Published in final edited form as:

Neuropsychology. 2012 July ; 26(4): 534–540. doi:10.1037/a0028686.

Variability in performance: Identifying early signs of future cognitive impairment

Alyssa A. Gamaldo, PhD¹, Yang An, MS¹, Jason C. Allaire, PhD², Melissa H. Kitner-Triolo, PhD¹, and Alan B. Zonderman, PhD¹

¹Laboratory of Behavioral Neuroscience, National Institute on Aging

²North Carolina State University

Abstract

Objective—The current study examined whether year-to-year variability in cognitive performance differ between individuals cognitively unimpaired and individuals who subsequently develop dementia.

Method—Analyses included a case-control sample of Baltimore Longitudinal Study of Aging (BLSA; M age = 69.90, SD = 8.92) participants. One hundred and thirty five clinically diagnosed demented participants were matched with 135 non-demented participants based on age at initial testing and sex. Cognitive performance was examined using measures of memory, executive function, attention, language, and global mental status performance. Cognitive performance was examined from baseline to 5 years before cognitive impairment (mean assessments = 3.03, SD = 2.80).

Results—Compared to unimpaired individuals, individuals diagnosed with dementia had greater variability on measures of attention, executive function, language, and semantic memory at least 5 years before the estimated onset of cognitive impairment, which may be indicative of maladaptive cognitive functioning. The dementia cases, however, had less variability on visual memory than the unimpaired group, which may suggest that these cases had more difficulty learning.

Conclusions—These results demonstrate that performance variability indexed over annual or biennial visits may be useful in identifying early signs of subsequent cognitive impairment.

Keywords

Cognitive Impairment; Dementia; Neuropsychological tests; Cognition; Variability

Introduction

Clinical aging research has commonly relied on mean levels of performance to differentiate unimpaired from impaired cognitive performance. A compelling question is whether within-person performance variability, average fluctuations around a person's average level of cognitive performance, can be a complementary measure for distinguishing unimpaired from impaired performance, particularly prior to the earliest stages of cognitive impairment. Cognitive performance variability in clinical research has not been explored extensively, but previous research has suggested that within-person fluctuations in performance can differentiate unimpaired cognition from dementia (Burton, Strauss, Hultsch, & Hunter,

2006; Hultsch, MacDonald, Hunter, Levy-Bencheton, Strauss, 2000; MacDonald, Li, & Bäckman, 2009). Furthermore, variability is a reliable index of impairment severity and distinguishes between neurodegenerative disorders, such as Alzheimer's (AD) and Parkinson's disease (PD; Burton et al., 2006).

Various approaches have been used to measure variability in cognitive performance. One approach estimates the average amount of deviation in an individual's performance over time (Hultsch et al., 2000; Allaire & Marsiske, 2005). A second approach estimates within-person fluctuations or changes over time in the context of the multi-level model (MLM; Sliwinski & Buschke, 2004). This second approach incorporates variability by exploring whether within-person changes on one variable is associated within-person changes on another variable, commonly referred to as a within-person coupling relationship. Although the latter approach can better capture a dynamic developmental change, it does not necessarily quantify intraindividual variability on a particular measure as the first approach. The current study estimates intraindividual variability by combining the two approaches by using MLM to estimate the average amount of deviation in an individual's performance over time across a variety of measures and explore group differences in these intraindividual variability estimates.

Variability in cognitive performance has been typically assessed using reaction time measures or latency-based tasks. Given that deficits in many cognitive abilities are often observed among individuals with dementia, it would be useful to determine whether variability over time can be observed and can differentiate unimpaired functioning from impairment across multiple cognitive domains. Furthermore, limited research has explored how far back the distinction in performance variability among individuals can be observed before one of those individuals crosses the threshold for clinical cognitive impairment. A majority of the literature has explored variability using multiple occasions spaced over trials or days (MacDonald et al., 2009). Few studies (Burton, Strauss, Hunter, & Hultsch, 2006; Knotek, Bayles, & Kaszniak, 1990; Murtha, Cismaru, Waechter, & Chertkow, 2002), however, have examined whether these within-person fluctuations can be observed over widely spaced intervals (i.e. annual or biannual intervals), which is more representative of routine examinations by practitioners. The current study explores whether variability across several neuropsychological measures that assess multiple cognitive domains differentiates cognitively unimpaired participants from individuals who subsequently develop dementia, particularly several years before the clinical onset of cognitive impairment. Our hypothesis is that in comparison to unimpaired participants, greater variability will be observed across the neuropsychological measures for impaired participants several years prior to the clinical onset of their impairment.

