Minimizing the Sample Sizes of Clinical Trials on Preclinical and Early Symptomatic Stage of Alzheimer Disease

Jingqin Luo[#,1,2,3], **Hua Weng**[#,3], **John C. Morris**[4,5,6], and **Chengjie Xiong**[3,4]

[1]Department of Surgery, Division of Public Health Sciences, Washington University School of Medicine, St. Louis, MO, 63110, USA

[2]Siteman Cancer Center Biostatistics Core Washington University School of Medicine, St. Louis, MO, 63110, USA

[3]Division of Biostatistics, Washington University School of Medicine, St. Louis, MO

[4]Knight Alzheimer Disease Research Center, Washington University School of Medicine, St. Louis, MO, 63110, USA

[5]Departments of Pathology and Immunology, Washington University School of Medicine, St. Louis, MO, 63110, USA

[6]Department of Neurology, Washington University School of Medicine, St. Louis, MO 63110, USA

[#] These authors contributed equally to this work.

Abstract

Background: Clinical trials of investigational drugs for Alzheimer disease (AD) increasingly focus on the prodromal (symptomatic) stage of the illness and now its preclinical (asymptomatic) stage. Sensitive and specific cognitive and functional endpoints are needed to track subtle cognitive and functional changes in the early and preclinical stages to minimize sample sizes in these trials.

Objectives: To identify informative items in a standard clinical assessment protocol and a psychometric battery that are predictive of onset of dementia symptom.

Design: Longitudinal retrospective study.

Setting: Washington University (WU) Knight Alzheimer Disease Research Center (ADRC).

Participants: A total of 735 individuals at least 65 years old and cognitively normal at baseline from a longitudinal clinical cohort at the WU Knight ADRC.

Corresponding author: Chengjie Xiong chengjie@wustl.edu**Address**: Division of Biostatistics, Campus Box 8067 4523 Clayton Ave., St. Louis, MO, 63110-1093 Phone: 314-362-3635; Fax: 314-362-2693.

Measurements: The annual clinical assessment included a wide spectrum of functional and cognitive domains; a comprehensive psychometric battery was completed about 2 weeks after the clinical evaluation. Psychometricians are blinded to the results of the clinical evaluation and to the prior performance of the participants on the psychometric tests.

Results: The mean age at baseline of the 735 participants was 74.30 and 62.31% were female. 240 individuals developed prodromal dementia symptoms (consistent with mild cognitive impairment due to AD and with very mild AD dementia) during longitudinal follow-up (mean follow-up=6.79 years). Among a total of 562 items in the clinical and cognitive assessments under analysis, 292 (52%) were identified as informative because their longitudinal changes were predictive of symptomatic onset. When these items were used to form the functional and cognitive composites, the longitudinal rates of changes were free of a learning effect and captured subtle longitudinal progression prior to symptomatic onset. The rates of change were much greater right after the symptomatic onset than those from the functional and cognitive composites formed using non-informative items. Although the sample sizes for prevention trials (prior to symptomatic onset) using the informative items still yield large numbers, the sample sizes for early treatment trial (after symptomatic onset) was much smaller than those derived from all the items or from the non-informative items alone.

Conclusions: The antecedent longitudinal changes in nearly half of the items in a clinical assessment protocol and a comprehensive cognitive battery did not show statistically significant ability to predict the dementia symptom onset, and hence may be non-informative to track the preclinical functional and cognitive progression of AD. The remaining items, on the other hand, captured some of the preclinical changes prior to the symptom onset, but performed much better right after the symptom onset. Currently ongoing prevention trials on preclinical AD of elderly individuals may need to re-assess the sample sizes and statistical power.

Introduction

Alzheimer disease (AD) is a neurodegenerative disorder characterized by the pathophysiological process of formation and accumulation of senile plaques and neurofibrillary tangles in the brain [1] and phenotypical process of progressive impairment of cognition, function and behavior. Considerable evidence accumulated in the past decade through mostly cross-sectional studies suggests that individuals who develop symptomatic AD exhibit cognitive deficits several years before the clinical diagnosis of mild cognitive impairment (MCI) due to AD or of very mild AD [2]. Because AD is an irreversible neurodegenerative disease that results from neuronal loss in one or multiple brain regions, an early detection and intervention offers the optimal hope for the disease treatment. Because current symptomatic therapies are initiated only after diagnosis, their modest benefit may be partly explained by the fact that some irreversible brain damage has already occurred by the time AD is clinically recognized. Given that no pharmaceutical treatments to date have demonstrated efficacy in reversing or stabilizing dementia progression in the mild or

moderate stages of AD, antecedent disease markers when individuals are still cognitively normal or at very early symptomatic stages are especially important to identify individuals at high risk for trials of putative disease-modifying therapies to allow optimal early intervention and prevention.

