AGING & MENTAL HEALTH
https://doi.org/10.1080/13607863.2021.1966746

Metrological properties of neuropsychological tests for measuring cognitive change in individuals with prodromal Alzheimer's disease

Xinnan Zhang[a], Yan Wu[a], Yao He[a], Xiaoyan Ge[a], Jing Cui[a], Hongjuan Han[b], Yanhong Luo[a], Long Liu[a], Zhixin Wang[a], Hongmei Yu[a,c] and the Alzheimer's Disease Neuroimaging Initiative

[a]Department of Health Statistics, School of Public Health, Shanxi Medical University, Taiyuan, China; [b]Department of Mathematics, School of Basic Medical Sciences, Shanxi Medical University, Taiyuan, China; [c]Shanxi Provincial Key Laboratory of Major Diseases Risk Assessment, Taiyuan, China

ABSTRACT

Objectives: In Alzheimer's Disease (AD) research, choosing appropriate method for measuring change in cognitive function over time can be challenging. The aim for this study was to examine the sensitivity of four neuropsychological tests used to measure cognition during the transition from mild cognitive impairment (MCI) to AD, and the impacts of associated covariates.

Methods: We enrolled 223 patients with MCI who progressed to AD and had completed multiple follow-up assessments in the Alzheimer's Disease Neuroimaging Initiative (ADNI) database. We constructed nonlinear mixed model for multivariate longitudinal data assuming that multiple neuropsychological tests would exhibit nonlinear transformation of a common factor in the latent cognitive process underlying the progression from MCI to AD.

Results: The Clinical Dementia Rating-Sum of the Boxes (CDR-SB) and Alzheimer's Disease Assessment Scale (11 items; ADAS-11) were more sensitive to cognitive changes in individuals with higher cognitive function, the Functional Activities Questionnaire (FAQ) was more sensitive to cognitive changes in individuals with middle cognitive function, and the Mini-Mental State Examination (MMSE) was more sensitive to cognitive changes in individuals with lower cognitive function. Gender ($p = 0.0139$) and educational level ($p = 0.0094$) had varying effects on different tests, such that men performed better on the FAQ and CDR-SB, and individuals with higher educational level tended to perform better on the FAQ and MMSE.

Conclusions: When choosing appropriate neuropsychological tests in cognitive measurements, the cognitive functional level of the patient as well as the impacts of covariates should be considered.

ARTICLE HISTORY
Received 21 April 2020
Accepted 1 August 2021

KEYWORDS
Aging;
cognition;
neuropsychological tests;
Alzheimer's disease;
mild cognitive impairment;
multivariate longitudinal data

Introduction

The rates of cognitive decline and dementia are increasing with the age of the overall population in many countries worldwide. Further, more than 60% of dementia cases are associated with Alzheimer's disease (AD) (Patterson, 2018). As the most common of the neurodegenerative diseases, AD is a progressive disorder characterized by cognitive decline, functional problems, and behavioral dysfunction (Smith & Buckwalter, 2005). Mild cognitive impairment (MCI) is a prodromal stage of dementia in which patients exhibit cognitive decline greater than that expected for their educational level and age, but retains all or almost all of the activities of daily living (Flicker et al., 1991; Petersen et al., 2001). Studies have shown that patients with MCI convert to AD at a rate of 10%–15% per year, and that up to 80% of MCI patients have converted to AD within 6 years of MCI diagnosis (Golob et al., 2007). It is of great significance to study the cognitive process from MCI to AD to prevent or delay the occurrence of dementia.

Several prospective studies have indicated that neuropsychological assessments can identify individuals with prodromal AD (i.e. the symptomatic predementia phase of AD or 'MCI due to AD') (Dubois & Albert, 2004). One study stressed the importance of measuring cognitive change using 'MCI due to AD' (i.e. prodromal AD) as a diagnostic criterion, illustrating that it provides evidence of longitudinal decline in cognition (Albert et al., 2011). Cognitive function can be evaluated using various different neuropsychological tests covering different cognitive domains (Dunson, 2003). Domains of cognitive weakness highlight the goals of cognitive training and clinical intervention, while domains of cognitive strength and preserved ability can be used to modify the clinical course of the disease (Lu & Lee, 2017; Pasternak & Smith, 2019). A brief cognitive screening tool can be used in the clinic to identify individuals with cognitive and functional changes, and track the progression of symptoms and the effects of treatment over time (Pasternak & Smith, 2019). However, there are no clear guidelines regarding how to choose specific neuropsychological tests for assessment. Further, conducting a battery of tests is time consuming, and lengthy tests can be difficult for individuals with cognitive impairment to endure (Dartigues et al., 1997). Thus, it is essential to choose appropriate neuropsychological tests, and particularly tests that are sufficiently sensitive to identify small changes in cognition (Morris et al., 1999).

Due to their metrological properties, some neuropsychological tests have ceiling or floor effects. Such tests generally have different levels of sensitivity to changes in the range of scores, which can be defined as differences in curvilinearity

CONTACT Hongmei Yu yu@sxmu.edu.cn

#Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. A complete listing of ADNI investigators can be found at: http://adni.loni.usc.edu/wpcontent/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf

(Proust-Lima et al., 2007). Moreover, in a series of neuropsychological tests, each test generally measures only one or several cognitive domains, even if the test is considered to be a global test, such as the Minimum-Mental State Examination (MMSE) (Folstein et al., 1975). Therefore, when analyzing the impacts of covariates on test scores over time, it is impossible to determine whether the results reflect the relationship between the covariate and global cognition, specific domains covered by the test, or different test abilities (Proust-Lima et al., 2008). Some researchers have hypothesized that a common factor among a group of tests might represent a latent cognitive process associated with the psychological phenomena that the tests are measuring (Fabrigoule et al., 1998). It may be possible to address this using the nonlinear mixed model for multivariate longitudinal data, which can be used to estimate metrological properties and to separate the effects of covariates on changes in common factors over time from specific effects on each test over time.

Our modeling approach is different from popular methods in many ways. First, the main interest of confirmatory factor analysis lies in the relationship between observed indicator variables and latent variables (Mueller & Hancock, 2015), while we focus on the common latent process change over time. Moreover, structural equation model generally assumes a Poisson or a Gaussian distribution for the outcomes when dealing with quantitative outcomes. In practice, the neuropsychological tests often have non-Gaussian distributions due to different behaviors with aging and different metrological properties (Amieva et al., 2005; Hall et al., 2001). Our model defined non-linear parameterized equations of observation to handle non-Gaussian continuous outcomes and to eliminate the spurious correlations when neglecting violation of the Gaussian assumption. Furthermore, in longitudinal settings, the item response theory tends to become very complicated and computationally intensive with a large number of model parameters. Our model estimated the nonlinear function linking the latent cognitive process with the outcomes inside a family of flexible continuous transformations. This remains computationally easy and improves the goodness of fit in comparison with the linear mixed model.

