants of patients with confirmed *PSEN1* E280A mutations were enrolled in the E280A Antioquia cohort study between 1995 and 2010 conducted by the Neuroscience group at University of Antioquia and approved by the medical ethics board of the University of Antioquia, Colombia.[34] Participants in the cohort study must be 17 years of age or older; there are no exclusion criteria regarding medical and neuropsychological monitoring. The participants or their guardians provided their informed consent. Participants without signs of dementia are not provided their genetic status. The original dataset is available from Grupo de Neurociencias de Antioquia, Universidad de Antioquia, Medellín, Colombia.

For the present study, only data from cohort study participants who met the following criteria were used in the analyses: (1) age 30 years or older at baseline (approximately 15 years prior to median age at clinical onset),[23] (2) no diagnosis of mild cognitive impairment or dementia due to Alzheimer's disease between the baseline and 2- and 5-year follow-up visits, (3) minimum of 2 or 5 years of longitudinal neuropsychological testing data, and (4) no report of retardation, cerebral paralysis, cerebral lesion, major psychiatric disease, serious systemic illness, uncontrolled seizures, or alcohol abuse, which would preclude participation in a typical clinical trial. The resulting dataset was composed of 56 *PSEN1* E280A carriers and 78

Figure 1. Participant Selection

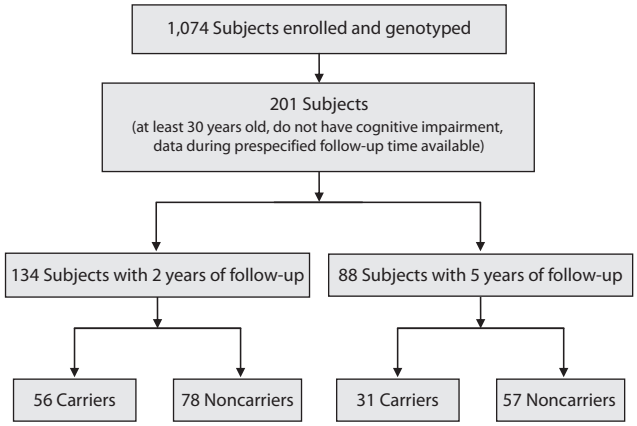


Table 1. Cognitive Assessments

Cognitive Assessment	Cognitive Domain Tested
Visual A cancellation tests	Attention
Trail Making Test-Part A	Attention
Memory of Three Phrases	Memory
Rey-Osterrieth Complex Figure-Recall	Memory
CERAD Word List Recall	Memory
CERAD Word List Recognition	Memory
Recall of Drawings	Memory
Categorical Verbal Fluency	Language ability
CERAD Boston Naming Test	Language ability
Constructional Praxis	Constructional abilities
Rey-Osterrieth Complex Figure-Copy	Constructional abilities
Raven's Progressive Matrices (Set A)	Abstract reasoning
Wechsler Intelligence Scale-Arithmetic (revised)	Calculation abilities
Wisconsin Card Sorting Test	Executive function
Phonological Verbal Fluency-*F* test	Executive function
Mini-Mental State Examination	Orientation, memory, attention and concentration, language ability

Abbreviation: CERAD = Consortium to Establish a Registry for Alzheimer's Disease.

noncarriers for the 24-month analyses and 31 carriers and 57 noncarriers for the 60-month analyses (Figure 1).

Cognitive and Clinical Evaluations

An initial interview and follow-up examination(s), including medical, psychological, and neuropsychological assessments, were performed by neurologists or psychologists trained in neuropsychology who were masked to participants' carrier status. The assessment protocol included the Consortium to Establish a Registry for Alzheimer's Disease (CERAD) neuropsychological battery as well as additional tests to further assess constructional abilities and abstraction, which were translated to Spanish and adapted to the cultural and linguistic idiosyncrasies of the target population (referred to as the CERAD-Col).[23,34–36]

The CERAD-Col assessment battery and details of its administration have been previously described[23,34–36] and are shown in Table 1. Dementia functional scales were also administered.

Diagnostic classification followed a procedure previously described.[23] Briefly, mild cognitive impairment criteria included (1) clinically significant cognitive decline as indicated by cognitive test scores of 2 standard deviations or more away from the mean normal value scores for noncarriers in at least 1 test on any cognitive domain, adjusted for age and education, and (2) subjective memory impairment corroborated by an informant. Dementia criteria included (1) impaired instrumental activities of daily living, (2) impaired activities of daily living, and (3) meeting *DSM-IV* criteria for dementia.

Data Analysis

We performed a search of every combination of 1 to 7 cognitive assessments and calculated the corresponding annualized mean-to-standard-deviation ratios (MSDRs) of the standardized change over time for the cognitively unimpaired *PSEN1* E280A mutation carriers aged 30 years and older during 2 (n = 57) and 5 years (n = 31) of follow-up. Mean-to-standard-deviation ratio values were adjusted for practice effects using data from the kindred mutation noncarriers (2-year follow-up, n = 78; 5-year follow-up, n = 56) by calculating the mean change of the composite score in the noncarriers and subtracting this value from the composite score change calculated in the mutation carriers. The MSDR was chosen as a measurement of sensitivity to the longitudinal decline for a cognitive test combination, representing the coefficient of change (the mean of standardized change divided by the standard deviation of standardized change) and was calculated as follows:

$$\text{MSDR} = \frac{\overline{X}}{\sigma_x}; \text{ change in composite score}$$

$$X_j = \sum_{i=1}^{n} x_i \Big/ n$$

where x_i is a change in standardized cognitive score i, n is a number of cognitive scores in the composite, X_j is the change in composite score of subject j, and σ_x is the standard deviation of the cognitive scores.