Although standard cognitive and functional instruments discriminate established symptomatic AD from normal aging, they are far from satisfactory in tracking the early changes of AD, partly because of the enormous ceiling and floor effects[3]. Because data with significant ceiling and floor effects have limited use in tracking the longitudinal changes of the disease, their use as part of the cognitive outcome in prevention or early treatment trials may lead to large sample sizes and represent a waste of time and precious research resources in terms of both research participants/informants and investigators. On the other hand, there is evidence in the literature suggesting even the current composite scores of these tests can help, to some degree, identify individuals at the preclinical or early stage of AD[4]. This implies that many individual items from the standard cognitive battery and functional tests may be informative in identifying individuals at high risk for symptomatic AD. Because these items are buried in and scattered across different tests that were not originally designed for tracking preclinical or very early changes of AD, their potential has not been fully appreciated due to the lack of optimal tools to identify them and to integrate them.

This paper aims to investigate whether longitudinal changes from the individual items in a standard clinical assessment protocol and a comprehensive cognitive battery used by the Washington University (WU) Knight Alzheimer's Disease Research Center (ADRC) were associated with onset of dementia symptoms. We hypothesize that the clinical protocol and cognitive battery contain many individual items that are neither sensitive nor specific for tracking early changes of AD. We further hypothesize that identification of items which are informative to predict symptomatic onset (called 'informative items') will lead to an improved estimate to the longitudinal cognitive decline, both prior to and after symptomatic onset. Finally, we evaluate whether the functional and cognitive endpoints defined by the informative items alone in future prevention or early treatment trials improve statistical power for efficacy comparison over the endpoints derived from the entire functional and cognitive batteries (which contain both informative and non-informative items) and those derived from the non-informative items alone.

Methods

Participants

The WU Knight ADRC has enrolled elderly individuals in a longitudinal clinical-pathologic cohort study of aging and dementia since 1979. Participants received annual clinical and psychometric examinations. The Clinical Dementia Rating (CDR)[5] staged the presence or absence of dementia and, when present, its severity such that CDR 0 indicates cognitively normality and CDR 0.5, 1, 2, and 3 indicates very mild, mild, moderate, and severe dementia, respectively. 735 individuals at least 65 years old and cognitively normal at baseline and followed longitudinally were included in the analyses. All study participants

provided written informed consent. The study was approved by the Institutional Review Board of WU School of Medicine.

Clinical and Psychometric Assessments

The clinical and psychometric assessments were conducted independently to permit the cognitive data to be evaluated without contamination and possible circularity that might result when cognitive scores were used in diagnostic classifications. The clinical assessment at the WU Knight ADRC assesses a wide spectrum of functional and cognitive domains (each with many items) on the participants, which also includes information about the participants provided by their informants. A total of 111 items from participants and 77 items about the participants from the informants were analyzed due to their availability of longitudinal item level data. Some tests/scales were administered on both research participants and informants, at least for a period of time that resulted in longitudinal item level data, including Geriatric Depression Scale (GDS) [6], the box of Judgment and Problem Solving in the CDR sum of boxes [7], and the Short Portable Mental Status Questionnaire [8]. Some tests were administered to participants only, including Mini Mental State Examination (MMSE)[9], Short Blessed Test (SBT)[10], Assessment of Aphasia[11] and a Drawing test. Some other tests were administered to informants only regarding the participant's behavioral features and functional abilities, including the neuropsychiatric inventory questionnaire (NPI-Q)[12], Functional Assessment Questionnaire (FAQ)[13], Ferman test [14], the box of personal care and community affairs in the CDR sum of boxes, orientation and daily activities with items from the Blessed Dementia Scale (BDS)[15] and a depressive features battery.

About 2 weeks after the clinical evaluation, participants completed an approximate 2-hour battery of psychometric tests. Psychometricians were blinded to the results of the clinical evaluation and previous performance of the participant on the psychometric tests. Episodic memory was assessed by the Wechsler Memory Scale Logical Memory, including immediate and delayed tests [16], Digit Span (both forward and backward)[16], Associate Learning subtests from the Wechsler Memory Scale (WMS)[17] and the Visual Retention Test (Form C, 10-second exposure)[18]. Two measures of semantic memory included the Information subset of the Wechsler Adult Intelligence Scale (WAIS) [19] and the Boston Naming Test [20]. WAIS Block Design was also included for measuring visuospatial ability[19]. Other tests in the psychometrics battery were Free and Cued Selective Reminding Test[21] and a mental control test of the Wechsler Memory Scale (WMS)[17]. The total number of items in the clinical and the psychometric battery that were analyzed can be found in Table 2.

Since 2005, the primary clinical and cognitive assessments of the WU Knight ADRC follow that of the National Alzheimer Coordinating Center Uniform Data Set[22], which include standard definitions and diagnostic criteria for detection of dementia and its differential diagnosis[23]. Prior to and after 2005 during more than 30 years of longitudinal follow-up, some of the tests and items were discontinued while others were added. More details regarding time of inclusion/discontinuation of all these items are summarized in Supplemental Table 1.