In this study, we selected longitudinal neuropsychological tests and covariates (sociodemographic information and genetic information) to construct a nonlinear mixed model for multivariate longitudinal data. With this model, we explored the sensitivity of a battery of neuropsychological tests generally used in patients with MCI to measure cognitive changes, and assessed the impacts of covariates on these tests.

Materials methods

Participants

Data used in the preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). The ADNI was launched in 2003 as a public-private partnership led by Principal Investigator Michael W. Weiner, MD (Weiner et al., 2013). The primary goal of the ADNI was to test whether serial magnetic resonance imaging (MRI), positron emission tomography (PET), other biological markers, and clinical and neuropsychological assessments could be combined to measure the progression of MCI and early AD. All participants provided written consent and the study protocol was approved by the institutional review committee of each participating center prior to implementing the specific protocol.

A total of 223 participants were enrolled in the study who were diagnosed MCI at baseline and eventually converted to AD during the follow-up period from 2006–2014. Neuropsychological data from the participants who had concerted to AD were excluded in the analysis. Follow-up assessments were performed semi-annually in the first 2 years and then yearly for the next 6 years. The inclusion and exclusion criteria of subjects were as follows. Inclusion criteria: 1) Hachinski Ischemic Score less than or equal to 4. 2) Age 65 years or above. 3) MMSE score greater than or equal to 24. 4) Geriatric Depression Scale score less than 6. 5) Good general health. 6) Medication stable for 4 weeks prior to screening. 7) No visual or auditory impairment. Exclusion criteria: 1) Meet the DSM-IV criteria for Dementia. 2) Contraindications for drug or MRI studies. 3) Mental disorders or substantial neurological diseases. The criteria for the diagnosis of MCI subjects were MMSE scores between 24 and 30 (inclusive), the Clinical Dementia Rating (CDR) of 0.5, a memory complaint, the preservation of activities of daily living, and no dementia. The diagnosis of AD was based on the National Institute of Neurologic and Communicative Disorders and Stroke-Alzheimer's Disease and Related Disorders Association (NINCDS/ADRDA) criteria.

Data collection

Neuropsychological evaluation: The Clinical Dementia Rating-Sum of the Boxes (CDR-SB) (Lynch et al., 2006) enables the evaluation of dementia severity using six "boxes" or dimensions, including Memory, Orientation, Judgment and Problem Solving, Community Affairs, Home and Hobbies, and Personal Care. The scores range from 0 to 18 (with higher scores indicating greater impairment).

The Alzheimer's Disease Assessment Scale (11 items; ADAS-11) (Mohs et al., 1997) is generally used to evaluate cognitive dysfunction in clinical settings, including clinical trials for drugs and other interventions. The ADAS-11 primarily measures language and memory, and includes the Word Recall Task, Naming Objects and Fingers, Following Commands, Constructional Praxis, Ideational Praxis, Orientation, the Word Recognition Task, Remembering Test Directions, Spoken Language, Comprehension, and Word-Finding Difficulty. Higher scores on the ADAS-11 (0–70 point score) indicate greater cognitive dysfunction.

The MMSE measures 10 items including orientation, memory, calculation, language, and visuospatial performance. The scores range from 0 to 30, with high scores illustrating better cognitive function.

The Functional Activities Questionnaire (FAQ) (Pfeffer et al., 1982) measures the degree of independence in terms of instrumental activities of daily living. It has 10 items, with the total score ranging from 0 to 30 points.

With the exception of the MMSE, higher scores represented poorer cognitive or behavioral function among all of the tests used of this study. Therefore, to make the evaluation criteria consistent with respect to the other tests, we subtracted the total MMSE scores from the actual measurement scores during modeling, and termed this value the transformed MMSE (t-MMSE). Because both the MMSE and FAQ scores range from 0 to 30 points, we multiplied the actual measurement scores of the other two scales by 30 and then divided these values by the total test scores to make the score range consistent among the tests.

Social demographic details and genetic information as covariates, including gender, age, family status (married, single), educational level (less than or equal to 12, 13–15, or more than or equal to 16 years), and ApoEε4 status (carrier, non carrier) were also collected.

Statistical analysis

To analyze the cognitive trajectory of participants from MCI to AD over the follow-up period, we used a nonlinear mixed model for multivariate longitudinal data. The statistical model assumed that the correlation between the four neuropsychological tests was induced by a latent common cognitive factor. The model consisted of two parts. The first was a linear mixed model that described the change over time in latent cognitive processes and evaluated the common effects of covariates on this latent trajectory (Laird & Ware, 1982). This linear mixed model included a random intercept and a slope for the time after inclusion in the study, as well as the following covariates: gender, age, family status, educational level, ApoEε4, and the interactions between these variables and time. To define individual deviations from a quadratic trend, we added a Brownian motion to relax the parametric form of the model (Mazur, 1959).

The second part of the model comprised test-specific measurement models that were flexible transformations linking the latent process and the neuropsychological tests. This part of the model also assessed the common effect of the covariates on each test and the specific effect of the covariates on it after adjustment for the latent cognitive process. These test-specific effects, called contrasts, were used to evaluate whether the covariates had a different and quantitative specific impact on each test aside from its global effect on latent cognitive processes. The nonlinear test-specific transformations were used to account for the global metrological properties of the tests (Proust-Lima et al., 2007). We selected Beta cumulative distribution functions as transformations because they are flexible and parsimonious. These functions provided a large variety of shapes (convex, concave, simply linear, or sigmoid) to account for the curvilinearity of the tests. The complete methodology details were described elsewhere (Proust et al., 2006; Proust-Lima et al., 2008).

We used SAS 9.4 for our preliminary analyses and the "Lcmm" package in R 3.6.1 for the nonlinear mixed model for multivariate longitudinal data. A value of $p < 0.05$ was considered significant.

Results

Participant characteristics

In this study, we enrolled 223 elderly people over the age of 65 who had converted to AD from MCI (Figure 1). Follow-up assessments were performed semi-annually in the first 2 years and then yearly for the next 6 years, with a median follow-up time of 18.33 months (interquartile range: 11.97-25.34 months). The participant characteristics are detailed in Table 1.

Comparison of the metrological properties of the four neuropsychological tests

Figure 2 displays the estimation of Beta transformations for the four cognitive tests. The shapes of the score distributions for all four tests showed nonlinear transformation. The transformations varied among the four cognitive tests. The shape of the CDR-SB was convex, showing that a decline of CDR-SB did not correspond to the same intensity of decline of the latent cognitive process in all range of the test. Indeed, a loss of 1 point in the latent cognitive process scale between −1 and 0 represented a loss of 1 CDR-SB point, whereas the same loss between 6 and 7 corresponded to a loss of 3 CDR-SB points. This indicated that changes in CDR-SB cannot be explained without accounting for the initial value, and that CDR-SB test was more sensitive to cognitive changes when participants had higher vs. lower levels of cognitive function. The situation of ADAS-11 was the same as above. The curve for t-MMSE was convex, which meant that the shape of the MMSE was concave. The MMSE dataset had a concave shape, indicating that the MMSE was more sensitive to cognitive changes in individuals with lower vs. higher levels of cognitive function. The FAQ dataset had a sigmoid shape, showing that it was sensitive to change in the middle range of cognitive function with respect to the latent cognitive process.

