

Profiles of cognitive change in preclinical and prodromal Alzheimer's disease using change-point analysis

Owen A Williams^{a,b}, Yang An^a, Nicole M Armstrong^a, Melissa Kitner-Triolo^a, Luigi Ferrucci^c,
Susan M Resnick^a

^aLaboratory of Behavioral Neuroscience, National Institute on Aging, Baltimore, Maryland
21224, USA

^bDepartment of Experimental Psychology, University of Oxford, Oxford, UK

^cLongitudinal Studies Section, Translational Gerontology Branch, National Institute on Aging,
Baltimore, MD, USA

Running title: Preclinical cognitive change-points

Corresponding author: Susan M Resnick, resnicks@nih.gov, 410-558-8618

Word count: 4341

Abstract word count: 211

Number tables: 5

Number figures: 2

Abstract

Background: Alzheimer's disease is now understood to have a long preclinical phase in which pathology starts to accumulate in the absence of clinical symptoms. Identifying the temporal stages of accelerated cognitive decline in this phase may help in developing more sensitive neuropsychological tools for early screening of preclinical cognitive decline. Change-point analyses are increasingly used to characterize the temporal stages of accelerated cognitive decline in the preclinical stages of Alzheimer's disease (AD). However, statistical comparisons of change-points between specific cognitive measures have not been reported.

Objective: To characterize and compare the temporal stages of accelerated decline in performance on multiple cognitive tests in a sample of participants from the Baltimore Longitudinal Study on Aging (BLSA) who later developed Alzheimer's disease.

Methods: 165 older adults (baseline age range: 61.1-91.2) from the Baltimore Longitudinal Study of Aging developed AD during follow-up. Linear and non-linear mixed models were fit for 11 cognitive measures to determine change-points in rates of decline before AD diagnosis. Bootstrapping was used to compare the timing of change-points across cognitive measures.

Results: Change-points followed by accelerated decline ranged from 15.5 years (Standard Error (S.E.) = 1.72) for Card Rotations to 1.9 years (S.E. = 0.68) for the Trail-Making A test before AD diagnosis. Accelerated decline in Card Rotations occurred significantly earlier than all other measures, including learning and memory measures.

Conclusion: Results suggest that visuospatial ability, as assessed by Card Rotations, may have the greatest utility as an early predictive tool in identifying preclinical AD.

Key Words: Preclinical, Prodromal, Cognitive Decline, Change-point analysis, Visuospatial ability, Alzheimer's Disease

Introduction

Alzheimer's disease (AD) is a progressive disease with a long preclinical phase in which pathological markers are present for years and even decades before clinical symptoms [1].

Decline in episodic memory is a hallmark of AD, but other cognitive domains are also vulnerable to AD [2]. Understanding the temporal stages of the early acceleration of declines in various cognitive domains in preclinical AD is important for identifying individuals vulnerable towards accumulating AD pathology and for characterizing AD progression prior to symptom onset.

The early preclinical phase of AD is characterized by beta amyloid ($A\beta$) and phosphorylated tau accumulation with subsequent acceleration of brain atrophy in the absence of clinical symptoms [1, 3-5]. With multiple anti-amyloid clinical trials failing to show that removal of $A\beta$ is associated with improved cognitive outcomes [6], one argument is that patients are not being targeted early enough [6, 7]. In the continuum of AD proposed by Jack Jr et al., preclinical AD is followed by mild cognitive impairment (MCI) [8] and then AD [9]. MCI patients exhibit faster cognitive decline compared to healthy controls for a range of cognitive domains including memory, executive function, attention and verbal fluency. However, the nature of accelerated decline in the preclinical stage of AD is less clearly understood. The failures of clinical trials at later disease stages have led to increasing focus on the earlier phases of disease, including the asymptomatic preclinical stage, with the hope that treatments at this stage may be more effective [10, 11]. Thus, it is critical to define the earliest, and possibly subtle, changes in cognitive performance in preclinical AD to identify individuals who would have the greatest potential to benefit from clinical interventions. To fully characterize cognitive changes in preclinical AD, it is important to examine a broad range of cognitive domains and neuropsychological measures that may be sensitive to the earliest changes.

One way of investigating the timing of cognitive decline prior to clinical AD diagnosis is to use change-point analyses. Change-point methods align participants by anchoring them at time of diagnosis to then examine trajectories of variables of interest retrospectively for timepoints of change prior to clinical diagnosis. Change-points are identified using piece-wise linear components separated by knots delineating between intervals with differing rates of change [12, 13]. Previous studies using change-point analyses in AD focused on verbal memory [14, 15], reporting steeper declines in Immediate Recall, measured by the picture version of the Free and Cued Selective Reminding test, between 1 and 8.1 years before clinical diagnosis. However, memory is not the only cognitive domain subject to decline prior to AD onset [2] and other domains have shown early change-points. In a systematic review of change-point studies in dementia and AD [16], the measure with the earliest change-point, at 9.6 years before AD diagnosis, was the Block design test assessing visuospatial ability [17]. Measures associated with other domains, i.e., language fluency and executive function showed change-points detected at 6.8 years [18] and 2.9 years [15] prior to AD diagnosis respectively.

While the systematic review by Karr et al. [16] allows for a cursory comparisons of change-points between measures associated with various cognitive domains, the authors highlight various methodological differences between studies that make it difficult to draw conclusions from such comparisons. For example, the maximum length of longitudinal testing prior to AD diagnosis ranged from 9-30 years with frequency of visits varying between studies. Furthermore, the mean baseline ages in all studies ranged from 70-82 years, and analyses were adjusted by different sets of covariates. Therefore, the temporal sequence of changes in different cognitive domains in preclinical AD remains unclear.

To elucidate the temporal sequence of cognitive changes prior to clinical AD onset, we investigated changes in rates of decline on multiple cognitive measures, representing specific cognitive domains, in individuals who eventually developed AD. The aims of this study were to identify how many change-points best characterize the trajectories of change in performance on several cognitive measures and to find the earliest point in time before AD diagnosis that accelerated decline in performance could first be detected for each measure.

Materials and Methods

Participants

Participants were from the Baltimore Longitudinal Study of Aging (BLSA), a longitudinal study started in 1958 [19]. Participants were community-dwelling volunteers who were healthy at enrollment. During each visit, they received comprehensive psychological evaluations. For this study, we selected participants who were diagnosed with AD during follow-up and only used data from visits when they had complete neuropsychological testing data across the 11 measures investigated. As such, data from the 11 measures were collected concurrently and thus have the same sample size. Research protocols were approved by local institutional review boards, and all participants gave written informed consent at each visit. Data from the BLSA are available on request from the BLSA website (blsa.nih.gov). All requests are reviewed by the BLSA Data Sharing Proposal Review Committee and are also subject to approval from the NIH institutional review board.