Other Covariates

Demographics such as baseline age, sex, and years of education were recorded at baseline. *APOE* genotyping was dichotomized into those with at least one copy of the E4 allele (E4 positive) vs. those without an E4 allele (E4 negative).

Selection of Informative items

Each item score was first converted into a binary scale, labeled as endorsement versus non-endorsement of the item, oriented in the way across all items that <u>non-endorsement</u> always indicates difficulty with function or cognition. For the items with a categorical score of three or more possible levels, a stringent dichotomization was applied. For instance, for the item named *int562*, participants were required to draw a triangle and then given a score as 0='correct', 1='partially correct' or 2="incorrect". In our analyses, this item was dichotomized as endorsement with a score of 1= 'correct' answer and non-endorsement with a score of 0= 'partially correct" or "incorrect' answer. Thus, a lower item score always corresponds to non-endorsement, i.e., difficulty with cognition, across the item pool.

The longitudinal trajectory of each individual item's score was examined for its association with the age of symptomatic onset, defined as the age with the first occurrence of CDR>0 over longitudinal follow-up. For each individual, we first computed the age of symptomatic onset, either observed within the follow-up or right censored if an individual was never rated as having CDR>0 during the entire follow-up. For each individual and each item, we then computed the age of <u>non-endorsement for each item</u>, defined as the age when the item was not endorsed and given a score of zero. An item's score usually fluctuated between endorsement and non-endorsement for some time before finally stabilizing at non-endorsement throughout the remaining time. With this type of fluctuation under consideration, the item-specific age of <u>non-endorsement</u> for each individual was treated as interval censored. The left side of the interval was the first age in the follow-up when the item was not endorsed, and the right side of the interval was the age at the first occurrence of <u>non-endorsement</u> after which the item remained not endorsed throughout the remaining follow-up. If non-endorsement of an item was observed right at baseline, the left side of the age interval of <u>non-endorsement</u> was defined as zero. If an individual endorsed an item during the entire follow-up, the left side of the interval would be set as the age of last assessment and thus the age of <u>non-endorsement</u> became right-censored. For each item, the item-specific age of <u>non-endorsement</u> and the age at symptomatic onset were correlated across all participants. A good item was expected to render a significant concordance correlation between the age of non-endorsement and the age of symptomatic onset, as measured by the Kendall's coefficient of concordance. This concordance correlation between the two interval censored variables was estimated through a bivariate smoothing of the joint density of the two variables (in logarithm scale) using a mixture of Gaussian densities fixed on a grid with weights determined by a penalized likelihood approach [24]. Items with a significant concordance correlation (p 0.05) were thereafter called informative for tracking early disease progression.

Estimating the longitudinal rate of functional and cognitive change

A participant's overall functional and cognitive ability, as evaluated by the clinical protocol and the psychometric battery, respectively, was calculated by the corresponding composite of z-scores from multiple components. A z-score of a component test in the clinical protocol or the psychometric battery was calculated across all the items belonging to the test, using the baseline mean and standard deviation of the component test. The average of z-scores across all components in the clinical protocol was calculated as the composite score to represent the overall functional outcome, and the average of z-scores across all components in the psychometric battery was calculated to represent the overall cognitive outcome. We also computed the composite score combining the clinical protocol and the psychometric battery by averaging the z-scores across all the components in the two protocols. We derived these composite scores in three different ways using: 1) all items, 2) informative items selected using Kendall's coefficient of concordance, and 3) the unselected non-informative items. Longitudinal changes in a participant's overall functional and cognitive performance were usually very subtle prior to symptomatic onset (first time occurrence of CDR>0), but the decline of the scores accelerated afterwards. In recognition of this, a piecewise random-intercept and random-slope linear mixed effects model was fit to the time segment from baseline to the symptomatic onset and then to the time segment after symptomatic onset for each of the three functional or cognitive composite scores. The estimated slopes (i.e., the longitudinal rate of change) both prior to and after symptomatic onset are reported. A negative slope indicates longitudinal cognitive decline.