The minimum values of the FAQ and CDR-SB were higher than the minimum values assigned to the latent cognitive process (-2 and −3, respectively), which indicated that these tests were not suitable for estimating cognitive changes in individuals with low levels of cognitive function. Thus, this finding illustrated the floor effect of these two tests. The t-MMSE had a floor effect indicated that the MMSE had a ceiling effect. In contrast, the minimum and maximum values of the ADAS-11 were close to the values of the common factors −10 and 10, respectively, indicating that this test was not affected by ceiling or floor effects.

Association between covariates and cognitive trajectories

Table 2 (part (a)) presents the model, including a quadratic function of time and four covariates (gender, ApoEε4 status, educational level, and family status) for the latent process using Beta transformations. The latent process also included interactions between ApoEε4 and time variables. The latent cognitive process was positively associated with follow-up time ($p < 0.0001$), and was also associated with the interaction between ApoEε4 and time ($p = 0.0277$). The interactions between gender and time, between family status and time, and between educational level and time were not found to be significant. ApoEε4 gene status was not related to the mean level of the common factor, while ApoEε4 gene status was strongly associated with change in the common factor over time. Cognitive function decreased more rapidly over time in ApoEε4 carriers compared with ApoEε4 non-carriers, as shown in Figure 3.

Table 3 (part (b) contrasts) summarizes the test-specific effects of covariates on each neuropsychological test. The contrasts between the tests and gender suggested that gender did not have the same impact on each neuropsychological test: men tended to perform better on the FAQ and CDR-SB than women. The contrasts between the tests and educational level illustrated that educational level did not have the same impact on each neuropsychological test: individuals with higher educational level tended to perform better on the FAQ and t-MMSE than lower educational level. Contrasts between tests and ApoEε4 or family status were not statistically significant, indicating that the effects of ApoEε4 and family status did not differ among the tests.

Figure 4 summarizes the normal quantile plots of the standardized marginal residuals for the four neuropsychological

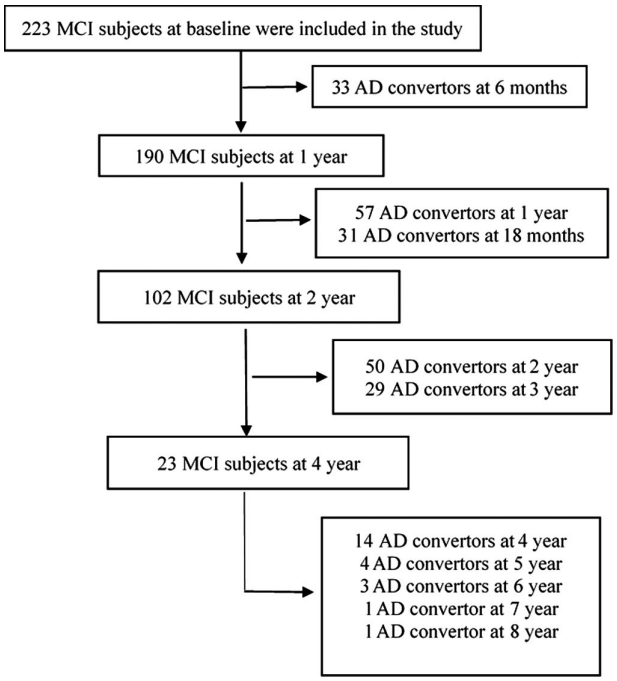

Figure 1. Diagram mapping the occurrence of Alzheimer's disease during the 8-year follow-up of the study.

Table 1. Baseline characteristics of demographic and cognitive performances.

Characteristics	Mean ± SD or n (%)
Age at baseline	75.51 ± 5.44
Gender	
Male	140 (62.8%)
Female	83 (37.2%)
Education (years)	
12	39 (17.5%)
13 ~ 15	102 (45.7%)
16	82 (36.8%)
Family status	
Married	182 (81.6%)
Single	41 (18.4%)
ApoEε4	
Carries	151 (67.7%)
Non carries	72 (32.3%)
MMSE score	26.83 ± 1.74
FAQ score	5.53 ± 4.92
CDR-SB score	1.90 ± 0.95
ADAS-11 score	13.17 ± 4.31

Note: SD: standard deviation; t-MMSE: transformed Mini-Mental State Examination; FAQ: Functional Activities Questionnaire; CDR-SB: Clinical Dementia Rating-Sum of the Boxes; ADAS-11: Alzheimer's Disease Assessment Scale (11 items).

tests. The normality assumption of the residuals seemed to be well satisfied for each of the four tests, illustrating that the model was a good fit. The latent process accounted for 16.42%–55.04% of the total variation in test scores at baseline and accounted for 71.31%–93.94% of the total variation in scores at 8 years after the initial assessment. As the follow-up duration increased, the proportion of change that could be interpreted to be associated with the latent process also gradually increased (Table 4). These data showed that the model was a good fit.

Discussion

We used a nonlinear mixed model for multivariate longitudinal data from the ADNI cohort. We estimated changes using a comprehensive indicator of cognition and studied the impacts of covariates on change in latent cognitive processes. At the same time, we found characteristics of the neuropsychological tests and provided theoretical support for the rational use of the tests.

A main result of this study is that the examined neuropsychological tests appear to be sensitive to the cognitive change associated with prodromal AD. The distribution of the dataset for each test emphasizes sensitivity to a different pattern of cognitive change over time. O'Bryant et al. (2008) reported that each point increase in CDR-SB scores were associated with a three-fold increase in the risk of a diagnosis of some form of dementia, supporting CDR-SB as an important tool for distinguishing between MCI and early dementia. Nicola Coley et al. (2011) announced that the CDR-SB can be used to effectively evaluate cognitive function with little floor or ceiling effects in people with mild to moderate AD. However, we found that the CDR-SB was affected by floor effect when assessing cognitive changes in individuals with low cognitive function. Cedarbaum JM et al. (2013) reported that there was a floor effect in the score of "Personal Care" box, indicating that it was often rated 0 in subjects with early cognitive changes. In addition, the CDR-SB scores were based on semi-structured interviews with patients and informants, which can lead to an additional level of subjectivity and hamper accurate assessment (Samtani et al., 2014).