Clinical diagnosis of Alzheimer's disease

Clinical and neuropsychological data from each participant were reviewed at a consensus case conference if their clinical dementia rating score [20] was 0.5 or greater or if they had more than

three errors on the Blessed Information-Memory-Concentration Test [21], and participants were evaluated by case conference upon death or withdrawal. MCI status was determined using the Petersen criteria [22]. Diagnoses of dementia and AD were based on criteria outlined in the Diagnostic and Statistical Manual of Mental Disorders, third edition, revised [23] and the National Institute of Neurological and Communication Disorders and Stroke – Alzheimer’s Disease and Related Disorders [24], respectively.

Neuropsychological measures

Participants were administered a comprehensive battery of cognitive measures assessing verbal learning and memory, figural memory, attention and processing speed, executive function, language, and visuospatial ability. Here, we provide a summary of each measure used as outcomes in the present study.

The California Verbal Learning Test (CVLT) [25] assesses episodic verbal learning and memory. There are 5 learning trials of 16 shopping items, presented orally, with four items from each of four semantic categories. The sum of the five trials provides a measure of new learning and immediate free recall. In addition, short- and long-delayed free recall, short- and long-delayed cued recall, and recognition memory are assessed. The two measures used were: total number of items recalled across the five immediate free recall trials (CVLT-IMM), and long-delay free recall (CVLT-LD), with maximum scores of 80, and 16, respectively.

The Benton Visual Retention Test (BVRT) [26] measures visual constructional skill and short-term figural memory. Participants study 10 line-drawings, including one to three geometric figures, for 10 seconds each, and then immediately reproduce them from memory using pencil

and paper. The designs become more difficult over the 10 trials. The measure used was the total number of errors.

The Trail-Making Test Parts A (TMT-A) and B (TMT-B) [27] assess attention, concentration, visuomotor scanning, perceptuomotor speed, working memory, and set-shifting. TMT-A involves drawing a line to connect randomly arranged numbers from 1 to 25 in sequential order. TMT-B involves connecting randomly arranged numbers and letters in alternating sequence (e.g., 1-A-2-B . . .). Time to completion for each test (in seconds) was used in the present study.

Letter [28] and Category [29] Fluency are measures of fluent language production and executive function. Participants were given 60 seconds to generate as many words as possible beginning with specific letters and specific categories. The mean numbers of correct words generated in 60 seconds, across the three trials each for letter and category fluency, were the measures of interest.

The Boston Naming Test (BNT) [30] is a measure of object recognition and semantic retrieval. Participants identify and name a series of line drawings of objects, beginning with common objects and ending with infrequent objects. The measure used was the number of words out of 60 correctly named without cues.

WAIS-R Similarities measures verbal concept formation and abstract reasoning [31].

Participants are asked how 14 pairs of two items are similar, starting with concrete items and becoming increasingly abstract. The measure of interest was the total score out of a maximum of 28.

A modified version [32] of the Educational Testing Service Card Rotations test was used to measure visuospatial ability. Participants were presented with a target figure and eight alternative figures in the same row. Subjects marked images that could be rotated in plane to match the

target, but not those that were mirror image figures. The total number correct minus total number incorrect across the two parts (14 targets per part) was the measure of interest, with a maximum score of 224.

The Mini Mental State Examination (MMSE) [33] assesses mental status, including orientation to time and place, immediate and delayed recall, attention and calculation, and language. Total score out of 30 was the measure of interest.

The reliability of cognitive measures used in the present study was assessed using the intra-class correlation (ICC). The ICC for each cognitive measure was computed by partitioning the variance of the data into between and within-individual variance after adjusting for longitudinal aging effects using linear mixed effects models.

Statistical analyses

To find the number of change-points where the rate of longitudinal decline changes significantly and the timing of these change-points relative to AD diagnosis, a series of linear and nonlinear mixed models with increasing complexity were fit with each of the cognitive measures as the outcome and the time (in years) to diagnosis of AD as the main predictor. We started with a no change-point model and then tested a one-change-point model and finally a two change-points model. The two-change-point model function is given by

$$Y_{ij} = \beta_0 + b_{0i} + \beta_1 time_{ij} + \beta_2 (time_{ij} - c_1)_+ + \beta_3 (time_{ij} - c_2)_+ + \varepsilon_{ij}$$

where $(x)_+ = x$, $x > 0$ and $(x)_+ = 0$, $x < 0$.

Y_{ij} is the cognitive outcome for i^{th} subject and j^{th} assessments, $time_{ij}$ is time to AD diagnosis

for i^{th} subject and j^{th} assessments. c_1 is first change-point, c_2 is second change-point, β_0 is intercept, β_1 is the slope before the first change-point, $(\beta_1 + \beta_2)$ is the slope between first and second change-point, $(\beta_1 + \beta_2 + \beta_3)$ is the slope after the second change-point. b_{0i} is a random effect that follows a normal distribution with mean 0 and standard deviation of σ . ϵ_{ij} is the error term. Among the three models, a no-change-point model is a linear model; one-change-point and two change-point models are considered models *with nonlinearity in parameters* because unknown parameters enter the models nonlinearly. All models included baseline age, sex (male vs. female), race (white vs. non-white), and years of education as main effects covariates. Model selections were based on the likelihood ratio test. The best model fit tells us how many change-points, if any, there are for each cognitive measure. We also tested 3-change-point models but in all but one case, these did not converge. In the sole case that converged, no significant improvement in fit was observed.

We used a bootstrapping approach to estimate our final parameter point estimates and standard errors so that these estimates can be captured more accurately, and robustness of the results can be assessed. Bootstrapping also provides us with the ability to compare change-points across different cognitive measures statistically.

Specifically, each resample step randomly draws subjects from the original dataset with replacement to get a new dataset with the same size. In our analysis, we used a total of 500 new bootstrapped datasets. The change-point model was fit to each of the 500 bootstrapped datasets separately, and the results give a distribution for each of the parameters in the model. Statistical inferences are made from these distributions. The difference between estimates of the change-points was computed in each iteration of the bootstrapped sample. The Wald test was then conducted on the whole distribution of differences to assess whether change-points were

statistically different between cognitive tests. The models were fit using PROC NLMIXED in SAS 9.4 (Cary, NC).