Powering future clinical trials on early and preclinical AD

We further examined the sample sizes required to adequately power a future randomized clinical trial (RCT) using the proposed functional and cognitive composites from the informative items as the primary efficacy outcome variable. We considered both a prevention trial on asymptomatic individuals (i.e., prior to symptom onset) and a treatment trial on early symptomatic AD (encompassing both mild cognitive impairment due to AD and very mild AD dementia). For the prevention trial, cognitively normal individuals must first be identified as having elevated risk for symptomatic AD (i.e., preclinical AD). Cerebrospinal fluid (CSF) and neuroimaging biomarkers of AD now are used in major secondary prevention RCTs, including the Anti-Amyloid Treatment in Asymptomatic Alzheimer's (A4) trial[25], the Dominantly Inherited Alzheimer Network-Trials Unit (DIAN-TU) trials[26], and the Alzheimer's Prevention Initiative (API) trial[27], to identify persons with preclinical AD. However, the AD biomarkers may not identify all individuals who will eventually develop symptomatic AD. Hence, we decided to use the longitudinal functional and cognitive data prior to symptomatic onset from the 240 converters who became CDR>0 during follow up to power a future prevention trial, and the longitudinal data after symptomatic onset to power a treatment trial on early symptomatic AD by estimating the corresponding longitudinal rates of functional and cognitive composites. Specifically, the estimates from the first time segment prior to symptomatic onset in the piecewise linear mixed model were used as the placebo effect for the prevention trial, and those estimated from the second time segment after symptomatic onset were used as the effect in the placebo arm for the therapeutic trial on early symptomatic AD. The sample sizes for detecting a range of effect sizes (ES) for a novel treatment with 80% power were calculated using a

standard normal test [28]. For comparison purpose, similar sample size calculations for the treatment trial were also done on the functional and cognitive composites using all the items as well as using the non-informative items only.

Statistical analyses

All statistical analyses were implemented in SAS® (version 9.4, SAS Institute, Cary, NC). All tests were two-sided and statistical significance was defined at the 5% level. Main analyses were also repeated on the converters who received an etiologic diagnosis of AD for comparison.

Results

735 individuals at least 65 years old and cognitively normal (with CDR 0) at baseline were assessed annually up to 29 years of follow-up (mean follow-up=6.79 years, SD=5.55 years). Baseline characteristics of the participants are summarized in Table 1. Over the course of follow-up, 240 individuals converted from being cognitively normal (i.e., CDR = 0) at baseline to an early dementia with CDR 0.5, and were termed 'converters' thereafter. The average age of all participants is 74.30 years (SD=8.86 years) at baseline. 62.31% of the participants are female. Among a total of 562 items, including 374 in the cognitive battery and 188 in the clinical protocol, only 292 items were identified as informative, with 174 of them in the cognitive battery and 118 in the clinical protocol. Table 2 shows the number of informative items from each of the tests in the two instruments. Additional analyses restricted to the converters who received a etiologic diagnosis of AD resulted in largely consistent findings. Supplemental Table 1 in the Appendix lists all individual items that were found to be informative, i.e., predicting the age of symptom onset.

For each composite, the piecewise linear mixed effects models resulted in two estimates to the slopes (i.e., longitudinal rates of change) along with their associated standard error (SE): the preclinical rate of change (Slope 1) from baseline to symptomatic onset, and the rate of disease progression (Slope 2) after symptomatic onset, as presented in Table 3. Results in Table 3 indicate that the estimated longitudinal rates of change in both the functional composite and cognitive composite, using the items identified as non-informative, are positive prior to symptomatic onset, indicating that asymptomatic individuals exhibited certain degree of learning over repeated testing of these items. When the informative items alone were used to form the functional and cognitive composites, however, their longitudinal rates of changes prior to the dementia symptom onset were negative. This implies that the informative items started to capture some of the subtle preclinical disease progression prior to symptomatic onset. Interestingly, the estimated longitudinal rates of change in both the functional composite and cognitive composites, using the items from the entire battery with both informative and non-informative ones, were also positive prior to symptomatic onset. This suggests that a large portion of the non-informative items in the battery easily overwhelmed the informative ones, leading to a collective learning effect over repeated testing for the entire batteries. Further, the longitudinal rates of change on both cognitive and functional composites after symptomatic onset using either all items in the item pool, or the informative items alone, or non-informative items alone, were all negative, suggesting that

the entire clinical protocol and the psychometric battery effectively tracked progression following symptomatic onset. More importantly, the longitudinal rates of change after symptomatic onset estimated using informative items only were larger in magnitude, both for cognitive and functional composites, than those estimated using all items, which were in turn larger (in magnitude) than the estimates using non-informative items.