Doraiswamy et al. (2001) concluded that the ADAS-11 was unlikely to be used for monitoring high functioning individuals or patients with MCI in short-term studies. This was due to the lack of appropriate sensitivity of certain items in early clinical AD studies. Specifically, a number of the ADAS-11 components showed ceiling effects in individuals with mild or moderate AD (Cano et al., 2010). However, our data suggested that ADAS-11 was suitable for assessing cognitive changes in individuals with high cognitive function in prodromal AD and it had no ceiling or floor effect. The heterogeneity of subjects and sample size may account for the different results observed in different cohorts.

In contrast, the MMSE should be chosen to evaluate cognitive changes in individuals with low cognitive function. The ceiling effect of MMSE indicated that it was not suitable for people with high cognitive level. Tombaugh and Mcintyre (1992) illustrated that the sensitivity of MMSE was higher in moderate to severe cognitive impairment and lower in mild impairment. The ceiling effect had the strongest relationship with concentration (sequence-seven), direction, and memory (Franco-Marina et al., 2010), indicating that these items did not correctly reflect the cognitive level of patients with mild impairment. This pointed out the research direction for improving the sensitivity of MMSE.

Many studies have used FAQ for cognitive screening or to assess the incidence of dementia (Laks et al., 2005; Tekin et al., 2001), because it can determine an individual's neuropsychiatric symptoms, such as delusions and aberrant motor function. Teng et al. (2010) reported that the FAQ as a standardized assessment of instrumental activities of daily living (ADL) was sufficiently sensitive and specific in the diagnosis of AD and MCI, and that tracking current events, bill paying, and transportation were the greatest diagnostic utility in the FAQ items. We found that the FAQ was very sensitive to cognitive change at middle cognitive level, while it had floor effect on cognitive change at low cognitive level. Considering that the FAQ mainly measures functional changes, the floor effect may originate from items that are not sensitive to cognitive changes. In this study, the majority of the neuropsychological tests had curvilinear performance. This curvilinearity signals the difficulty in

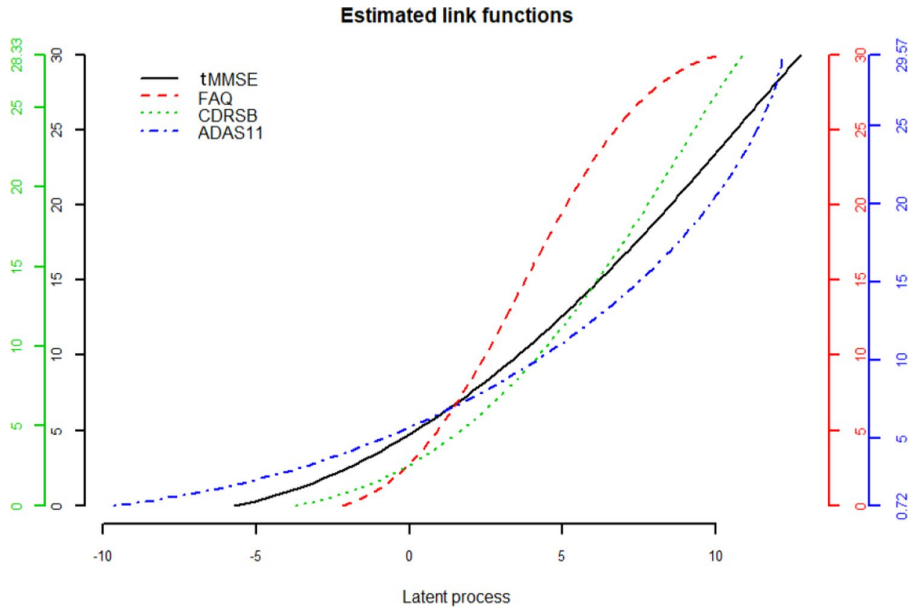

Figure 2. Estimated beta transformations for each test.Note: t-MMSE: transformed Mini-Mental State Examination; FAQ: Functional Activities Questionnaire; CDR-SB: Clinical Dementia Rating-Sum of the Boxes; ADAS-11: Alzheimer's Disease Assessment Scale (11 items).

Table 2. Common effect of the covariates on latent cognitive process (a).

Parameter	Estimate	SE	Wald	*P*
Gender	−0.0978	0.1977	−0.4940	0.6210
ApoEε4	0.3420	0.1825	1.8740	0.0609
Education	−0.0635	0.1214	−0.5230	0.6009
t^2	−0.0273	0.0153	−1.7880	0.0738
t	1.2183	0.1077	11.3120	<0.0001
Family status	−0.2256	0.2428	−0.9290	0.3529
t^2*ApoEε4	0.0314	0.0143	2.2010	0.0277

Note: Log likelihood =-11875.62; Number of parameters = 46; AIC = 23843.24; BIC = 23999.97.
Reference: female; ε4 gene noncarriers; 12 years and under; married.

Figure 3. Predicted mean evolution for the latent cognitive process in ε4 gene carriers and ε4 gene non-carriers.

choosing a cut-off value of cognitive decline for diagnosing prodromal AD (Mura et al., 2014). A decrease of 1 point on the test scale does not have the same meaning in the all range of the test, and it is decided by the initial level of test. It may be interesting to select several possible cut-off values to provide the predictions of the probability of developing AD according to the metrological properties of the neuropsychological tests.

Regarding the impacts of covariates on neuropsychological tests, our results showed that gender and educational level had different effects on different tests. Gender had a significant impact on both the FAQ and CDR-SB such that men performed better than women on both scales. Men had an advantage in a number of cognitive domains, including visuospatial and episodic memory, language, and semantic ability (Irvine et al., 2012). The reasons for this advantage include estrogen reduction in postmenopausal women, higher cognitive reserve in men, and sex differences in AD pathology (Laws et al., 2016).

Some studies have shown that a low level of education was a risk factor for cognitive decline (Sachdev et al., 2012). However, others have found that educational level was not related to cognitive decline, which was consistent with our results (Kryscio et al., 2006). We found that MMSE and FAQ scores significantly varied by educational level, indicating that these scales were sensitive to different levels of education. Tappen et al. (2010) used hierarchical regression and demonstrated that education had significant effects on the FAQ and MMSE, suggesting that education correction was necessary. A higher level of cognitive reserve may translate to improved mental efficiency and flexibility in terms of dealing with tasks (Lenehan et al., 2015), thus enabling individuals to better cope or compensate for the subtle brain changes imposed by aging (Foverskov et al., 2018), and thus delay the onset of impairment and AD symptoms (Soldan et al., 2017).

The statistical methodology we used has several advantages. First, it handles unbalanced and non-Gaussian repeated measurements of bounded quantitative outcomes, and takes into account the metrological properties of neuropsychological tests (Mura et al., 2014). It is worth noting that the latent cognitive process was defined by a series of neuropsychological tests. Therefore, the connection between neuropsychological test scores and latent cognitive processes enabled us to estimate the sensitivity of each test in detecting cognitive changes. Another advantage of this model is the way of explaining covariate effects. Using fixed contrasts, this model is able to

Table 3. Test-specific effects on each of the neuropsychological tests through contrasts (b).