The MSDR is quite similar to an effect size, as components of the MSDR are used to calculate it and, the larger the MSDR value, the greater the sensitivity to detecting and tracking cognitive decline over time. Prior to calculating the MSDRs, each cognitive assessment was standardized on a 0–1 scale, similar to a z score. For assessments that did not have a predefined maximum score (such as Categorical Fluency Test), a value of 2 standard deviations above the mean was used as the maximum.

Results from these analyses were used as 1 way to assess the combinations and determine an optimal composite. Tests that were consistently represented in the combinations with the highest sensitivity and that also demonstrated consistency within separate years of the 2- and 5-year follow-up time period were identified as robust items for measuring change. The optimal combination was then evaluated for construct validity and was used to calculate the sample size required

Table 2. Baseline Characteristics of Participants Included in the Power Analysis for 24-Month and 60-Month Analysis in a Randomized Controlled Trial[a]

	2-Year Analysis			5-Year Analysis		
Characteristic	Carrier (n = 56)	Noncarrier (n = 78)	*P* Value[b]	Carrier (n = 31)	Noncarrier (n = 57)	*P* Value[b]
Age	43.93 ± 6.49	44.99 ± 10.07	.49	41.37 ± 4.36	45.35 ± 11.05	.06
Education	7.84 ± 4.57	7.29 ± 3.97	.92	6.74 ± 4.93	6.86 ± 4.96	.92
Sex (male/female), %	39/61	22/78	.03	35/65	37/63	.89
MMSE	26.63 ± 4.01	28.95 ± 1.79	1.27e-5	26.97 ± 2.36	28.53 ± 2.10	2.01e-3
Verbal Fluency (animal)	15.51 ± 5.08	18.33 ± 4.45	1.01e-3	17 ± 4.77	18.82 ± 3.91	.06
CERAD Boston Naming Test	12.02 ± 2.23	12.64 ± 2.01	.01	12.3 ± 2.14	12.44 ± 1.93	.77
CERAD Word List Immediate						
Correct	13.49 ± 5.90	17.78 ± 4.83	1.18e-5	15.43 ± 4.88	17.24 ± 4.81	.11
Intrusions	3.55 ± 4.54	1.45 ± 2.09	5.06e-4	2.87 ± 2.75	1.56 ± 3.27	.07
CERAD Word List Recall						
Correct	3.75 ± 2.89	6.49 ± 2.22	1.08e-8	4.23 ± 2.74	6.47 ± 2.07	.73e-5
Intrusions	1.34 ± 1.62	0.6 ± 1.44	7.20e-3	1.03 ± 1.50	0.56 ± 1.40	.15
CERAD Word List Recognition						
Correct "yes"	8.6 ± 1.70	9.55 ± 1.03	1.19e-4	9.13 ± 1.11	9.35 ± 1.44	.49
Correct "no"	8.81 ± 1.82	9.9 ± 0.41	1.23e-6	9.3 ± 1.18	9.73 ± 1.10	.01
Constructional Praxis	9.26 ± 1.57	9.51 ± 1.6	.38	9.7 ± 1.21	9.6 ± 1.62	.77
Trail Making Test-Part A						
Errors	1.11 ± 3.63	0.28 ± 1.39	.08	0.37 ± 1.55	0.13 ± 0.49	.31
Time	115.43 ± 80	81.39 ± 61.55	.01	93.07 ± 50.94	86.96 ± 50.95	.62
Recall of Drawings	7.2 ± 7.01	14.06 ± 7.15	3.08e-7	8.18 ± 7.76	13.63 ± 7.51	2.16e-3
Raven's Progressive Matrices	7.92 ± 2.28	8.41 ± 2.08	.21	8.03 ± 2.06	8.36 ± 2.23	.51

[a]All data are mean ± SD unless otherwise stated.
[b]Baseline group differences were compared using 2-tailed *t* test, and χ^2 test.
Abbreviations: CERAD = Consortium to Establish a Registry for Alzheimer's Disease, MMSE = Mini-Mental State Examination.

in a 60-month trial to detect a 25% treatment effect, with 80% power and $P = .05$, as well as to calculate the detectable treatment effect a 60-month trial with 75 *PSEN1* E280A mutation carriers per treatment arm.

Weighting the Optimal Composite Cognitive Test Score

After identifying the optimal composite cognitive test score, we examined whether the MSDR could be increased (and therefore, the sensitivity improved) by weighting the individual assessments included in the composite. A search of every potential weighting combination to optimize the sensitivity such that

$$X_j = \sum_{i=1}^{n} w_{ij} x_{ij},$$

where w_{ij} is the weight for test i and subject j and $w_{ij} \geq 0$ and $\sum_{i=1}^{n}$, was conducted in *PSEN1* E280A carriers during the 2- and 5-year follow-up period. Note that w_{ij} is the same for every subject (ie, although weights differ between the tests/subtest, the weight for each test/subtest is constant across all subjects). Data from the noncarriers were used to adjust for practice effects. The combination of weights that resulted in the largest adjusted MSDR was then used to calculate the sample size needed to detect a 25% treatment effect and estimate the treatment effect that could be detected in 75 *PSEN1* E280A mutation carriers with 80% power and $P = .05$.