Results

Sample characteristics are shown in Table 1. The sample consisted of 165 participants with an AD diagnosis with a total of 988 visits. Average baseline age was 76.5 years (standard deviation, [SD]=7.4), the average follow-up interval was 8.3 years (SD=6.0, range=0-24.1), and average age at AD diagnosis was 86.5 years (SD=6.1). Eighty-three (50.3%) participants were female, and 85.5% of the sample were white. The average years of education was 16.7 years (SD=2.6, range=8-21).

Reliability of cognitive measures

The intra-class correlations for different cognitive measures vary from 0.55 to 0.84 with the majority of the measures with ICC higher than 0.7. Table 2 shows the ICC for each cognitive measure.

Change-point model comparisons for each cognitive measure

Table 3 contains the results of the model fit statistics (likelihood ratio test) comparing the fit of three models for each cognitive measure. Models with 1-change-point provided better fit compared with no change-point models for all cognitive measures. Models with 2-change-points provided a better fit for CVLT-IMM, CVLT-LD, Category Fluency, Letter Fluency, BNT, Similarities, Card Rotations, and the MMSE. However, for the BVRT, TMT-A, and TMT-B, the 2-change-point model did not significantly improve the model fit, indicating a 1-change-point model was the best fitting model for these measures.

Temporal position of change-points and subsequent rates of decline

Table 4 shows the results of parameter estimates from bootstrapping the best fitting model for each cognitive measure, including the estimated change-points, the rate of change at each segment of the trajectory, and corresponding standard errors (SE). Figure 1 shows the estimated trajectories superimposed over the raw data for each cognitive measure.

Learning and memory

CVLT-IMM: The estimated first change-point was 11.65 (SE=0.80) years before AD diagnosis, when the trajectory transitioned from non-significant increase in performance over time (0.23 items per year, SE=0.28) to significant moderate decline in performance (-1.44 items per year, SE=0.25). The second change-point was 2.80 (SE=0.78) years before AD diagnosis, when decline accelerated again (-3.74 items per year, SE=0.48).

CVLT-LD: The estimated first change-point was 7.58 (SE=0.49) years before AD diagnosis, when the trajectory transitioned from trending-significant minor decrease in performance over time (-0.080 items per year, SE=0.041) to significant moderate decline in performance (-0.48 items per year, SE=0.13). The second change-point was 4.21 (SE=0.34) years before AD diagnosis, when decline accelerated again (-1.00 items per year, SE=0.091).

BVRT: The estimated change-point was 4.83 (SE=0.81) years before AD diagnosis, when the trajectory transitioned from significant modest decline in performance over time (0.36 errors per year, SE=0.034) to significant accelerated decline in performance (0.99 errors per year, SE=0.12).

Attention and executive function

TMT-A: The estimated change-point was 1.90 (SE=0.68) years before AD diagnosis, when the trajectory transitioned from significant modest decline in performance over time (0.59 seconds per year, SE=0.14) to significant accelerated decline in performance (9.01 seconds per year, SE=3.12).

TMT-B: The estimated change-point was 4.82 (SE=0.73) years before AD diagnosis, when the trajectory transitioned from significant modest decline in performance over time (1.68 seconds per year, SE=0.42) to significant accelerated decline in performance (11.86 seconds per year, SE=2.21).

Verbal fluency

Category fluency: The estimated first change-point was 9.89 (SE=0.90) years before AD diagnosis, when the trajectory transitioned from significant minor decrease in performance over time (-0.13 words per year, SE=0.035) to significant moderate decline in performance (-0.34 words per year, SE=0.052). The second change-point was 3.17 (SE=0.41) years before AD diagnosis, when decline accelerated again (-0.97 words per year, SE=0.095).

Letter fluency: The estimated first change-point was 10.03 (SE=1.39) years before AD diagnosis, when the trajectory transitioned from non-significant minor decrease in performance over time (-0.010 words per year, SE=0.047) to significant moderate decline in performance (-0.22 words per year, SE=0.058). The second change-point was 2.25 (SE=0.91) years before AD diagnosis, when decline accelerated again (-0.69 words per year, SE=0.20).

Object recognition and naming

BNT: The estimated first change-point was 6.04 (SE=0.74) years before AD diagnosis, when the trajectory transitioned from significant minor decrease in performance over time (-0.10 words per year, SE=0.036) to significant moderate decline in performance (-0.89 words per year, SE = 0.17). The second change-point was 1.51 (SE=0.57) years before AD diagnosis, when decline accelerated again (-2.06 words per year, SE=0.57).

Abstract reasoning

Similarities: The estimated first change-point was 10.65 (SE=1.24) years before AD diagnosis, when the trajectory transitioned from non-significant increase in performance over time (0.038 points per year, SE=0.054) to significant moderate decline in performance (-0.16 points per year, SE=0.043). The second change-point was 1.72 (SE=0.35) years before AD diagnosis, when decline accelerated again (-0.84 points per year, SE=0.23).

Visuospatial ability

Card Rotations: The estimated first change-point was 15.48 (SE=1.72) years before AD diagnosis, when the trajectory transitioned from non-significant increase in performance over time (1.66 points per year, SE=1.59) to significant moderate decline in performance (-1.38 points per year, SE=0.35). The second change-point was 4.33 (SE=1.18) years before AD diagnosis, when decline accelerated again (-4.74 points per year, SE=1.08).

Global Cognitive Performance

MMSE: The estimated first change-point was 9.13 (SE=0.92) years before AD diagnosis, when the trajectory transitioned from non-significant increase in performance over time (-0.038 units per year, SE=0.024) to significant moderate decline in performance (-0.18 units per year,

SE=0.033). The second change-point was 1.77 (SE=0.47) years before AD diagnosis, when decline accelerated again (-1.03 units per year, SE=0.15).

Comparing change-points across cognitive measures

Figure 2 provides a schematic overview of estimated change-points for each cognitive measure.

Table 5 shows the p-value results from the Wald test of results from using bootstrapping to compare the first change-points for each measure against all other measures to identify the earliest changing measures. The measure with the earliest change-point was Card Rotations, which was significantly earlier than all other measures. The next measure to show an early change-point was CVLT-IMM, which was significantly earlier than CVLT-LD, BVRT, TMT-A, TMT-B, BNT, and MMSE but not significantly earlier than measures of verbal fluency, or Similarities. The measure to show the latest change-point in relation to AD diagnosis was TMT-A. Table 6 shows the p-value results from conducting the Wald test on bootstrapping samples to compare the second change-points for each measure. The second change-point for Card Rotations was not significantly earlier than the second change-points for CVLT measures.