Parameter	Estimate	SE	Wald	*P*
Contrasts on gender ($p = 0.0139$)				
t-MMSE	−0.2622	0.2000	−1.3110	0.1900
FAQ	0.4134	0.1575	2.6250	0.0087
CDR-SB	0.2790	0.1226	2.2760	0.0229
ADAS-11	−0.4301	0.2270	−1.8950	0.0581
Contrasts on ApoEε4 ($p = 0.9035$)				
t-MMSE	−0.1210	0.1825	−0.6630	0.5072
FAQ	0.0396	0.1431	0.2770	0.7821
CDR-SB	0.0560	0.1112	0.5040	0.6145
ADAS-11	0.0254	0.2078	0.1220	0.9026
Contrasts on education ($p = 0.0094$)				
t-MMSE	−0.3532	0.1256	−2.8130	0.0049
FAQ	0.2549	0.0972	2.6240	0.0087
CDR-SB	0.1150	0.0750	1.5340	0.1250
ADAS-11	−0.0167	0.1385	−0.1210	0.9040
Contrasts on family status ($p = 0.5627$)				
t-MMSE	−0.1837	0.2456	−0.7480	0.4544
FAQ	0.0207	0.1931	0.1070	0.9147
CDR-SB	0.2026	0.1499	1.3520	0.1765
ADAS-11	−0.0396	0.2782	−0.1420	0.8868

Note: t-MMSE: transformed Mini-Mental State Examination; FAQ: Functional Activities Questionnaire; CDR-SB: Clinical Dementia Rating-Sum of the Boxes; ADAS-11: Alzheimer's Disease Assessment Scale (11 items).
Reference: female; ε4 gene noncarriers; 12 years and under; married.

Table 4. Interpreting variance percentage in the common latent process model for the four tests at given time.

Tests	Baseline	2 years follow-up	4 years follow-up	6 years follow-up	8 years follow-up
t-MMSE	19.40	41.22	56.94	67.78	75.28
FAQ	35.51	61.60	75.16	82.80	87.45
CDRSB	55.04	78.10	87.06	91.45	93.94
ADAS11	16.42	36.40	51.91	63.20	71.32

Note: t-MMSE: transformed Mini-Mental State Examination; FAQ: Functional Activities Questionnaire; CDR-SB: Clinical Dementia Rating-Sum of the Boxes; ADAS-11: Alzheimer's Disease Assessment Scale (11 items).

distinguish connection between covariates and the latent process and the connection between covariates and various neuropsychological test. This gave us the opportunity to show that covariates had different and quantitative specific impact on each neuropsychological test. Finally, it can avoid the misspecification of the linear mixed model in evaluating predictors of cognitive decline (Proust-Lima et al., 2011). As a consequence, the nonlinear mixed model for multivariate longitudinal data was suitable with respect to the aim of our study.

It is well known that neuropsychological assessments play a prominent role in the detection, evaluation, diagnosis and treatment of neurodegenerative diseases (Boller & Barba, 2001). We chose several neuropsychological tests which were commonly used to measure cognition (Baldwin & Farias, 2009), and understood the characteristics of these neuropsychological

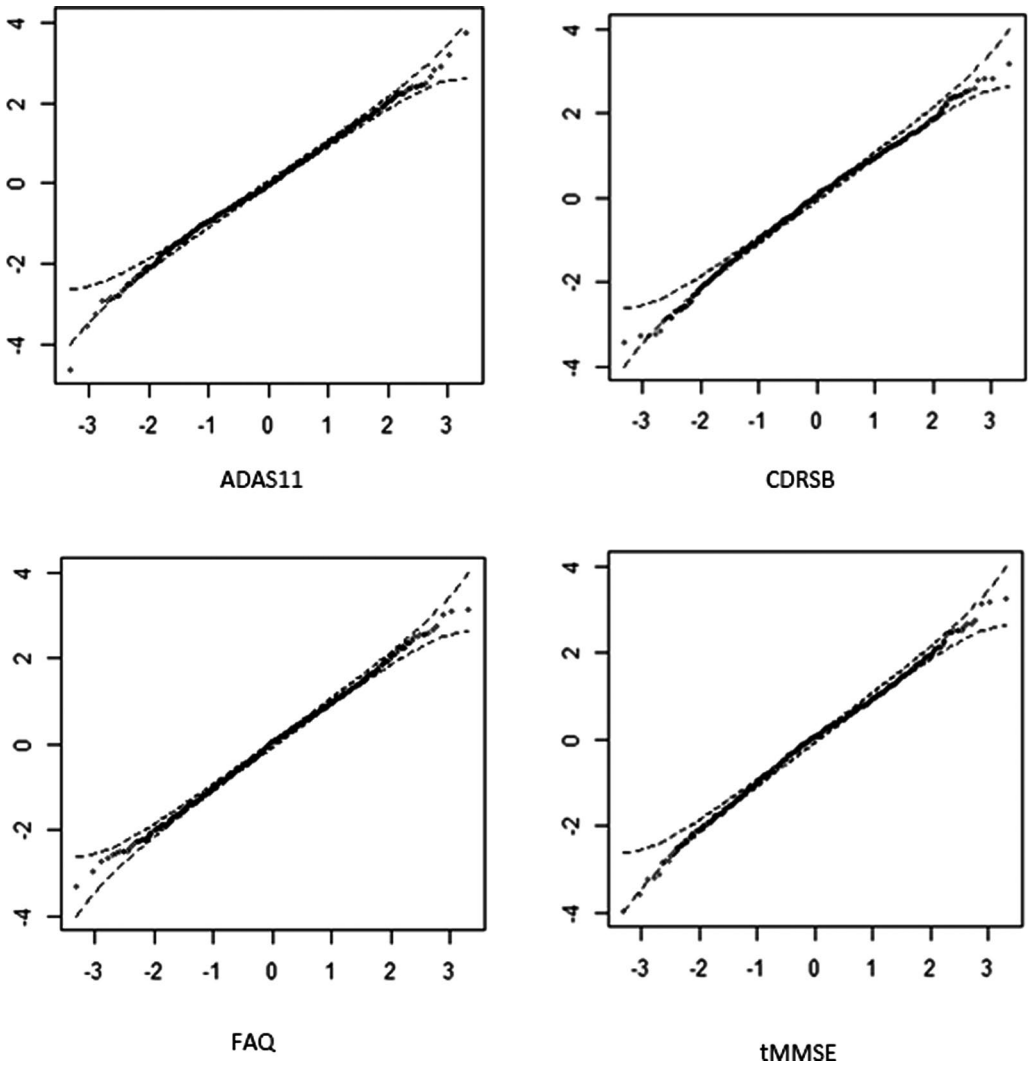

Figure 4. Normal quantile plot of the standardized marginal residuals for each test.Note: t-MMSE: transformed Mini-Mental State Examination; FAQ: Functional Activities Questionnaire; CDR-SB: Clinical Dementia Rating-Sum of the Boxes; ADAS-11: Alzheimer's Disease Assessment Scale (11 items).

tests through a nonlinear mixed model for multivariate longitudinal data. We found that patients at different levels of cognitive function showed different sensitivities to cognitive changes for the neuropsychological tests, which highlighted that we should chose appropriate neuropsychological tests according to the level of functioning of each patient. This study has clinical relevance and utility in that the individual test is enough for clinical purpose. The clinical diagnosis from history taking and reports on cognitive and functional changes can be supported by test scores. Discrepancy between observation of deterioration and tests scores can alert the clinicians to other causes for deterioration, for example depression.