Evaluating the Optimal Composite Cognitive Test Score

To confirm the sensitivity of the composite cognitive test score, the MSDR of the composite was compared to the MSDR of the CERAD Word List Recall, an episodic memory assessment, and the MSDR of the entire CERAD-Col battery. Additionally, to evaluate the stability of the composite cognitive test score, selected test items from the composite test score were replaced with a different test item from the same cognitive domain, and the resulting MSDRs and required sample sizes were compared.

RESULTS

Participant Characteristics

At baseline, the *PSEN1* E280A carriers and kindred noncarriers did not differ in terms of age or level of education. The carrier group included in the 2-year analysis had a higher ratio of males compared to the noncarrier group, but this difference was not present in the 5-year analysis (Table 2).

Individual Cognitive Assessment Properties

The most sensitive individual neuropsychological tests for differentiating *PSEN1* E280A carriers from noncarriers at baseline included the CERAD Word List Recall, CERAD Word List Recognition, Recall of Drawings, Mini-Mental State Examination (MMSE) Total, and MMSE Orientation to Time ($P \leq .05$). The individual neuropsychological tests most sensitive to longitudinal decline during the 5-year follow-up period (unadjusted for practice effect) included Memory of Three Phrases, Wechsler-Arithmetic, Recall of Drawings, Rey-Osterrieth Complex Figure-Copy, Constructional Praxis (cube), Raven's Progressive Matrices (Set A), CERAD Word List Recognition-total correct, Constructional Praxis, and MMSE Orientation to Time (Table 3). After adjusting for practice effects using data from the noncarriers, the individual neuropsychological tests most sensitive to

Table 3. The Most Sensitive Individual Item Mean to Standard Deviation Ratios (MSDRs) During 5-Year Analysis

Cognitive Assessment	Unadjusted MSDR	Cognitive Assessment	Adjusted MSDR[a]
Memory of Three Phrases	0.99	Wechsler-Arithmetic	1.09
Wechsler-Arithmetic	0.96	Memory of Three Phrases	0.98
Recall of Drawings	0.85	Recall of Drawings	0.94
Rey-Osterrieth Complex Figure-Copy	0.85	Rey-Osterrieth Complex Figure-Copy	0.92
Constructional Praxis (cube)	0.84	Raven's Progressive Matrices	0.85
Raven's Progressive Matrices	0.81	MMSE Orientation to Time	0.77
CERAD Word List Recognition-total correct	0.73	Constructional Praxis (cube)	0.72
Constructional Praxis	0.73	Constructional Praxis	0.68
MMSE Orientation to Time	0.73	Trail Making Test A-Time	0.67

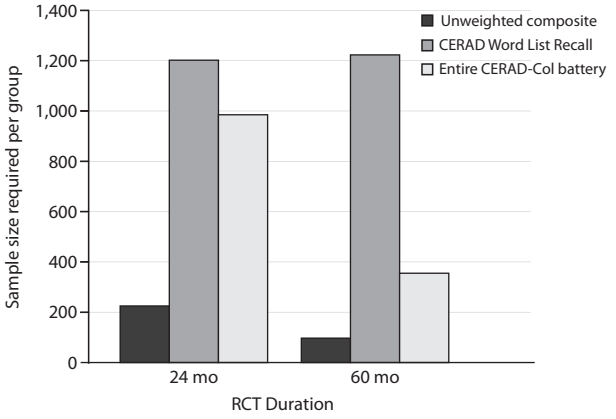
[a]Adjusted for practice effect using longitudinal data from *PSEN1* noncarriers.
Abbreviations: CERAD = Consortium to Establish a Registry for Alzheimer's Disease, MMSE = Mini-Mental State Examination.

Table 4. Estimated Sample Size (completers) Required to Detect 25% Treatment Effect, With 80% Power and α = .05

Composite Measure[a]	Month	Total n	Adjusted MSDR	Estimated Sample Size (completers) per Group, n
Unweighted	60	56	1.62	97
Unweighted	24	95	1.06	225
Weighted	60	56	1.93	69
Weighted	24	95	1.19	179

[a]Composite measure includes CERAD Word List Recall, CERAD Boston Naming Test (high frequency items), Mini-Mental State Examination Orientation to Time, Constructional Praxis, and Raven's Progressive Matrices.
Abbreviations: CERAD = Consortium to Establish a Registry for Alzheimer's Disease, MSDR = mean to standard deviation ratio.

Figure 2. Numbers of *PSEN1* E280A Mutation Carriers per Group Over the Age of 30 Years Required to Detect a 25% Treatment Effect in a 24- and 60-Month Randomized Controlled Trial Using Unweighted Composite, CERAD Word List Recall, or Entire CERAD Battery End Points, With 80% Power and α = .05 Type I Error



Abbreviation: CERAD = Consortium to Establish a Registry for Alzheimer's Disease

longitudinal decline during the 5-year follow-up were nearly identical to those from the unadjusted analysis (Table 3).