Discussion

In a sample of participants with consensus diagnoses of clinical AD, we used extensive longitudinal cognitive data to examine the temporal sequence of stages of decline in 11 cognitive measures. Change-points identifying steeper rates of cognitive decline ranged from 15.5 years before AD diagnosis for the Card Rotations test to 1.9 years before AD diagnosis for TMT-A. While episodic memory assessed by CVLT measures was not the domain to show the earliest changes in rates of decline, changes were still detected up to 11.7 years before AD diagnosis.

Using change-point analyses in this way can reveal the temporal ordering of domain-specific accelerated decline in preclinical AD.

The change-point for Card Rotations (15.5 years before AD diagnosis) was significantly earlier than change-points for all other cognitive measures, including CVLT measures of episodic memory. This extends previous findings from a systematic review [16] in which it was casually observed that visuospatial ability shows the earliest acceleration of cognitive decline prior to AD. The underlying mechanisms that may lead to early accelerated decline in visuospatial ability in preclinical AD may be understood from the roles of the precuneus and other parietal regions in visuospatial tasks that involve spatial manipulation (as is the case with the Card Rotations Task in BLSA) [34]. The precuneus is also part of a large network that includes medial temporal lobe and frontal lobe regions that support spatial navigation [35]. The precuneus is one of the earliest brain regions to show accumulation of A β in preclinical AD [36, 37] and deficits in spatial navigation are one of the earliest impairments leading to loss of independence. Taken together, the functional importance of the precuneus in visuospatial processing and its susceptibility to early AD pathology support our finding that visuospatial ability would be affected early in preclinical AD.

CVLT-IMM showed the second earliest change-point at 11.7 years before diagnosis. This change was significantly earlier than CVLT-LD, which had a change-point at 7.6 years before AD diagnosis. The difference between change-points for CVLT-IMM and CVLT-LD is consistent with previous studies that reported faster rates of verbal learning compared to delayed free recall declines at earlier stages of disease progression [38, 39], and confirms the importance of early learning deficits in detecting individuals at risk of developing AD [40]. However, change-points for CVLT-IMM were not significantly earlier than those for measures of verbal fluency, the

Similarities Test, or the MMSE, suggesting that some aspects of executive function, i.e., verbal concept formation and abstract reasoning, as well as aspects of mental status may exhibit changes in the rates of decline as early as some memory-based learning tasks. These results contrast with earlier reports using BLSA data indicating that memory is affected earlier than executive function [15]. One possible explanation for the different pattern of results in the present analysis is the larger number of participants with longer follow-up compared to previous reports.

Every cognitive measure examined showed at least one change-point, with the majority of measures exhibiting two change-points. In measures with two change-points, the first change-point was always more than five years before diagnosis while the second change-point was less than five years before diagnosis and was followed by even faster rates of decline than the first change-point. There were three initial slopes (for CVLT-IMM, Similarities, and Card rotations) which showed non-significant improvement in performance before the first change-point preceded by significant decline. It is possible that such improvement in performance is due to practice effects (i.e. multiple exposures to cognitive tests improves performance). We performed sensitivity analyses including effects of practice (first versus subsequent administration)[41], which did not appreciably change the change-point results (data not shown). The second change-points appear to represent the transition from preclinical AD to the prodromal stage as indicated by their temporal proximity to average time of consensus-based diagnoses of MCI onset in the sample (2.5 years before AD diagnosis, 1.8 S.D.). Accelerated cognitive declines in the years immediately prior to symptom onset and AD diagnosis are consistent with other reports that MCI participants show steeper rates of decline compared to healthy controls for a range of cognitive domains including memory, executive function, attention and verbal fluency [42] .

The temporal ordering of the first and second change-points across cognitive measures were similar. However, the second change-point for Card Rotations was not significantly earlier than the second change-points for measures of memory or fluency, suggesting that there is little difference in the sensitivity of these measures as the time to diagnosis becomes shorter. As noted by Grober et al., [15] the temporal unfolding of cognitive decline identified by change-point studies implies that the predictive utility of different measures would be expected to vary by time from AD diagnosis. As such, the temporal ordering of change-points in the present study would suggest that measures of visuospatial ability and memory may serve as predictive tools for the development of AD as much as 15 years before diagnosis with other measures becoming more relevant closer to diagnosis. However, some measures of processing speed, i.e., TMT-A, may only have predictive utility less than five years before diagnosis, during a period when MCI may already be detectable. This interpretation is supported by previous reports of the predictive power of different cognitive measures [43].

A limitation of this study is that BLSA participants are a highly educated group who were mostly white. Future research using a similar statistical framework should be carried out on different populations to confirm the temporal ordering of accelerated cognitive decline is stable across more diverse groups. Furthermore, the mean baseline age was relatively old at 76 years, which may limit generalizability. The relatively older baseline age is due to the restriction of this analysis to the first timepoint when all tests were successfully completed. While some measures might have been available at earlier visits, comparisons across tests in the change-point analysis require all measures to come from the same visit. The older baseline age is also important to consider as some participants may have already been in the preclinical stage of AD before cognitive testing occurred. Therefore, it is possible that longer follow-ups of younger participants

might reveal even earlier change-points. Comparisons of cognitive trajectories between AD patients and controls were not carried out in the present study because the statistical framework relied on aligning participant's longitudinal trajectories based on the year in which they were diagnosed with AD. As such, it was not possible to include a sample of healthy controls for comparison as they cannot be aligned in the same way. Further, some older individuals who appear cognitively normal may be in a preclinical phase of AD and accelerated cognitive change in this group could drive change-points in the overall group of normal participants.

Additionally, the focus of this study was on defining and comparing population-level change-points and as such we did not examine individual differences in the timing of change-points, which requires longer follow-up at the individual level and adds computational challenges. Future work is needed to assess individual differences in change-points and we acknowledge that other statistical methods may be more suited for this, for example Bayesian Markov Chain Monte Carlo approaches. However, the strengths of this study are the comprehensive cognitive battery, frequent visit schedule (subjects were tested annually or biannually in this sample) and consensus-based determination of symptom onset and AD diagnosis. While we do not have pathological data for all subjects, a subset of participants with postmortem data has confirmed that clinical diagnoses correspond with AD pathology base diagnoses with an 85% accuracy in the BLSA (unpublished data). In addition, the use of the bootstrapping analysis not only allows us to capture parameter estimates and standard errors more accurately, but enables the statistical comparison of change-points among different cognitive measures, greatly extending the work of previous studies [16].