There were some limitations to this study. First, the ADNI cohort comprised convenience samples as opposed to being an epidemiological cohort. This likely led to the recruitment of a high proportion of participants who were experiencing cognitive decline and who were taking multiple medications to treat AD, MCI, or other conditions that may affect the results. Second, this study only chose neuropsychological tests as indexes of cognition for analysis and it was necessary to combine clinical information and other indicators in future research. Third, as this study was not designed to compare different measurements in specific cognitive domains, we cannot reasonably highlight one measure in terms of its utility for assessing particular symptoms in dementia patients.

Conclusions

Our study provides reasonable explanations for the selection of appropriate neuropsychological tests and measurements of cognitive change among individuals with prodromal AD at different cognition levels. The CDR-SB and the ADAS-11 are appropriate to whom with high cognitive level. The FAQ could be a good candidate to measure cognitive changes in individuals with middle cognitive level. The MMSE is adapted for individuals with low cognitive level. When conducting cognitive measurement, the bias caused by covariates (such as gender and education) on the results of tests should be considered. Clinicians should select appropriate neuropsychological tests according to specific cognitive stages and consider the impacts of covariates when measuring patient cognition.

Acknowledgements

Data collection and sharing for this project was funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U01 AG024904) and DOD ADNI (Department of Defense award number W81XWH-12-2-0012). The ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: AbbVie, Alzheimer's Association; Alzheimer's Drug Discovery Foundation; Araclon Biotech; BioClinica, Inc.; Biogen; Bristol-Myers Squibb Company; CereSpir, Inc.; Cogstate; Eisai Inc.; Elan Pharmaceuticals, Inc.; Eli Lilly and Company; EuroImmun; F. Hoffmann-La Roche Ltd and its affiliated company Genentech, Inc.; Fujirebio; GE Healthcare; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research & Development, LLC.; Johnson & Johnson Pharmaceutical Research & Development LLC.; Lumosity; Lundbeck; Merck & Co., Inc.; Meso Scale Diagnostics, LLC.; NeuroRx Research; Neurotrack Technologies; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal Imaging; Servier; Takeda Pharmaceutical Company; and Transition Therapeutics. The Canadian Institutes of Health Research is providing funds to support ADNI clinical sites in Canada. Private sector contributions are facilitated by the Foundation for the National Institutes of Health (www.fnih.org). The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer's Therapeutic Research Institute at the University of Southern California. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of Southern California. We also thank the reviewers for their helpful suggestions.

Disclosure statement

The authors report no conflict of interest.

Author's contribution

Hongmei Yu designed the study. Xinnan Zhang, Yan Wu, Yao He, Xiaoyan Ge, Jing Cui, Hongjuan Han, Long Liu, and Yanhong Luo provided study materials and guidance. Xinnan Zhang and Zhixin Wang were responsible for analyzing the data. Xinnan Zhang wrote a draft of the paper. All authors discussed and approved the final manuscript.

Funding

This work was supported by the National Natural Science Foundation of China (Grant nos. 81673277, 81973154, 81903418).

ORCID

Xinnan Zhang http://orcid.org/0000-0002-7050-5691
Yao He http://orcid.org/0000-0003-2907-505X
Hongmei Yu http://orcid.org/0000-0002-9852-2485

References

Albert, M. S., DeKosky, S. T., Dickson, D., Dubois, B., Feldman, H. H., Fox, N. C., Gamst, A., Holtzman, D. M., Jagust, W. J., Petersen, R. C., Snyder, P. J., Carrillo, M. C., Thies, B., & Phelps, C. H. (2011). The diagnosis of mild cognitive impairment due to Alzheimer's disease: Recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimer's & Dementia: The journal of the Alzheimer's Association*, *7*(3), 270–279. https://doi.org/10.1016/j.jalz.2011.03.008

Amieva, H., Jacqmin-Gadda, H., Orgogozo, J.-M., Le Carret, N., Helmer, C., Letenneur, L., Barberger-Gateau, P., Fabrigoule, C., & Dartigues, J.-F. (2005). The 9 year cognitive decline before dementia of the Alzheimer type: A prospective population-based study. *Brain: A Journal of Neurology*, *128*(5), 1093–1101. https://doi.org/10.1093/brain/awh451

Baldwin, S., & Farias, S. T. (2009). Neuropsychological assessment in the diagnosis of Alzheimer's disease. *Current protocols in neuroscience*, Chapter 10, Unit10.3.

Boller, F., & Barba, G. D. (2001). Neuropsychological tests in Alzheimer's disease. *Aging Clinical and Experimental Research*, *13*(3), 210–220. https://doi.org/10.1007/BF03351479

Cano, S. J., Posner, H. B., Moline, M. L., Hurt, S. W., Swartz, J., Hsu, T., & Hobart, J. C. (2010). The ADAS-cog in Alzheimer's disease clinical trials: Psychometric evaluation of the sum and its parts. *Journal of Neurology, Neurosurgery, and Psychiatry*, *81*(12), 1363–1368. https://doi.org/10.1136/jnnp.2009.204008

Cedarbaum, J. M., Jaros, M., Hernandez, C., Coley, N., Andrieu, S., Grundman, M., & Vellas, B. (2013). Rationale for use of the Clinical Dementia Rating Sum of Boxes as a primary outcome measure for Alzheimer's disease clinical trials. *Alzheimer's & Dementia: The Journal of the Alzheimer's Association*, *9*(1S), S45–S55. https://doi.org/10.1016/j.jalz.2011.11.002

Coley, N., Andrieu, S., Jaros, M., Weiner, M., Cedarbaum, J., & Vellas, B. (2011). Suitability of the Clinical Dementia Rating-Sum of Boxes as a single primary endpoint for Alzheimer's disease trials. *Alzheimer's & Dementia: The Journal of the Alzheimer's Association*, *7*(6), 602–610.e2. https://doi.org/10.1016/j.jalz.2011.01.005

Dartigues, J. F., Commenges, D., Letenneur, D., Barberger-Gateau, P., Gilleron, V., Fabrigoule, C., Mazaux, J. M., Orgogozo, J. M., & Salamon, R. (1997). Cognitive

predictors of dementia in elderly community residents. *Neuroepidemiology*, *16*(1), 29–39. https://doi.org/10.1159/000109668

Doraiswamy, P. M., Kaiser, L., Bieber, F., & Garman, R. L. (2001). The Alzheimer's Disease Assessment Scale: Evaluation of psychometric properties and patterns of cognitive decline in multicenter clinical trials of mild to moderate Alzheimer's disease. *Alzheimer Disease and Associated Disorders*, *15*(4), 174–183.