Empirically Deriving the API Composite Cognitive Test Score

The combination most sensitive to detecting preclinical cognitive decline related to autosomal dominant Alzheimer's disease, adjusting for aging and practice effects, that has construct validity and also includes tests/subtests that are robust across follow-up time periods consisted of MMSE Orientation to Time, CERAD Boston Naming Test (high frequency items), CERAD Word List Recall, Constructional Praxis, and Raven's Progressive Matrices (Set A). On the basis of the 5-year follow-up data, the total 60-month MSDR of the composite cognitive test score is 1.62. In comparison, the most sensitive individual cognitive assessment is the Memory of Three Phrases, with a total 60-month MSDR of 0.99, making the composite cognitive test score considerably more sensitive to tracking preclinical cognitive decline in autosomal dominant Alzheimer's disease mutation carriers. On the basis of the 2-year longitudinal data, a shorter study with the same composite cognitive test score would result in a total MSDR of 1.06. This is important to consider, as the MSDR is a coefficient of change (the mean change divided by the standard deviation of change), in which a larger value indicates the sensitivity of the measure, thereby impacting the required sample size and detectable treatment effect.[24]

On the basis of the MSDR of the API composite cognitive test score, 97 *PSEN1* E280A mutation carriers who complete the trial per group aged 30 years and older are needed to detect a 25% treatment effect in a 60-month randomized controlled trial (Table 4). In contrast, if the CERAD Word List Recall is used, 1,223 mutation carriers per group aged 30 years and older are needed to detect a 25% treatment effect in a 60-month randomized controlled trial (MSDR = 0.38), while use of the entire CERAD-Col battery would require 355 carriers per group MSDR = 0.76) (Figure 2).

Using the API composite cognitive test score, we estimate that a trial of 75 mutation carriers who complete the trial per treatment arm aged 30 years and older would permit us to detect a 29% treatment effect in a 60-month trial (Table 5). In comparison, using the CERAD World List Recall would permit us to detect a 103% treatment effect, while the CERAD-Col would permit us to detect a 55% treatment effect.

Results from the sensitivity analyses suggest that substituting either the CERAD Boston Naming Test (total score) or Categorical Fluency Test (animals) for the CERAD Boston Naming Test (high frequency items) requires a larger sample size to detect a 25% treatment effect in a 24-month trial. Substituting the CERAD Boston Naming Test (total score) for the high frequency items would require 398 participants per group, while replacing the CERAD Boston Naming Test (high frequency items) with Category Fluency

Table 5. Estimated Minimal Detectable Treatment Effect Required Using With 75 Completers per Group, With 80% Power and α = .05

Composite Measure[a]	Month	Total n	Adjusted MSDR	Detectable Treatment Effect
Unweighted	60	56	1.62	29
Unweighted	24	95	1.06	44
Weighted	60	56	1.93	24
Weighted	24	95	1.19	39

[a]Composite measure includes CERAD Word List Recall, CERAD Boston Naming Test(high frequency items), Mini-Mental State Examination Orientation to Time, Constructional Praxis, and Raven's Progressive Matrices.
Abbreviations: CERAD = Consortium to Establish a Registry for Alzheimer's Disease, MSDR = mean to standard deviation ratio.

Test (animals) requires 404 participants per group. Similarly, but to a lesser degree, in a 60-month trial, substituting the CERAD Boston Naming Test (total score) for the high frequency items would require a larger sample size of 98 participants per group, while replacing the CERAD Boston Naming Test (high frequency items) with Category Fluency Test (animals) requires 102 participants per group.

Results from the weighting analyses indicated that applying a higher weighting to the CERAD Boston Naming Test (high frequency items) improved the MSDR. Although the weightings had similar patterns at 2 and 5 years, they were not identical. Compared to the original, nonweighted composite cognitive test score, the weighted composite cognitive test score requires 69 *PSEN1* E280A mutation carriers per group aged 30 years and older to detect a 25% treatment effect in 60-month trial. Using the weighted composite in a trial of 75 mutation carriers per group aged 30 years and older would permit us to detect a 24% treatment effect in a 60-month trial (Table 5).

DISCUSSION

We empirically identified an API composite cognitive test score sensitive to preclinical cognitive decline in autosomal dominant Alzheimer's disease mutation carriers within 15 years of their estimated age at clinical onset. We propose that this composite is well suited for preclinical autosomal dominant Alzheimer's disease trials to evaluate treatment effects with smaller sample sizes and improved statistical power compared to the most sensitive individual cognitive assessment or larger test batteries, and in a manner that is reasonably likely to predict a treatment's clinical benefit. The API composite cognitive test score and the analytic approach used in its development appear to fit into the US Food and Drug Administration's framework in the recent draft guidance regarding a cognitive assessment being a primary efficacy measure in preclinical Alzheimer's disease trials.[37] Moreover, the optimal combination of assessments empirically identified in preclinical autosomal dominant Alzheimer's disease mutation carriers is quite similar to the composite cognitive test score identified in older adults who progressed to clinical stages of late-onset Alzheimer's disease (LOAD).[38]

The empirically identified composite cognitive test score consisted of 5 test items targeting several different cognitive domains. The composite has greater statistical power to detect a treatment effect compared to that of a single test item (CERAD Word List Recall), supporting the notion that combining test items can result in better captured variance for tracking preclinical cognitive decline. The optimal composite cognitive test score is more sensitive than that of the entire neuropsychological test battery. This is consistent with the hypothesis that including test items that capture overlapping variation or that are not sensitive to preclinical Alzheimer's disease into the test score can lower the overall sensitivity of the score to track preclinical cognitive decline. The other reason the entire battery may be less sensitive than a subset is that the entire test battery may include assessments that are psychometrically noisy as well as tests that have excellent psychometric properties, thus increasing the overall variability.