In summary, we found that the cognitive measure to show the earliest change in rates of decline in preclinical AD was visuospatial ability rather than episodic memory. Using change-point

analyses with bootstrapping can reveal the temporal patterns of accelerated cognitive decline in preclinical AD and may help guide the development of tools for participant screening in clinical trials.

Acknowledgements

We thank the staff of the BLSA and LBN cognitive testing group for their assistance and the BLSA participants for their dedication to this study.

Funding

This research was supported entirely by the Intramural Research Program of the National Institutes of Health, National Institute on Aging.

Declaration of interest

Declaration of interest: None.

References

- [1] Sperling RA, Aisen PS, Beckett LA, Bennett DA, Craft S, Fagan AM, Iwatsubo T, Jack Jr CR, Kaye J, Montine TJ (2011) 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* 7, 280-292.
- [2] Bäckman L, Jones S, Berger A-K, Laukka EJ, Small BJ (2005) Cognitive impairment in preclinical Alzheimer's disease: A meta-analysis. *Neuropsychology* 19, 520-531.
- [3] Armstrong NM, Huang C-W, Williams OA, Bilgel M, An Y, Doshi J, Erus G, Davatzikos C, Wong DF, Ferrucci L, Resnick SM (2019) Sex differences in the association between amyloid and longitudinal brain volume change in cognitively normal older adults. *Neuroimage-Clin* 22, 101769.
- [4] Jack Jr CR, Bennett DA, Blennow K, Carrillo MC, Dunn B, Haeberlein SB, Holtzman DM, Jagust W, Jessen F, Karlawish J (2018) NIA-AA Research Framework: Toward a biological definition of Alzheimer's disease. *Alzheimers Dement*, 14, 535-562.
- [5] Villemagne VL, Burnham S, Bourgeat P, Brown B, Ellis KA, Salvado O, Szoëke C, Macaulay SL, Martins R, Maruff P (2013) Amyloid β deposition, neurodegeneration, and cognitive decline in sporadic Alzheimer's disease: a prospective cohort study. *Lancet Neurol* 12, 357-367.
- [6] van Dyck CH (2018) Anti-amyloid- β monoclonal antibodies for Alzheimer's disease: pitfalls and promise. *Biol Psychiatry* 83, 311-319.
- [7] Sarazin M, Dorothée G, de Souza LC, Aucoeur P (2013) Immunotherapy in Alzheimer's disease: do we have all the pieces of the puzzle? *Biol Psychiatry* 74, 329-332.
- [8] Petersen RC, Smith GE, Waring SC (1999) Mild cognitive impairment: clinical characterization and outcome. *Arch Neurol* 56, 303-308.
- [9] Jack Jr CR, Albert MS, Knopman DS, McKhann GM, Sperling RA, Carrillo MC, Thies B, Phelps CH (2011) Introduction to the recommendations from the National Institute on Aging-

Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimers Dement* 7, 257-262.

[10] Sevigny J, Chiao P, Bussière T, Weinreb PH, Williams L, Maier M, Dunstan R, Salloway S, Chen T, Ling Y (2016) The antibody aducanumab reduces A β plaques in Alzheimer's disease. *Nature* 537, 50.

[11] Sperling RA, Rentz DM, Johnson KA, Karlawish J, Donohue M, Salmon DP, Aisen P (2014) The A4 study: stopping AD before symptoms begin? *Science translational medicine* 6, 228fs213-228fs213.

[12] Hall CB, Ying J, Kuo L, Lipton RB (2003) Bayesian and profile likelihood change point methods for modeling cognitive function over time. *Comput Stat Data An* 42, 91-109.

[13] Hall CB, Ying J, Kuo L, Sliwinski M, Buschke H, Katz M, Lipton RB (2001) Estimation of bivariate measurements having different change points, with application to cognitive ageing. *Stat Med* 20, 3695-3714.

[14] Grober E, An Y, Lipton RB, Kawas C, Resnick S (2019) Timing of onset and rate of decline in learning and retention in the pre-dementia phase of Alzheimer's disease. *J Int Neuropsych Soc*, 25, 699-705.

[15] Grober E, Hall CB, Lipton RB, Zonderman AB, Resnick SM, Kawas C (2008) Memory impairment, executive dysfunction, and intellectual decline in preclinical Alzheimer's disease. *J Int Neuropsych Soc* 14, 266-278.

[16] Karr JE, Graham RB, Hofer SM, Muniz-Terrera G (2018) When does cognitive decline begin? A systematic review of change point studies on accelerated decline in cognitive and neurological outcomes preceding mild cognitive impairment, dementia, and death. *Psych Aging* 33, 195.

[17] Thorvaldsson V, MacDonald SW, Fratiglioni L, Winblad B, Kivipelto M, Laukka EJ, Skoog I, Sacuiu S, Guo X, Östling S (2011) Onset and rate of cognitive change before dementia diagnosis: findings from two Swedish population-based longitudinal studies. *J Int Neuropsych Soc* 17, 154-162.

- [18] Laukka EJ, MacDonald SWS, Fratiglioni L, Bäckman L (2012) Preclinical Cognitive Trajectories Differ for Alzheimer's Disease and Vascular Dementia. *J Int Neuropsych Soc* 18, 191-199.
- [19] Shock NW (1984) Normal human aging: The Baltimore longitudinal study of aging. US Department of Health and Human Services, Public Health Service, National Institutes of Health, National Institute on Aging, Gerontology Research Center.
- [20] Morris JC (1993) The Clinical Dementia Rating (CDR): current version and scoring rules. *Neurology*, 41, 1588-1592.
- [21] Fuld PA (1978) Psychological testing in the differential diagnosis of the dementias. *Alzheimer's disease: senile dementia and related disorders* 7, 185-193.
- [22] Petersen RC (2004) Mild cognitive impairment as a diagnostic entity. *Journal of internal medicine* 256, 183-194.
- [23] American Psychiatric Association (1987) Diagnostic and Statistical Manual of Mental Health Disorders (DSM-III-R), American Psychiatric Association.
- [24] McKhann G, Drachman D, Folstein M, Katzman R, Price D, Stadlan EM (1984) Clinical diagnosis of Alzheimer's disease Report of the NINCDS-ADRDA Work Group* under the auspices of Department of Health and Human Services Task Force on Alzheimer's Disease. *Neurology* 34, 939-939.
- [25] Delis DC, Kramer JH, Kaplan E, Thompkins BAO (1987) CVLT: California verbal learning test-adult version: manual, Psychological Corporation.
- [26] Benton AL (1974) Visual retention test, Psychological Corporation, New York, NY.
- [27] Reitan RM (1992) Trail Making Test: Manual for administration and scoring, Reitan Neuropsychology Laboratory.