Dubois, B., & Albert, M. L. (2004). Amnestic MCI or prodromal Alzheimer's disease?*The Lancet. Neurology*, *3*(4), 246–248. https://doi.org/10.1016/S1474-4422(04)00710-0

Dunson, D. B. (2003). Dynamic latent trait models for multidimensional longitudinal data. *Journal of the American Statistical Association*, *98*(463), 555–563. https://doi.org/10.1198/016214503000000387

Fabrigoule, C., Rouch, I., Taberly, A., Letenneur, L., Commenges, D., Mazaux, J. M., Orgogozo, J. M., & Dartigues, J. F. (1998). Cognitive process in preclinical phase of dementia. *Brain : A Journal of Neurology*, *121 (Pt 1))*, 135–141. https://doi.org/10.1093/brain/121.1.135

Flicker, C., Ferris, S. H., & Reisberg, B. (1991). Mild cognitive impairment in the elderly: Predictors of dementia. *Neurology*, *41*(7), 1006–1006. https://doi.org/10.1212/wnl.41.7.1006

Folstein, M. F., Folstein, S. E., & McHugh, P. R. (1975). Mini-mental state. A practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research*, *12*(3), 189–198. https://doi.org/10.1016/0022-3956(75)90026-6

Foverskov, E., Glymour, M. M., Mortensen, E. L., Holm, A., Lange, T., & Lund, R. (2018). Education and cognitive aging: Accounting for selection and confounding in linkage of data from the Danish registry and survey of health, ageing and retirement in Europe. *American Journal of Epidemiology*, *187*(11), 2423–2430. https://doi.org/10.1093/aje/kwy162

Franco-Marina, F., García-González, J. J., Wagner-Echeagaray, F., Gallo, J., Ugalde, O., Sánchez-García, S., Espinel-Bermúdez, C., Juárez-Cedillo, T., Rodríguez, M. A. V., & García-Peña, C. (2010). The mini-mental state examination revisited: Ceiling and floor effects after score adjustment for educational level in an aging Mexican population. *International Psychogeriatrics*, *22*(1), 72–81. https://doi.org/10.1017/S1041610209990822

Golob, E. J., Irimajiri, R., & Starr, A. (2007). Auditory cortical activity in amnestic mild cognitive impairment: Relationship to subtype and conversion to dementia. *Brain : A Journal of Neurology*, *130*(Pt 3), 740–752. https://doi.org/10.1093/brain/awl375

Hall, C. B., Ying, J., Kuo, L., Sliwinski, M., Buschke, H., Katz, M., & Lipton, R. B. (2001). Estimation of bivariate measurements having different change points, with application to cognitive ageing. *Statistics in Medicine*, *20*(24), 3695–3714. https://doi.org/10.1002/sim.1113

Irvine, K., Laws, K. R., Gale, T. M., & Kondel, T. K. (2012). Greater cognitive deterioration in women than men with Alzheimer's disease: A meta analysis. *Journal of Clinical and Experimental Neuropsychology*, *34*(9), 989–998. https://doi.org/10.1080/13803395.2012.712676

Kryscio, R. J., Schmitt, F. A., Salazar, J. C., Mendiondo, M. S., & Markesbery, W. R. (2006). Risk factors for transitions from normal to mild cognitive impairment and dementia. *Neurology*, *66*(6), 828–832. https://doi.org/10.1212/01.wnl.0000203264.71880.45

Laird, N. M., & Ware, J. H. (1982). Random-Effects models for longitudinal data. *Biometrics*, *38*(4), 963–974. https://doi.org/10.2307/2529876

Laks, J., Batista, E. M. R., Guilherme, E. R. L., Contino, A. L. B., Faria, M. E. V., Rodrigues, C. S., Paula, E. d., & Engelhardt, E. (2005). Prevalence of cognitive and functional impairment in community-dwelling elderly: Importance of evaluating activities of daily living. *Arquivos de Neuro-Psiquiatria*, *63*(2a), 207–212. https://doi.org/10.1590/S0004-282X2005000200003

Laws, K. R., Irvine, K., & Gale, T. M. (2016). Sex differences in cognitive impairment in Alzheimer's disease. *World Journal of Psychiatry*, *6*(1), 54–65. https://doi.org/10.5498/wjp.v6.i1.54

Lenehan, M. E., Summers, M. J., Saunders, N. L., Summers, J. J., & Vickers, J. C. (2015). Relationship between education and age-related cognitive decline: A review of recent research. *Psychogeriatrics*, *15*(2), 154–162. https://doi.org/10.1111/psyg.12083

Lu, P. H., & Lee, G. J. (2017). The role of neuropsychology in the assessment of the cognitively impaired elderly. *Neurologic Clinics*, *35*(2), 191–206. https://doi.org/10.1016/j.ncl.2017.01.002

Lynch, C. A., Walsh, C., Blanco, A., Moran, M., Coen, R. F., Walsh, J. B., & Lawlor, B. A. (2006). The clinical dementia rating sum of box score in mild dementia. *Dementia and Geriatric Cognitive Disorders*, *21*(1), 40–43. https://doi.org/10.1159/000089218

Mazur, P. (1959). On the theory of Brownian motion. *Physica*, *25*(1-6), 149–162. https://doi.org/10.1016/S0031-8914(59)91931-7

Mohs, R. C., Knopman, D., Petersen, R. C., Ferris, S. H., Ernesto, C., Grundman, M., Sano, M., Bieliauskas, L., Geldmacher, D., Clark, C., & Thal, L. J. (1997). 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 Disease and Associated Disorders*, *11* (Suppl 2), S13–S21. https://doi.org/10.1097/00002093-199700112-00003

Morris, M. C., Evans, D. A., Hebert, L. E., & Bienias, J. L. (1999). Methodological Issues in the Study of Cognitive Decline. *American Journal of Epidemiology*, *149*(9), 789–793. https://doi.org/10.1093/oxfordjournals.aje.a009893

Mueller, R. O., & Hancock, G. R. (2015). Factor analysis and latent structure analysis: Confirmatory factor analysis. In J. D. Wright (Ed.), *International encyclopedia of the social & behavioral sciences* (2nd ed., pp. 686–690). Elsevier.