Because cognitively normal individuals exhibited learning effects prior to symptomatic onset on the functional and cognitive composites using items identified as non-informative or using the entire item pool, these composites are hence of limited utility in designing future prevention trials on asymptomatic individuals. Given that the functional composite and cognitive composite, using only items identified as informative, did show decline longitudinally prior to the symptom onset, we explored the feasibility of using these two composites to power a future prevention trial of AD on asymptomatic individuals. In a hypothetical two-arm future prevention trial with a 1:1 sample size ratio and annual functional and cognitive assessments over 4 years, we calculated the total sample size with 80% power using the functional and cognitive composites from the informative items alone. Table 4a presents the results to detect a set of ES, expressed as percentages of improvement by the active treatment arm over the longitudinal rates of functional and cognitive progression in the placebo. After symptomatic onset, the larger magnitudes of the rate of progression using the functional and cognitive composites from the informative items (in comparison to those using either the entire item pool or the non-informative items alone) suggest that they may improve (i.e., reduce) the sample sizes for future therapeutic trials in early symptomatic AD. We again assumed a two-arm RCT treating participants with early symptomatic AD with a 1:1 sample size ratio between a novel therapeutic arm and a placebo arm and annual assessments over 4-year follow-up. Table 4b presents the total sample size required to detect a set of ES with 80% power using the functional and cognitive composites from the informative items alone, and for comparison, the total sample sizes required using the functional and cognitive composites from the non-informative items alone, as well as from the entire pools of items available. Given the same ES (in %) to be detected and the same statistical power, the sample sizes using the functional composite of the informative items alone from the clinical protocol as the efficacy outcome are less than a third of those using the non-informative items. For the cognitive composite as the primary efficacy endpoint, using informative items in the cognitive battery also leads to a dramatic reduction in the sample size when compared to using the non-informative items. For example, for an effect size of 50%, the RCT with the functional endpoint can be adequately powered with a total 940 participants using informative items in the clinical protocol alone, a more than 70% reduction or about 7% reduction to the sample sizes using non-informative items alone (n=3174 participants) or the entire item pool (n=1010 participants), respectively. For the cognitive endpoint with an ES of 50%, the RCT can be adequately powered with a total of 1250 participants using informative items alone, only 11.5% of the sample size using non-informative items alone (n=10830 participants) and 66% of the sample size using the entire item pool (n=1884 participants).

Discussion

A major paradigm shift in RCTs of investigational drugs for AD is the current focus on the preclinical or very early symptomatic stages. Major secondary prevention RCTs, including the A4 trial, the DIAN-TU trials, and the API trial, are currently ongoing. Given that the recently revised Food and Drug Administration (FDA) guidelines for RCTs for early AD mandate that treatments only be approved if they demonstrate cognitive and functional benefits, a well designed future RCT requires not only longitudinal cognitive and functional assessments, but also the linkage between the preclinical or early symptomatic stages and the rate of cognitive and functional decline in the placebo arm. A common challenge to all ongoing prevention RCTs on preclinical AD or treatment RCTs for early symptomatic AD is the optimum cognitive and functional endpoints that can best power the trials[29]. The prevention or early treatment RCTs on AD mandate instruments that are much more sensitive and specific to the subtle early and preclinical longitudinal changes of AD than the existing ones (e.g., ADAS-cog[30]). As a matter of fact, several recent studies have failed to detect significant decline in the placebo groups of RCTs on mild cognitive impairment or even established mild to moderate AD populations with the existing cognitive and functional instruments. The situation will get even worse when it comes to designing primary preventive trials for AD. Because of the lack of highly reliable and well validated sensitive and specific cognitive and functional tests, it is difficult to establish *a priori* who in a population will ultimately develop AD symptoms and over what time frame. RCTs must therefore study a huge number of individuals for many years in order to guarantee that a significant number in any treatment arm will develop dementia symptoms so that meaningful statistical conclusions can be drawn. Such large and long duration studies are prohibitively costly and prone to high dropouts.

A major reason that currently used clinical and cognitive instruments in AD research lack sensitivity and specificity to identify individuals who are at high risk of developing dementia symptoms is the ceiling and floor effects [3] as well as the learning effect due to the repeated administering of the same instruments. The fact that the A4, the DIAN-TU, and the API trials have all chosen to employ different cognitive endpoints highlights an urgent need to comprehensively analyze the longitudinal item level functional and cognitive data and to inform the RCTs with optimum cognitive and functional endpoints and adequate statistical power. We hence analyzed the longitudinal item level data from the clinical assessment protocol and the cognitive battery administered at the WU Knight ADRC to identify informative items that were most predictive of early symptomatic onset. We found that approximately half of the items among a total of 562 items were uninformative to predict symptomatic onset, likely because these items showed very little or only random changes over the preclinical time window prior to the onset of symptoms. Unsurprisingly, we found that, for both the cognitive and the functional composites that were formed using the non-informative items, the estimated longitudinal rate of change prior to symptomatic onset was positive, suggesting a learning effect during the preclinical stage of the disease. Importantly, we found that for both the cognitive and the functional composites using the informative items, the estimated longitudinal rate of change prior to symptomatic onset became negative, although the rate of changes for these composites using all items from the entire batteries

was still positive. This suggests that the contamination of the batteries by non-informative items prevents adequate tracking of subtle preclinical disease progression. After symptomatic onset, as expected, we found that the functional and cognitive composites using the informative items alone rendered a larger rate of decline (in magnitude) in comparison to the corresponding composites using only the non-informative items, as well as using the entire item pool with both informative and non-informative items.