- [28] Benton AL (1968) Differential behavioral effects in frontal lobe disease. *Neuropsychologia* 6, 53-60.
- [29] Newcombe F (1969) Missile wounds of the brain: A study of psychological deficits.
- [30] Kaplan E, Goodglass H, Weintraub S (1983) *The Boston Naming Test* (2nd ed.), Lea & Febiger, Philadelphia, PA.
- [31] Wechsler D (1981) *WAIS-R manual: Wechsler adult intelligence scale-revised*, Psychological Corporation.
- [32] Wilson JR, De Fries J, Mc Clearn G, Vandenberg S, Johnson R, Rashad M (1975) Cognitive abilities: Use of family data as a control to assess sex and age differences in two ethnic groups. *The International Journal of Aging and Human Development* 6, 261-276.
- [33] Folstein MF, Folstein SE, McHugh PR (1975) "Mini-mental state": a practical method for grading the cognitive state of patients for the clinician. *J. Psychiatr. Res.* 12, 189-198.
- [34] Cavanna AE, Trimble MR (2006) The precuneus: a review of its functional anatomy and behavioural correlates. *Brain*, 129, 564-583.
- [35] Coughlan G, Laczó J, Hort J, Minihane A-M, Hornberger M (2018) Spatial navigation deficits—overlooked cognitive marker for preclinical Alzheimer disease? *Nat. Rev. Neurol*, 14, 496- 506
- [36] Bilgel M, Prince JL, Wong DF, Resnick SM, Jernigan BM (2016) A multivariate nonlinear mixed effects model for longitudinal image analysis: Application to amyloid imaging. *Neuroimage*, 134, 658-670.
- [37] Rodrigue K, Kennedy K, Devous M, Rieck J, Hebrank A, Diaz-Arrastia R, Mathews D, Park D (2012) β -Amyloid burden in healthy aging: regional distribution and cognitive consequences. *Neurology*, 78, 387-395.

- [38] Bilgel M, An Y, Lang A, Prince J, Ferrucci L, Jedynak B, Resnick (2014) Trajectories of Alzheimer disease-related cognitive measures in a longitudinal sample. *Alzheimers Dement*, 10, 735-742.
- [39] Grober E, Kawas C (1997) Learning and retention in preclinical and early Alzheimer's disease. *Psychol. Aging*, 12, 183.
- [40] Chang Y-L, Bondi MW, Fennema-Notestine C, McEvoy LK, Hagler Jr DJ, Jacobson MW, Dale AM, ADNI, (2010) Brain substrates of learning and retention in mild cognitive impairment diagnosis and progression to Alzheimer's disease. *Neuropsychologia*, 48, 1237-1247.
- [41] Vivot A, Power MC, Glymour MM, Mayeda ER, Benitez A, Spiro III A, Manly JJ, Proust-Lima C, Dufouil C, Gross AL (2016) Jump, hop, or skip: modeling practice effects in studies of determinants of cognitive change in older adults. *Am J Epidemiol*, 183, 302-314.
- [42] Johnson JK, Gross AL, Pa J, McLaren DG, Park LQ, Manly JJ, (2012) Longitudinal change in neuropsychological performance using latent growth models: a study of mild cognitive impairment. *Brain Imaging Behav*, 6, 540-550.
- [43] Saxton J, Lopez OL, Ratcliff G, Dulberg C, Fried LP, Carlson MC, Newman A, Kuller L (2004) Preclinical Alzheimer disease: neuropsychological test performance 1.5 to 8 years prior to onset. *Neurology*, 63, 2341-2347.

Tables

Table 1. Sample Characteristics of Participants from the Baltimore Longitudinal Study of Aging with a Diagnosis of AD

| | |
|---|-----------------------|
| N Subjects | 165 |
| Total N Assessments | 988 |
| Number of Visits per Subject, Mean (S.D.), Range | 6.0 (4.3), 1-22 |
| Female, n (%) | 83 (50.3) |
| White, n (%) | 141 (85.5) |
| Baseline Age, Mean (S.D.), Range | 76.5 (7.4), 61.1-91.2 |
| Education (years), Mean (S.D.), Range | 16.7 (2.6), 8 -21 |
| Age at MCI Onset, Mean (S.D.), Range | 84.0 (6.1), 65-98 |
| Age at AD Diagnosis, Mean (S.D.), Range | 86.5 (6.1), 65 - 102 |
| Time to AD Diagnosis, Mean (S.D.), Range | 10.0 (6.5), 0-22.9 |
| Follow up Interval, Mean (S.D.), Range | 8.3 (6.0), 0-24.1 |
| Interval between MCI onset and AD diagnosis, Mean (S.D.), Range | 2.5 (1.8), 0-10 |

Table 2. The Reliability of each Cognitive Measure as assessed using the Intra-class

Correlation (ICC)

| Cognitive Measure | ICC |
|-------------------|------|
| CVLT-IMM | 0.71 |
| CVLT-LD | 0.73 |
| BVRT | 0.73 |
| TMT-A | 0.55 |
| TMT-B | 0.67 |
| Categories | 0.73 |
| Letters | 0.74 |
| BNT | 0.84 |
| Similarities | 0.71 |
| Card Rotations | 0.79 |
| MMSE | 0.57 |

Abbreviations: BNT, Boston Naming Test; BVRT, Benton Visual Retention Test; CVLT, California Verbal Learning Test; CVLT-IMM, CVLT immediate free recall; CVLT-LD, CVLT long delayed free recall; MMSE, Mini Mental State Exam, TMT-A, Trail-Making Test-A, TMT-B, Trail-Making Test-B

Table 3: Fit statistics for models using no change-points, one change-point, or two change-points. Model fit was assessed using the likelihood ratio test, with bold text indicating the current model provides significantly better fit compared to the one prior.