Mura, T., Proust-Lima, C., Jacqmin-Gadda, H., Akbaraly, T. N., Touchon, J., Dubois, B., & Berr, C. (2014). Measuring cognitive change in subjects with prodromal Alzheimer's disease. *Journal of Neurology. Journal of Neurology, Neurosurgery & Psychiatry*, *85*(4), 363–370. https://doi.org/10.1136/jnnp-2013-305078

O'Bryant, S. E., Waring, S. C., Cullum, C. M., Hall, J., Lacritz, L., Massman, P. J., Lupo, P. J., Reisch, J. S., Doody, R., & Texas Alzheimer's Research Consortium. (2008). Staging dementia using clinical dementia rating scale sum of boxes scores: A Texas Alzheimer's research consortium study. *Archives of Neurology*, *65*(8), 1091–1095. https://doi.org/10.1001/archneur.65.8.1091

Pasternak, E., & Smith, G. (2019). Chapter 6 - Cognitive and neuropsychological examination of the elderly. In S. T. Dekosky & S. Asthana (Eds.), *Handbook of clinical neurology* (Vol. 167, pp. 89–104). Elsevier.

Patterson, C. (2018). World Alzheimer report 2018. Alzheimer's Disease International. Retrieved September from https://www.alz.co.uk/research/world-report-2018

Petersen, R. C., Doody, R. S., Kurz, A., Mohs, R. C., Morris, J. C., Rabins, P. V., Ritchie, K., Rossor, M., Thal, L., & Winblad, B. (2001). Current concepts in mild cognitive impairment. *JAMA Neurology*, *58*(12), 1985–1992.

Pfeffer, R. I., Kurosaki, T. T., Harrah, C. H., Chance, J. M., & Filos, S. (1982). Measurement of functional activities in older adults in the community. *Journal of Gerontology*, *37*(3), 323–329. https://doi.org/10.1093/geronj/37.3.323

Proust, C., Jacqmin-Gadda, H., Taylor, J. M. G., Ganiayre, J., & Commenges, D. (2006). A nonlinear model with latent process for cognitive evolution using multivariate longitudinal data. *Biometrics*, *62*(4), 1014–1024. https://doi.org/10.1111/j.1541-0420.2006.00573.x

Proust-Lima, C., Amieva, H., Dartigues, J.-F., & Jacqmin-Gadda, H. (2007). Sensitivity of four psychometric tests to measure cognitive changes in brain aging-population-based studies. *American Journal of Epidemiology*, *165*(3), 344–350. https://doi.org/10.1093/aje/kwk017

Proust-Lima, C., Amieva, H., Letenneur, L., Orgogozo, J.-M., Jacqmin-Gadda, H., & Dartigues, J.-F. (2008). Gender and education impact on brain aging: A general cognitive factor approach. *Psychology and Aging*, *23*(3), 608–620. https://doi.org/10.1037/a0012838

Proust-Lima, C., Dartigues, J.-F., & Jacqmin-Gadda, H. (2011). Misuse of the linear mixed model when evaluating risk factors of cognitive decline. *American Journal of Epidemiology*, *174*(9), 1077–1088. https://doi.org/10.1093/aje/kwr243

Sachdev, P. S., Lipnicki, D. M., Crawford, J., Reppermund, S., Kochan, N. A., Trollor, J. N., Draper, B., Slavin, M. J., Kang, K., Lux, O., Mather, K. A., & Brodaty, H. (2012). Risk profiles for mild cognitive impairment vary by age and sex: The Sydney memory and ageing study. *The American Journal of Geriatric Psychiatry : Official Journal of the American Association for Geriatric Psychiatry*, *20*(10), 854–865. https://doi.org/10.1097/JGP.0b013e31825461b0

Samtani, M. N., Raghavan, N., Novak, G., Nandy, P., & Narayan, V. A. (2014). Disease progression model for Clinical Dementia Rating-Sum of Boxes in mild cognitive impairment and Alzheimer's subjects from the Alzheimer's Disease Neuroimaging Initiative. *Neuropsychiatric Disease and Treatment*, *10*, 929–952.

Smith, M., & Buckwalter, K. C. (2005). Behaviors associated with dementia: Whether resisting care or exhibiting apathy, an older adult with dementia is attempting communication. Nurses and other caregivers must learn to 'hear' this language. *The American Journal of Nursing*, *105*(7), 40–52. https://doi.org/10.1097/00000446-200507000-00028

Soldan, A., Pettigrew, C., Cai, Q., Wang, J., Wang, M.-C., Moghekar, A., Miller, M. I., & Albert, M. (2017). Cognitive reserve and long-term change in cognition in aging and preclinical Alzheimer's disease. *Neurobiology of Aging*, *60*, 164–172. https://doi.org/10.1016/j.neurobiolaging.2017.09.002

Tappen, R. M., Rosselli, M., & Engstrom, G. (2010). Evaluation of the Functional Activities Questionnaire (FAQ) in cognitive screening across four American ethnic groups. *The Clinical Neuropsychologist*, *24*(4), 646–661. https://doi.org/10.1080/13854040903482855

Tekin, S., Fairbanks, L. A., O'Connor, S., Rosenberg, S., & Cummings, J. L. (2001). Activities of daily living in Alzheimer's disease: Neuropsychiatric, cognitive, and medical illness influences. *The American Journal of Geriatric Psychiatry*, *9*(1), 81–86. https://doi.org/10.1097/00019442-200102000-00013

Teng, E., Becker, B. W., Woo, E., Knopman, D. S., Cummings, J. L., & Lu, P. H. (2010). Utility of the functional activities questionnaire for distinguishing mild cognitive impairment from very mild Alzheimer disease. *Alzheimer Disease & Associated Disorders*, *24*(4), 348–353. https://doi.org/10.1097/WAD.0b013e3181e2fc84

Tombaugh, T. N., & Mcintyre, N. J. (1992). The mini-mental state examination: a comprehensive review . *Journal of the American Geriatrics Society*, *40*(9), 922–935. https://doi.org/10.1111/j.1532-5415.1992.tb01992.x

Weiner, M. W., Veitch, D. P., Aisen, P. S., Beckett, L. A., Cairns, N. J., Green, R. C., Harvey, D., Jack, C. R., Jagust, W., Liu, E., Morris, J. C., Petersen, R. C., Saykin, A. J., Schmidt, M. E., Shaw, L., Shen, L., Siuciak, J. A., Soares, H., Toga, A. W., & Trojanowski, J. Q. (2013). The Alzheimer's disease neuroimaging initiative: A review of papers published since its inception. *Alzheimer's & Dementia: The Journal of the Alzheimer's Association*, *9*(5), e111–e194. https://doi.org/10.1016/j.jalz.2013.05.1769