Using these results to design a future prevention trial on asymptomatic individuals who will eventually develop symptomatic AD, we found that even with the informative items alone, the sample sizes required to adequately power such a trial remain formidable. For example, for detecting an effect size of 50% improvement as compared to placebo, the prevention RCT needs to enroll a total of 302322 participants using the functional endpoint, and a total of 6078 participants using the cognitive composite. These numbers are very hard to achieve, and much larger than the sample sizes currently estimated in some of the ongoing secondary prevention trials on elderly individuals. These large sample sizes also imply that, although the informative items from many standard cognitive tests are able to capture some of the cognitive decline prior to the symptom onset, the magnitude of decline captured is too small and the variation is too large. Hence, to best design future prevention trials on AD, completely new cognitive items and tests may need to be developed. These new items and tests should be designed in a way that will specifically target the cognitive traits and domains most vulnerable of very early change during the preclinical stage of AD, and minimize the learning effect over repeated administering. Significant resources are needed to develop such preclinical cognitive batteries and test their psychometric properties on individuals at high risk for preclinical AD.

For treatment trials on early symptomatic AD, however, sample sizes calculated using informative items only were dramatically smaller than those using non-informative items or using the entirety of the items in the batteries. In addition, these sample sizes are more feasible to achieve. For example, for detecting an effect size of 50%, the RCT with the functional composite can be adequately powered with a total 940 participants using informative items of the clinical protocol alone, and with a total of 1250 participants using the cognitive composite from informative items alone. When a meta-composite is used to combine both the cognitive and functional composites from the informatics items alone, the RCT can be adequately powered with 764 individuals with early symptomatic AD.

Major strengths of the study include a relatively large sample size of cognitively normal elderly individuals (at baseline) who were carefully characterized by annual clinical and cognitive assessments over a relatively long follow-up of up to 29 years. The relatively large number of individuals (n=240) who developed dementia symptoms during the follow-up allowed reasonably accurate estimates to the longitudinal rates of change in both function and cognition domains, both prior to (for designing prevention trials) and after symptom onset (for designing early treatment trials). Our longitudinal item analysis is also novel in the sense it directly correlated the item level changes in scores to the onset of symptoms. Realizing the vast variabilities in item-level scores over time, we analyzed the item scores over time as interval-censored variables, and used the Kendall's coefficient of concordance

to quantify the correlation between the interval censored variables and the age of symptomatic onset.

Limitations of this study include the convenience nature of the study sample, which were mostly restricted to the elder adult population in the St. Louis metropolitan area and may prevent the findings from being generalized to the more general population. Whereas the WU Knight ADRC clinical assessment protocols and cognitive battery are comprehensive, covering all major cognitive domains, they did not include some cognitive scales that are often used in the treatment trials of AD, such as the ADAS-cog, and hence the items from such scales cannot be evaluated for their utility in AD prevention trials.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgement

References

1. Price JL, Morris JC: Tangles and plaques in nondemented aging and "preclinical" Alzheimer's disease. Ann Neurol 1999, 45(3):358–368. [PubMed: 10072051]

2. Sperling RA, Aisen PS, Beckett LA et al.: Toward defining the preclinical stages of Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. Alzheimers Dement 2011, 7(3):280–292. [PubMed: 21514248]

3. Sheehan B: Assessment scales in dementia. Ther Adv Neurol Disord 2012, 5(6):349–358. [PubMed: 23139705]

4. Twamley EW, Ropacki SA, Bondi MW: Neuropsychological and neuroimaging changes in preclinical Alzheimer's disease. J Int Neuropsychol Soc 2006, 12(5):707–735. [PubMed: 16961952]

5. Morris JC: The Clinical Dementia Rating (CDR): current version and scoring rules. Neurology 1993, 43(11):2412–2414.

6. JA SJaY: Geriatric Depression Scale (GDS): Recent evidence and development of a shorter version Clinical Gerontology : A Guide to Assessment and Intervention. In. New York: The Haworth Press; 1986: 165–173.

7. Hughes CP, Berg L, Danziger WL, Coben LA, Martin RL: A new clinical scale for the staging of dementia. Br J Psychiatry 1982, 140:566–572. [PubMed: 7104545]

8. Pfeiffer E: A short portable mental status questionnaire for the assessment of organic brain deficit in elderly patients. J Am Geriatr Soc 1975, 23(10):433–441. [PubMed: 1159263]

9. Folstein MF, Folstein SE, McHugh PR: "Mini-mental state". A practical method for grading the cognitive state of patients for the clinician. J Psychiatr Res 1975, 12(3):189–198. [PubMed: 1202204]

10. Katzman R, Brown T, Fuld P et al.: Validation of a short Orientation-Memory-Concentration Test of cognitive impairment. Am J Psychiatry 1983, 140(6):734–739. [PubMed: 6846631]

11. Goodglass HaK E: The Assessment of Aphasia and Related Disorders, 2nd Edition Philadelphia: Lea & Febiger; 1972.

12. Cummings JL: The Neuropsychiatric Inventory: assessing psychopathology in dementia patients. Neurology 1997, 48(5 Suppl 6):S10–16.