| | | No change-points | 1 change-point | 2 change-points |
|----------------|---------------------|------------------|--------------------------------------|--------------------------------------|
| CVLT-IMM | -2 LL/DF P value | 6918.5/8 | 6745.4/10 <.0001 | 6720.9/12 <.0001 |
| CVLT-LD | -2 LL/DF P value | 4629.7/8 | 4436.5/10 <.0001 | 4428.9/12 0.022 |
| BVRT | -2 LL/DF P value | 5081.5/8 | 5013.0/10 <.0001 | 5011.5/12 0.47 |
| TMT-A | -2 LL/DF P value | 8405.4/8 | 8302.8/10 <.0001 | 8299.1/12 0.16 |
| TMT-B | -2 LL/DF P value | 8701.8/8 | 8626.8/10 <.0001 | 8626.5/12 0.86 |
| Categories | -2 LL/DF P value | 4245.0/8 | 4073.8/10 <.0001 | 4064.1/12 0.0078 |
| Letters | -2 LL/DF P value | 4582.4/8 | 4530.4/10 <.0001 | 4523.9/12 0.039 |
| BNT | -2 LL/DF P value | 5152.7/8 | 4963.6/10 <.0001 | 4949.9/12 0.0011 |
| Similarities | -2 LL/DF P value | 4221.8/8 | 4154.9/10 <.0001 | 4147.3/12 0.022 |
| Card Rotations | -2 LL/DF P value | 7548.6/8 | 7468.6/10 <.0001 | 7457.6/12 0.0041 |
| MMSE | -2 LL/DF P value | 3924.1/8 | 3726.4/10 <.0001 | 3719.0/12 0.025 |

Abbreviations: BNT, Boston Naming Test; BVRT, Benton Visual Retention Test; CVLT, California Verbal Learning Test; CVLT-IMM, CVLT immediate free recall; CVLT-LD, CVLT long delayed free recall; MMSE, Mini Mental State Exam, TMT-A, Trail-Making Test-A, TMT-B, Trail-Making Test-B; $-2LL/DF$, $-2 \times \log \text{likelihood} / \text{Degree of Freedom}$. Significant p-values indicate better model fit over the previous model, e.g. for CVLT-IMM, a one change-point model has a better fit than a no change-point model and a two change-point model has better fit than a one change-point model

Table 4: Estimated change-points in years from AD diagnosis and slopes in test units change per year. All estimates, standard errors, and p values are from bootstrapping.

| Estimate (S.E.) | Change-point 1 | Change-point 2 | Slope 1 | Slope 2 | Slope 3 |
|-----------------|----------------|----------------|-----------------------------|------------------------------|-----------------------------|
| CVLT-IMM | -11.65 (0.80) | -2.80 (0.78) | 0.23 (0.28) P = 0.41 | -1.44 (0.25) P < .0001 | -3.74 (0.48) P < .0001 |
| CVLT-LD | -7.58 (0.49) | -4.21 (0.33) | -0.080 (0.041) P = 0.051 | -0.48 (0.13) P = 0.0002 | -1.00 (0.091) P < .0001 |
| BVRT * | -4.83 (0.81) | NA | 0.36 (0.034) P < .0001 | 0.99 (0.12) P < .0001 | NA |
| TMT-A * | -1.90 (0.68) | NA | 0.59 (0.14) P < .0001 | 9.01 (3.12) P = 0.0039 | NA |
| TMT-B* | -4.82 (0.73) | NA | 1.68 (0.42) P < .0001 | 11.86 (2.21) P < .0001 | NA |
| Categories | -9.89 (0.90) | -3.17 (0.41) | -0.13 (0.035) P = 0.0002 | -0.34 (0.052) P < .0001 | -0.97 (0.095) P < .0001 |
| Letters | -10.03 (1.39) | -2.25 (0.91) | -0.010 (0.047) P = 0.83 | -0.22 (0.058) P = 0.00016 | -0.69 (0.20) P = 0.00048 |
| BNT | -6.04 (0.74) | -1.51 (0.57) | -0.10 (0.036) P = 0.0043 | -0.89 (0.17) P < .0001 | -2.06 (0.57) P = 0.00033 |
| Similarities | -10.65 (1.24) | -1.72 (0.53) | 0.038 (0.054) P = 0.48 | -0.16 (0.043) P = 0.00026 | -0.84 (0.23) P = 0.00019 |
| Card Rotation | -15.48 (1.72) | -4.33 (1.18) | 1.66 (1.59) P = 0.30 | -1.38 (0.35) P < .0001 | -4.74 (1.08) P < .0001 |
| MMSE | -9.13 (0.92) | -1.77 (0.44) | -0.038 (0.024) P = 0.11 | -0.18 (0.033) P < .0001 | -1.03 (0.15) P < .0001 |

* These measures only had a single change-point. Abbreviations: : BNT, Boston Naming Test; BVRT, Benton Visual Retention Test;

CVLT, California Verbal Learning Test; CVLT-IMM, CVLT immediate free recall; CVLT-LD, CVLT long delayed free recall;

MMSE, Mini Mental State Exam, TMT-A, Trail-Making Test-A; TMT-B, Trail-Making Test-B; S.E., standard error.

Table 5: The first change-point for each measure was compared between all measures using bootstrapping. P-values from the Wald Test are presented to show which measures were significantly different from each other.

| | CVLT-IMM | CVLT-LD | BVRT | TMT-A | TMT-B | Categories | Letters | BNT | Similarities | Card Rotation | MMSE |
|---------------|----------|---------|--------|--------|---------|------------|---------|--------|--------------|---------------|---------|
| CVLT-IMM | 1 | <.0001 | <.0001 | <.0001 | <.0001 | 0.13 | 0.31 | <.0001 | 0.46 | 0.045 | 0.031 |
| CVLT-LD | | 1 | 0.0029 | <.0001 | 0.0022 | 0.031 | 0.10 | 0.087 | 0.025 | <.0001 | 0.13 |
| BVRT | | | 1 | 0.0039 | 0.99 | <.0001 | 0.0018 | 0.26 | 0.00012 | <.0001 | 0.00069 |
| TMT-A | | | | 1 | 0.00063 | <.0001 | <.0001 | <.0001 | <.0001 | <.0001 | <.0001 |
| TMT-B | | | | | 1 | <.0001 | 0.00057 | 0.27 | <.0001 | <.0001 | 0.00034 |
| Categories | | | | | | 1 | 0.93 | 0.0014 | 0.64 | 0.0041 | 0.56 |
| Letters | | | | | | | 1 | 0.019 | 0.73 | 0.010 | 0.56 |
| BNT | | | | | | | | 1 | 0.0017 | <.0001 | 0.013 |
| Similarities | | | | | | | | | 1 | 0.023 | 0.33 |
| Card Rotation | | | | | | | | | | 1 | 0.0015 |
| MMSE | | | | | | | | | | | 1 |

Abbreviations: BNT, Boston Naming Test; BVRT, Benton Visual Retention Test; CVLT, California Verbal Learning Test; CVLT-IMM, CVLT immediate free recall; CVLT-LD, CVLT long delayed free recall; MMSE, Mini Mental State Exam, TMT-A, Trail-Making Test-A; TMT-B, Trail-Making Test-B.