Our optimal composite cognitive test scores incorporate cognitive assessments from several different domains, complementing those findings of recent studies, which suggest that cognitive decline in preclinical Alzheimer's disease presents in multiple domains[13,39] in addition to decline in episodic memory (although it remains a defining trait of preclinical Alzheimer's disease),[11–15,23] along with other studies focusing on cognitive domain specific composite scores based on data from the Alzheimer's Disease Neuroimaging Initiative.[25,26] These research results suggest that composite end points may offer greater power and sensitivity to detect cognitive changes.

Confirming the findings in this study, we obtained a very similar composite cognitive test score from an independent analysis[38] performed in older adults who later progressed to clinical stages of LOAD. Both optimal cognitive composite test scores consisted of assessments from the same domains/assessments, with the exception of the present study, which included the test Constructional Praxis, whereas the other study[38] included the visuospatial ability test Symbol Digits Modalities, despite substantial differences in the cohorts' neuropsychological test batteries. In addition, the results from the API efforts complement a recent study[13] that suggested multiple cognitive domains decline in preclinical Alzheimer's disease, including verbal and working memory, visuospatial, and global functioning. The significant overlap between the 2 optimal composite test scores suggests the similar patterns of cognitive decline between LOAD and autosomal dominant Alzheimer's disease, despite evidently different ages at onset and possible different time courses and underlying etiologies and biological processes. Likewise, researchers preparing the Alzheimer's Disease Cooperative Study "A4" trial[4] in cognitively healthy older adults with amyloid burden pathology have implemented a similar approach using other datasets and found comparable results to those reported in this study.

Despite the similarity of the composite cognitive test scores, the MSDR of the API composite cognitive test score empirically derived by using the *PSEN1* E280A cohort data is considerably higher than that in older adults who progress to the clinical stages of LOAD described in a separate report.[38]

This is consistent with the fact that *PSEN1* E280A mutation carriers are certain to develop symptomatic dementia in a predictable timeline and clinical course. Additionally, since the *PSEN1* E280A carriers are relatively young, the cognitive decline observed is most likely only due to the predisposition to Alzheimer's disease, that is, preclinical Alzheimer's disease decline, as opposed to a confounding aging effect that is observed in older adults who progress to clinical stages of LOAD.

Although individual neuropsychological tests have varying levels of sensitivity to detecting and tracking preclinical Alzheimer's disease decline (measured by their MSDRs), this analytic approach allowed us to empirically characterize the composite cognitive test score, resulting in the high overall sensitivity to track preclinical decline by simultaneously determining a combination of individual tests that complement each other to capture as much variability as possible. As a result, more sensitive cognitive tests/subtests may not be included in the composite end point, since these items may correlate with another assessment that captures the same information and has a higher MSDR. The tests/subtests that are included and have smaller MSDR may measure variability not captured by other assessments in the combination. This is different from the approach in which each sensitive test is determined individually at a time and then simply combined to form a composite. The result of the latter approach may be a composite with a lower sensitivity due to overlap in elements of variability captured by the tests, making them redundant to each other and, in turn, weaken overall sensitivity of the composite. Similarly, the latter approach may result in the loss of opportunity to identify tests that may be less sensitive on their own but may add to the composite by capturing additional aspects of variability not captured by other assessments. Another possible reason our proposed analytic approach resulted in increased sensitivity is that it helps reduce the impact of error due to other idiosyncratic single test items or subdomains in the composite. Moreover, weighting the individual tests in the composite allowed the tests that capture additional variability or are more sensitive to preclinical autosomal dominant Alzheimer's disease decline to have a greater effect on the composite test score, resulting in even higher statistical power to detect preclinical cognitive decline.

In this study, we aimed to characterize the aspects of the disease that decline consistently across individuals in order to assess effectiveness of a treatment in slowing decline in a preclinical Alzheimer's disease trial, rather than discrimination between those who subsequently progressed and those that did not, or the neuropathological underpinnings of Alzheimer's disease that result in a change in cognitive functioning. This approach also allows for the incorporation of data from various points along the preclinical Alzheimer's disease stages, just as in a preclinical trial, some participants may progress to cognitive impairment within months, while others remain cognitively healthy for many years. In addition, we chose to adjust for practice effects[40] to better capture the cognitive decline specific to Alzheimer's disease. The noncarrier group showed an increase on the API composite cognitive score, while the carrier group showed a reduction. This suggests that, unlike the noncarrier group, the carrier group was not able to benefit as much from prior exposure to the tests. Although it is important to account for differences in study participants' baseline cognitive function, the present study did not adjust for such differences, given that they can be accounted for when analyzing the trial data to determine whether a treatment is effective at slowing cognitive decline.