13. Pfeffer RI, Kurosaki TT, Harrah CH, Jr., , Chance JM, Filos S: Measurement of functional activities in older adults in the community. J Gerontol 1982, 37(3):323–329. [PubMed: 7069156]

14. Ferman TJ, Smith GE, Boeve BF et al.: DLB fluctuations - Specific features that reliably differentiate DLB from AD and normal aging. Neurology 2004, 62(2):181–187. [PubMed: 14745051]

15. Blessed G, Tomlinson BE, and Roth M. : The association between quantitative measures of dementia and of senile change in the cerebral grey matter of elderly subjects. British Journal of Psychiatry 1968, 1114:797–811.

16. Wechsler D: Wechsler Memory Scale-Revised. In. San Antonio, Texas: Psychological Corporation; 1987.

17. Wechsler DaS CP: Wechsler Memory Scale In. New York: Psychological Corporation; 1973.

18. Benton AL: The Revised Visual Retention Test: Clinical and experimental applications. In. New York: Psychological Corporation; 1963.

19. Goodglass HaK E, : Boston Diagnostic Aphasia Examination Booklet. In., 3rd ed. edn. Philadelphia: Lea & Febiger; 1983.

20. Wechsler D: Wechsler Adult Intelligence Scale In. New York: Psychological Corporation; 1955.

21. Grober E, Buschke H, Crystal H, Bang S, Dresner R: Screening for dementia by memory testing. Neurology 1988, 38(6):900–903. [PubMed: 3368071]

22. Morris JC, Weintraub S, Chui HC et al.: The uniform data set (UDS): Clinical and cognitive variables and descriptive data from Alzheimer disease centers. Alz Dis Assoc Dis 2006, 20(4): 210–216.

23. American Psychiatric Association: Diagnostic and Statistical Manual of Mental Disorders, 4 edn. Washington DC: American Psychiatric Association; 2000.

24. Bogaerts K, Lesaffre E: Estimating local and global measures of association for bivariate interval censored data with a smooth estimate of the density. Statistics in Medicine 2008, 27(28):5941– 5955. [PubMed: 18623606]

25. Sperling RA, Rentz DM, Johnson KA et al.: The A4 Study: Stopping AD Before Symptoms Begin? Science Translational Medicine 2014, 6(228).

26. Mills SM, Mallmann J, Santacruz AM et al.: Preclinical trials in autosomal dominant AD: Implementation of the DIAN-TU trial. Rev Neurol-France 2013, 169(10):737–743.

27. Ayutyanont N, Langbaum JBS, Hendrix SB et al.: 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. J Clin Psychiat 2014, 75(6):652–660.

28. Xiong C, Zhu K, Yu K, Miller JP: Statistical In: Handbook of Statistics. Volume 27, edn. Edited by Rao CRMJ, Rao DC. London: Elsevier; 2007: 429–463.

29. Ringman JM, Grill J, Rodriguez-Agudelo Y, Chavez M, Xiong C: Commentary on "a roadmap for the prevention of dementia II: Leon Thal Symposium 2008." Prevention trials in persons at risk for dominantly inherited Alzheimer's disease: opportunities and challenges. Alzheimers Dement 2009, 5(2):166–171. [PubMed: 19328453]

30. Mohs RC, Knopman D, Petersen RC et al.: Development of cognitive instruments for use in clinical trials of antidementia drugs: additions to the Alzheimer's Disease Assessment Scale that broaden its scope. The Alzheimer's Disease Cooperative Study. Alzheimer Dis Assoc Disord 1997, 11 Suppl 2:S13–21. [PubMed: 9236948]

Table 1.

Baseline Characteristics of the cohort (N = 735, mean/SD follow-up time = 6.79/5.55 y)

Variable	All participants (N = 735)	Converters (N = 240)	Non-converters (N = 495)	P-value[*]
Age (N=735) (mean, SD)	74.30 (8.86)	77.57 (8.45)	72.71 (8.63)	<0.0001
Gender (N=735) (number, % of female)	458 (62.3%)	142 (59.2%)	316 (63.8%)	0. 225
Education (unit: years) (N=735) (mean, SD)	14.88 (3.08)	14.69 (3.48)	14.97 (2.87)	0.251
Race (N=687) Caucasian (number, %)	629 (91.6%)	212 (94.6%)	417 (90.1%)	0.121
African American (number, %)	50 (7.3%)	10 (4.5%)	40 (8.6%)	
other (number, %)	8 (1.2%)	2 (0.9%)	6 (1.3%)	
APOE4 positive (N=685) (number, %)	207 (30.2%)	72 (32.5%)	135 (29.2%)	0.425

[*] The P-values for age, educ were from two-sided two-sample t-test and others from two-sided Fisher's exact test.

Table 2.

Numbers of total and informative items in the WU Knight ADRC clinical protocol and cognitive battery (tests in italic are no longer in use).