Table 6: The second change-point for each measure was compared between all measures using bootstrapping. P-values from the Wald Test are presented to show which measures were significantly different from each other.

| | CVLT-IMM | CVLT-LD | Categories | Letters | BNT | Similarities | Card Rotation | MMSE |
|---------------|----------|---------|------------|---------|--------|--------------|---------------|--------|
| CVLT-IMM | 1 | 0.072 | 0.63 | 0.59 | 0.17 | 0.22 | 0.25 | 0.21 |
| CVLT-LD | | 1 | 0.043 | 0.034 | <.0001 | 0.00020 | 0.92 | <.0001 |
| Categories | | | 1 | 0.33 | 0.013 | 0.019 | 0.35 | 0.017 |
| Letters | | | | 1 | 0.47 | 0.61 | 0.12 | 0.60 |
| BNT | | | | | 1 | 0.77 | 0.029 | 0.70 |
| Similarities | | | | | | 1 | 0.044 | 0.95 |
| Card Rotation | | | | | | | 1 | 0.030 |
| MMSE | | | | | | | | 1 |

Abbreviations: BNT, Boston Naming Test; CVLT, California Verbal Learning Test; CVLT-IMM, CVLT immediate free recall; CVLT-LD, CVLT long delayed free recall; MMSE, Mini Mental State Exam.

Figure Legends

Figure 1: Line graphs showing the modelled population-level longitudinal trajectories from change-point models (in red) superimposed over spaghetti plots of the raw data (gray) for each cognitive measure. Vertical black lines indicate the change-points and the vertical red lines indicate timing of AD diagnosis. The X-axis represents years before AD diagnosis. Measurement units: CVLT-IMM, total correct out of 80; CVLT-LD, total correct out of 16; BVRT, total number of errors; TMT-A, seconds to complete; TMT-B, seconds to complete; Categories, mean number correct words; Letters, mean number correct words; BNT, correct out of 60; Similarities, total out of 28; Card Rotations, total out of 224; MMSE total out of 30.

Figure 2: Dot plot comparing the estimated change-points relative to AD diagnosis across cognitive measures. Cognitive measures are presented in order of first change-points. Red dots represent first change-points and blue points represent second change-points. Extended lines show standard errors. MCI indicates the average time of mild cognitive impairment symptom onset before AD diagnosis in this sample.

Figures

Figure 1

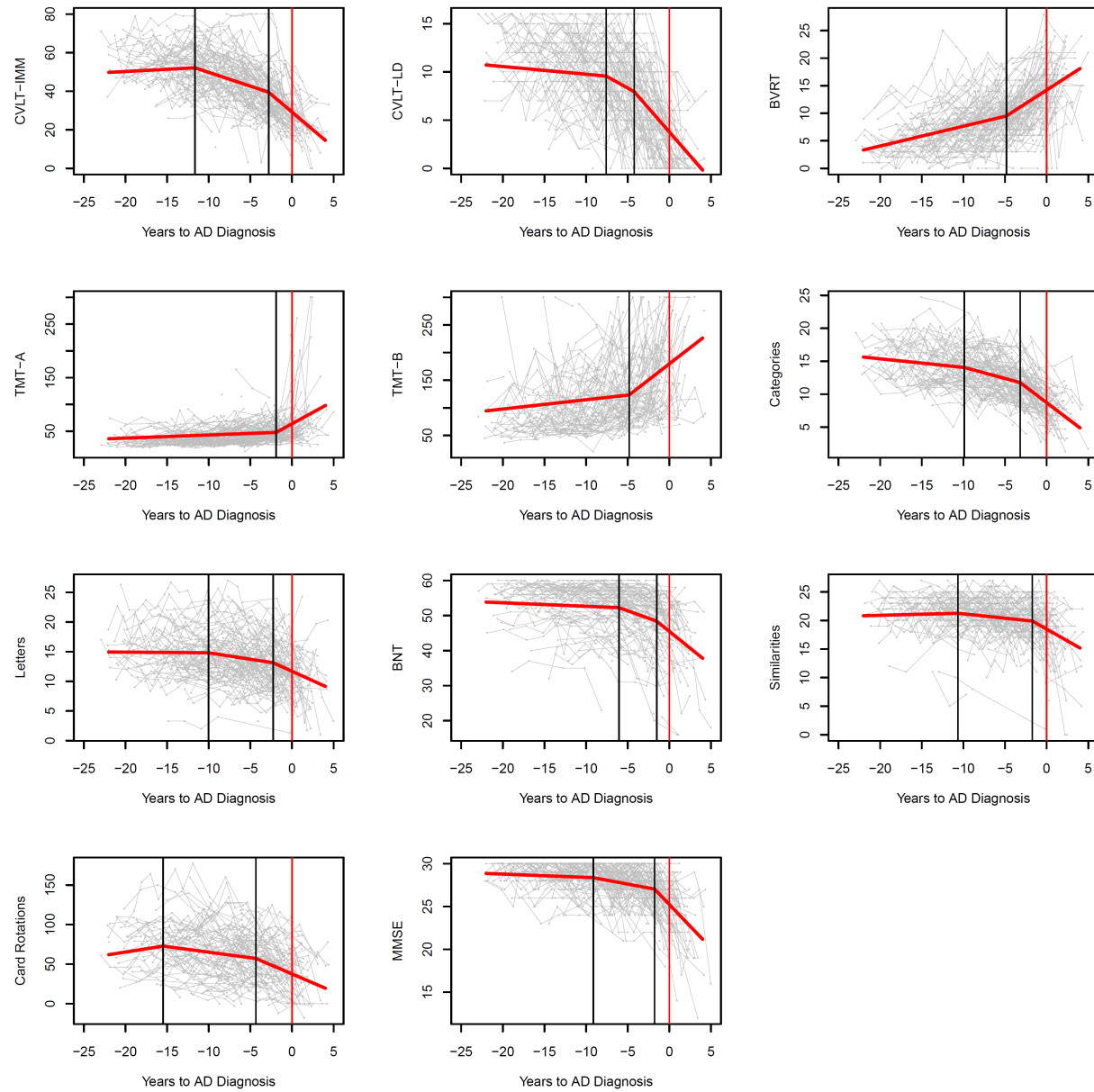


Figure 2