There are some limitations to the present study. For instance, development of the optimal composite cognitive test score was constrained by the starting neuropsychological test battery used in the Antioquia cohort study and the composite development sample size available. That said, we achieved remarkably similar results with independent efforts to empirically deriving a composite cognitive test score based on data from individuals who progress to the clinical stages of LOAD[38] despite differences in the cohorts' starting neuropsychological test battery. Likewise, scientists preparing for the Alzheimer's Disease Cooperative Study A4 trial in cognitively healthy individuals with significant fibrillar amyloid burden have undertaken a similar effort using other datasets and have produced results comparable to those reported here.[4] The generalizability and sensitivity of the composite cognitive test score to other autosomal dominant Alzheimer's disease mutations remain unknown, given the limited preclinical longitudinal data available. That said, recent evidence[41] has suggested that there is no significant difference in cognitive measures when comparing *PSEN1* mutation carriers to PSEN2 and amyloid precursor protein mutation carriers. Additional efforts are underway to confirm the generalizability and power of the API composite cognitive test score in other populations followed to clinical progression (which may include different assessment batteries) and to estimate the statistical power in different preclinical Alzheimer's disease participant groups (eg, apolipoprotein E [*APOE*] ε4 homozygotes or heterozygotes at different ages, older adults with or without biomarker evidence of Alzheimer's disease). The results from these analyses, along with sample size estimates, will be reported separately.

In summary, we examined longitudinal data from cognitively unimpaired *PSEN1* E280A mutation carriers within 15 years of their estimated mean age of dementia onset and conducted a search of every combination of 1 to 7 cognitive assessments to identify the optimal combination that is sensitive to tracking preclinical Alzheimer's disease decline over a 2- and 5-year time period, while controlling for practice effects using data from kindred mutation noncarriers. The empirically identified API composite cognitive test score is being used as the primary end point in the first API trial in cognitively unimpaired autosomal dominant Alzheimer's disease carriers within 15 years of their estimated age at clinical onset. This composite end point requires fewer participants to detect a treatment effect compared to using the most sensitive individual cognitive test or the entire

CERAD-Col neuropsychological test battery. A similar composite cognitive test score was independently derived in cognitively unimpaired older adults who subsequently progressed to clinical stages of LOAD.[38] As a result of these efforts, other preclinical trial investigators are extending the API composite cognitive test score development strategy for use in their planned trials and studies.

Author affiliations: Banner Alzheimer's Institute (Drs Ayutyanont, Langbaum, Chen, Fleisher, Tariot, and Reiman); Neurogenomics Division, Translational Genomics Research Institute (Dr Reiman); Arizona Alzheimer's Consortium (Drs Ayutyanont, Langbaum, Chen, Fleisher, Tariot, and Reiman), Phoenix; Department of Mathematics and Statistics, Arizona State University, Tempe (Dr Chen); Department of Psychiatry, University of Arizona, Tucson (Drs Tariot and Reiman); Pentara Corporation, Salt Lake City, Utah (Dr Hendrix); Department of Neurology, University of California, San Diego (Dr Fleisher); F. Hoffmann-La Roche, Ltd, South San Francisco, California (Dr Ward and Mr Friesenhahn); and Grupo de Neurociencias de Antioquia, Universidad de Antioquia, Medellín, Colombia (Drs Acosta-Baena, Madrigal, and Lopera and Mss Aguirre, Muñoz, Tirado, and Moreno).
Author contributions: Drs Lopera and Reiman are co-senior authors of this article.
Potential conflicts of interest: **Dr Hendrix** is a consultant for Banner Alzheimer's Institute. **Dr Ward** and **Mr Friesenhahn** are employees and stock shareholders of F Hoffmann-La Roche. **Dr Acosta-Baena** is an employee of Grupo de Neurociencias de Antioquia, Universidad de Antioquia. **Dr Tariot** receives consulting fees from Abbott, AbbVie, AC Immune, Boehringer-Ingelheim, California Pacific Medical Center, Chase, CME Inc, Medavante, Otsuka, and Sanofi-Aventis; receives consulting fees and research support from Avanir, Bristol-Meyers Squibb, Cognoptix, Janssen, Merck, and Hoffmann-La Roche; receives research support from AstraZeneca, Baxter Healthcare, Functional Neuromodulation, GE, Genentech, Pfizer, and Targacept; receives other research support from the National Institute on Aging and Arizona Department of Health Services; owns stock options in Adamas; and is listed as a contributor to a patent owned by the University of Rochester, "Biomarkers of Alzheimer's Disease." **Drs Ayutyanont**, **Langbaum**, **Chen**, **Fleisher**, **Madrigal**, **Lopera**, and **Reiman** and **Mss Aguirre**, **Muñoz**, **Tirado**, and **Moreno** report no potential or other conflicts of interest.
Funding/support: This work was supported by the National Institute on Aging (R01AG031581 and P30AG19610 to Dr Reiman); National Institutes of Health (RF1 AG041705); the State of Arizona, Arizona Alzheimer's Consortium (to Dr Ayutyanont); and contributions from the Banner Alzheimer's Foundation, the Nomis Foundation (salary support for Drs Ayutyanont, Langbaum, Hendrix, Chen, Fleisher, Ward, Tariot, and Reiman and Mr Friesenhahn); and Colciencias (1115-408-20543 and 1115-408-20512) (for data collection, data maintenance, and salary support for Drs Acosta-Baena, Madrigal, and Lopera and Mss Aguirre, Muñoz, Tirado, and Moreno).
Disclaimer: Dr Reiman and Dr Tariot, both JCP editorial board members, were not involved in the editorial evaluation or decision to publish this article.
Previous presentations: Portions of this study were presented at the 2011 Alzheimer's Association International Conference; July 16–21, 2011; Paris, France and the 2011 Clinical Trials on Alzheimer's Disease (CTAD); November 3–5, 2011; San Diego, California.
Acknowledgments: The authors thank Madelyn Gutierrez, BA; Eliana Henao, MS; and Adriana Ruiz, MS (Grupo de Neurociencias de Antioquia, Universidad de Antioquia, Medellín, Colombia) for their assistance and support in coordination and data collection. They have no conflict of interest to report. They also thank Laura Jakimovich, RN, MS; Carolyn Langlois, MA; and Auttawut Roontiva, MS (Banner Alzheimer's Institute, Phoenix Arizona and Arizona Alzheimer's Consortium, Phoenix) for their support and assistance in coordination, and data analysis. They have no conflict of interest to report.