Instrument	Test names	number of all items	Number of informative items
Clinical protocol: Participants	Geriatric Depression Scale	15	7
	Judgment and Problem Solving	9	8
	Memory Evaluation	21	17
	Mini Mental State Examination	26	20
	Short Blessed Test	8	6
	Assessment of Aphasia	29	16
	Drawing	3	2
	Sum of total	**111**	**76**
Clinical Protocol: Informants	Geriatric Depression Scale	15	6
	Judgment and Problem Solving	5	4
	Memory Evaluation	10	1
	Neuropsychiatric Inventory	12	11
	Functional Assessment Questionnaire	10	7
	Ferman test	4	2
	Personal Care	4	4
	Community Affaires	5	3
	Orientation and Daily Activities	10	4
	Depressive features	2	0
	Sum of total	**77**	**42**
Cognitive battery	Associate Leaning	30	16
	Benton Visual Retention	*10*	*6*
	Boston naming	*60*	*14*
	Boston naming	*30*	*21*
	Boston naming	30	20
	Digits backward a	6	1
	Digits backward b	6	2
	Digits forward a	7	3
	Digits forward b	7	1
	Free and cued selective reminding	48	33
	logical memory, delayed recall (original)	*24*	*9*
	logical memory, delayed recall (revised)	25	6
	logical memory, immediate recall (original)	*24*	*8*
	logical memory, immediate recall (revised)	25	10
	Mental control	3	1
	WAIS block design	10	3

Instrument	Test names	number of all items	Number of informative items
	WAIS information	29	20
	Sum of total	**374**	**174**

Table 3.

Estimates to the longitudinal rate of change prior to and after symptomatic onset, labeled as "Slope 1" and "Slope 2", respectively, using all items in the item pool, informative items only, and non-informative items only.

Test	Estimates (SE)	All items	Informative Items only	Non-informative items only
Functional composite (from the clinical protocol)	Slope 1	0.000375 (0.000211)	−0.00014 (0.000166)	0.0000935 (0.000304)
	Slope 2	−0.02053 (0.001536)	−0.02266 (0.001722)	−0.01624 (0.001308)
Cognitive composite (from the cognitive battery)	Slope 1	0.000268 (0.000416)	−0.00177 (0.000469)	0.002734 (0.000513)
	Slope 2	−0.00651 (0.00777)	−0.01110 (0.001051)	−0.00315 (0.000636)
Functional and cognitive composites combined	Slope 1	−0.00031 (0.000389)	−0.00242 (0.000430)	0.002039 (0.000455)
	Slope 2	−0.00663 (0.000837)	−0.01068 (0.001033)	−0.00290 (0.000652)

[*] Slope 1: slope before dementia onset with the first CDR>0 diagnosis; Slope 2: slope after symptomatic onset.

Table 4.

(a) Total sample size needed for a future prevention trial of AD to detect the effect sizes (% improvement) in the rate of disease progression prior to symptomatic onset, using the functional and cognitive composites formed by the *informative* items alone from the clinical protocol and the psychometric battery, respectively. (sample size results were only provided using the functional and cognitive composites from the informative items alone, because those using all items in the entire battery and the non-informative items alone yielded positive estimates to the longitudinal rate of change in the placebo arm, which indicated a learning effect over repeated testing).

(b) Total sample sizes needed for future treatment RCTs in early symptomatic AD to detect the effect sizes (% improvement) in rate of disease progression after symptomatic onset, using the functional and cognitive composites formed by all items in the item pool ("All"), informative items alone ("Informative"), and non-informative items alone ("Non-informative").

Test	Effect size (%)	Total N
Functional composite (from tde clinical protocol)	20	1.89E+06
	30	839778
	40	472376
	50	302322
	60	209946
	70	154248
	80	118096
	90	93312
Cognitive composite (from tde cognitive battery)	20	37972
	30	16878
	40	9496
	50	6078
	60	4222
	70	3102
	80	2376
	90	1878
Functional and cognitive composites combined	20	18892
	30	8398
	40	4726
	50	3026
	60	2102
	70	1544
	80	1184
	90	936

Test	effect size (%)	All N Total	Informative N Total	Non-informative N Total
Functional composite	20	6294	5854	19824
	30	2798	2604	8812
	40	1576	1466	4958
	50	1010	940	3174
	60	702	654	2206
	70	516	480	1620
	80	396	368	1242
	90	314	292	982
Cognitive composite	20	11758	7802	67666
	30	5228	3470	30076
	40	2942	1952	16918
	50	1884	1250	10830
	60	1308	870	7522
	70	962	640	5526
	80	738	490	4232
	90	584	388	3344
	20	7588	4764	54830
	30	3374	2120	24370
	40	1900	1194	13710
Functional and cognitive composites combined	50	1216	764	8776
	60	846	532	6094
	70	622	392	4478
	80	478	300	3430
	90	378	238	2710