REFERENCES

1. Reiman EM, Langbaum JBS, Tariot PN. Alzheimer's prevention initiative: a proposal to evaluate presymptomatic treatments as quickly as possible. *Biomarkers Med*. 2010;4(1):3–14.
2. Reiman EM, Langbaum JB, Fleisher AS, et al. Alzheimer's Prevention Initiative: a plan to accelerate the evaluation of presymptomatic treatments. *J Alzheimers Dis*. 2011;26(suppl 3):321–329.
3. Mills SM, Mallmann J, Santacruz AM, et al. Preclinical trials in autosomal dominant AD: implementation of the DIAN-TU trial. *Rev Neurol (Paris)*. 2013;169(10):737–743.
4. Sperling R, Donohue M, Aisen P. The A4 Trial: Anti-amyloid treatment of asymptomatic Alzheimer's disease. *Alzheimers Dement*. 2012;8(4):P425–P426.
5. Roses AD, Lutz MW, Amrine-Madsen H, et al. A TOMM40 variable-length polymorphism predicts the age of late-onset Alzheimer's disease. *Pharmacogenomics J*. 2010;10(5):375–384.
6. DeKosky ST. Maintaining adherence and retention in dementia prevention trials. *Neurology*. 2006;67(suppl 3):S14–S16.
7. Martin BK, Szekely C, Brandt J, et al; ADAPT Research Group. Cognitive function over time in the Alzheimer's Disease Anti-inflammatory Prevention Trial (ADAPT): results of a randomized, controlled trial of naproxen and celecoxib. *Arch Neurol*. 2008;65(7):896–905.
8. Silverberg NB, Ryan LM, Carrillo MC, et al. Assessment of cognition in early dementia. *Alzheimers Dement*. 2011;7(3):e60–e76.
9. Cano SJ, Posner HB, Moline ML, et al. The ADAS-cog in Alzheimer's disease clinical trials: psychometric evaluation of the sum and its parts. *J Neurol Neurosurg Psychiatry*. 2010;81(12):1363–1368.
10. Aisen PS, Andrieu S, Sampaio C, et al. Report of the task force on designing clinical trials in early (predementia) AD. *Neurology*. 2011;76(3):280–286.
11. Elias MF, Beiser A, Wolf PA, et al. The preclinical phase of Alzheimer disease: a 22-year prospective study of the Framingham Cohort. *Arch Neurol*. 2000;57(6):808–813.
12. Saxton J, Lopez OL, Ratcliff G, et al. Preclinical Alzheimer disease: neuropsychological test performance 1.5 to 8 years prior to onset. *Neurology*. 2004;63(12):2341–2347.
13. Johnson DK, Storandt M, Morris JC, et al. Longitudinal study of the transition from healthy aging to Alzheimer disease. *Arch Neurol*. 2009;66(10):1254–1259.
14. Wilson RS, Leurgans SE, Boyle PA, et al. Cognitive decline in prodromal Alzheimer disease and mild cognitive impairment. *Arch Neurol*. 2011;68(3):351–356.
15. Amieva H, Le Goff M, Millet X, et al. Prodromal Alzheimer's disease: successive emergence of the clinical symptoms. *Ann Neurol*. 2008;64(5):492–498.
16. Grober E, Hall CB, Lipton RB, et al. Memory impairment, executive dysfunction, and intellectual decline in preclinical Alzheimer's disease. *J Int Neuropsychol Soc*. 2008;14(2):266–278.
17. Small BJ, Bäckman L. Longitudinal trajectories of cognitive change in preclinical Alzheimer's disease: a growth mixture modeling analysis. *Cortex*. 2007;43(7):826–834.
18. Stonnington CM, Chen K, Lee W, et al. Fibrillar amyloid correlates of preclinical cognitive decline. *Alzheimers Dement*. 2014;10(1):e1–e8.
19. Pike KE, Ellis KA, Villemagne VL, et al. Cognition and beta-amyloid in preclinical Alzheimer's disease: data from the AIBL study. *Neuropsychologia*. 2011;49(9):2384–2390.
20. Ellis KA, Lim YY, Harrington K, et al; AIBL Research Group. Decline in cognitive function over 18 months in healthy older adults with high amyloid-β. *J Alzheimers Dis*. 2013;34(4):861–871.
21. Lim YY, Maruff P, Pietrzak RH, et al; AIBL Research Group. Effect of amyloid on memory and non-memory decline from preclinical to clinical Alzheimer's disease. *Brain*. 2014;137(pt 1):221–231.
22. Sperling RA, Johnson KA, Doraiswamy PM, et al; AV45-A05 Study Group. 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(3):822–831.
23. Acosta-Baena N, Sepulveda-Falla D, Lopera-Gómez CM, et al. Pre-dementia clinical stages in presenilin 1 E280A familial early-onset Alzheimer's disease: a retrospective cohort study. *Lancet Neurol*. 2011;10(3):213–220.
24. Hendrix SB. Measuring clinical progression in MCI and pre-MCI populations: enrichment and optimizing clinical outcomes over time. *Alzheimers Res Ther*. 2012;4(4):24.
25. Crane PK, Carle A, Gibbons LE, et al; Alzheimer's Disease Neuroimaging Initiative. Development and assessment of a composite score for memory in the Alzheimer's Disease Neuroimaging Initiative (ADNI). *Brain Imaging Behav*. 2012;6(4):502–516.
26. Gibbons LE, Carle AC, Mackin RS, et al; Alzheimer's Disease Neuroimaging Initiative. A composite score for executive functioning, validated in Alzheimer's Disease Neuroimaging Initiative (ADNI) participants with baseline mild cognitive impairment. *Brain Imaging Behav*. 2012;6(4):517–527.
27. Dowling NM, Hermann B, La Rue A, et al. Latent structure and factorial invariance of a neuropsychological test battery for the study of preclinical Alzheimer's disease. *Neuropsychology*. 2010;24(6):742–756.
28. Hayden KM, Kuchibhatla M, Romero HR, et al. Pre-clinical cognitive phenotypes for Alzheimer Disease: a latent profile approach [published

online ahead of print September 27, 2013]. *Am J Geriatr Psychiatry*.
29. Balsis S, Unger AA, Benge JF, et al. Gaining precision on the Alzheimer's Disease Assessment Scale-cognitive: a comparison of item response theory-based scores and total scores. *Alzheimers Dement*. 2012;8(4): 288–294.
30. Xiong C, van Belle G, Chen K, et al. Combining multiple markers to improve the longitudinal rate of progression-application to clinical trials on the early stage of Alzheimer's disease. *Stat Biopharm Res*. 2013;5(1):54–66.
31. Reiman EM, Langbaum JBS. Brain imaging in the evaluation of putative Alzheimer's disease slowing, risk-reducing and prevention therapies. In: Jagust WJ, D'Esposito M, eds. *Imaging the Aging Brain*. New York, NY: Oxford University Press; 2009:319–350.
32. Reiman EM, Langbaum JBS, Fleisher AS, et al. Alzheimer's Prevention Initiative: a plan to accelerate the evaluation of presymptomatic treatments. In: Ashford JW, ed. *Handbook of Imaging the Alzheimer Brain*. Amsterdam, the Netherlands: IOS Press; 2011.
33. Ayutyanont N, Lopera F, Hendrix S, et al. Composite cognitive endpoints with improved power to detect presymptomatic Alzheimer's disease treatment effects: findings in the Colombian kindred with the E280A presenilin 1 mutation and the Alzheimer's Prevention Initiative. *Alzheimers Dement*. 2011;7(4):S608.
34. Lopera F, Ardilla A, Martínez A, et al. Clinical features of early-onset Alzheimer disease in a large kindred with an E280A presenilin-1 mutation. *JAMA*. 1997;277(10):793–799.
35. Arango Lasprilla JC, Iglesias J, Lopera F. Neuropsychological study of familial Alzheimer's disease caused by mutation E280A in the presenilin 1 gene. *Am J Alzheimers Dis Other Demen*. 2003;18(3):137–146.
36. Rosselli MC, Ardila AC, Moreno SC, et al. Cognitive decline in patients with familial Alzheimer's disease associated with E280a presenilin-1 mutation: a longitudinal study. *J Clin Exp Neuropsychol*. 2000;22(4): 483–495.
37. Food and Drug Administration. Guidance for industry Alzheimer's disease: Developing drugs for the treatment of early stage disease. Draft Guidance, 2013. US Department of Health and Human Services Food and Drug Administration, Center for Drug Evaluation and Research (CDER). http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM338287.pdf. Updated January 16, 2013. Accessed March 10, 2014.
38. Langbaum JB, Hendrix SB, Ayutyanont N, et al. An empirically derived composite cognitive endpoint with improved power to track and evaluate treatments for preclinical Alzheimer's disease. *Alzheimers Dement*. In press.
39. Riley KP, Jicha GA, Davis D, et al. Prediction of preclinical Alzheimer's disease: longitudinal rates of change in cognition. *J Alzheimers Dis*. 2011;25(4):707–717.
40. Wilson RS, Li Y, Bienias JL, et al. Cognitive decline in old age: separating retest effects from the effects of growing older. *Psychol Aging*. 2006;21(4): 774–789.
41. Bateman RJ, Xiong C, Benzinger TL, et al; Dominantly Inherited Alzheimer Network. Clinical and biomarker changes in dominantly inherited Alzheimer's disease. *N Engl J Med*. 2012;367(9):795–804.

Editor's Note: We encourage authors to submit papers for consideration as a part of our Focus on Alzheimer's Disease and Related Disorders section. Please contact Eric M. Reiman, MD, at ereiman@psychiatrist.com.