Methods

Participants

From a sample of 3,036 Baltimore Longitudinal Study of Aging (BLSA) participants, the current study included a sample of 270 participants who had complete data for multiple cognitive measures. The data included in the current study represents data collected in the BLSA from 1970 to 2006. One hundred and thirty-five participants who were diagnosed with dementia (58 females, 77 males), which were labeled as impaired, were matched with 135 cognitively unimpaired participants (58 females, 77 males) based on sex and age at the first testing occasion. The average age of the sample at initial testing was 69.90 ($SD = 8.92$, range = 50.00 – 87.00). Participants in both groups were mostly white (Unimpaired: 84%; Impaired: 96%) and highly educated (Unimpaired: mean years of education = 16.22, $SD = 3.12$; Impaired: mean years of education = 16.89, $SD = 2.29$). The Johns Hopkins University

and the MedStar institutional review boards approved this study, and all participants provided written informed consent.

Measures

A cognitive battery was administered annually or biennially to assess a broad range of cognitive abilities. The Benton Visual Retention Test –Total Errors (BVRT; Sivan, 1982) assessed short-term visual memory. Administration A has been used in the BLSA since 1960, where participants study geometric designs for 10 seconds and then reproduce the designs from memory. The first two designs contain 1 geometric figure and the remaining 8 designs contain two main and one peripheral figure, for 10 designs in total. A modified error scoring system is used, such that higher scores indicate poorer visual memory.

The California Verbal Learning Test (CVLT; Delis, Kramer, Kaplan, & Ober, 1987) was used to assess verbal memory. There were five learning trials, where participants were read a shopping list of 16 items (List A; herbs, fruits, articles of clothing, tools). After each trial the participant freely recalled as many words as they could, as the psychometric technician wrote the correctly recalled words as well as any intrusions and/or perseverations. Participants were then read a new list of 16 shopping items (List B) and, subsequently asked to recall the list of items. After the List B recall, participants were again asked to freely recall List A, followed by a cued recall trial where the four corresponding semantic cue were given. After roughly a 20 minute delay, in which other cognitive tasks were administered, participants were again asked to freely recall List A, followed by cued recall for List A. The following measures were used in analyses: the total number of words correctly recalled across the five List A learning trials (CVLT-LA); The total number of List A words correctly recalled after List B (short-delay free recall: CVLT – SD); the total number of List A words correctly recalled after the 20 minute delay (long-delay free recall: CVLT – LD).

The Trail Making Test was used to assess attention, speed (Trails A), and executive function (Trails B), specifically attention switching and visuomotor scanning. For Trails A participants drew a line as quickly as possible to consecutively numbered circles. When errors were committed, participants corrected the error by returning to his/her last correct response and continuing from there. The stop-watch recorded the time while corrections were made. Trails B is similar to Trails A; however, participants drew a line as quickly as possible to consecutively alternating numbered and letter circles (Reitan, 1992). Again, participants corrected all errors while the stop watch ran. The amount of time taken to complete each part of the task was included in the analyses, and higher scores indicated poorer performance.

Verbal Fluency (Benton & Hamsher, 1976; Goodglass & Kaplan, 1983) was used to assess language, semantic memory, and executive function. The measures consisted of Letter and Category Fluency. For Letter Fluency, participants were given a letter of the alphabet (F, A, S) and asked to generate words beginning with that letter in a minute. For Semantic Fluency, participants were given a category (fruits, animals, and vegetables) and asked to come up with words belonging to the category in a minute. The total number of appropriate words generated was included in the analyses.

Boston Naming Task (BNT; Goodglass & Kaplan, 1983) was used to assess language. Participants were shown 60 line drawings of objects and asked to name them. Participants who did not name the object in 20 seconds were given a stimulus cue (i.e. a category cue or brief description of the purpose of the object). If the object was not recalled after the stimulus cue was given or if the participant could not describe the object, s/he was given a phonemic cue (i.e. pronunciation of the first 2-3 letters of the object). The number of correct responses given before cues was included in the analyses.

The Mini-Mental State Examination (MMSE; Folstein, Folstein, & McHugh, 1975) was used to assess mental status. The MMSE is commonly used to measure orientation, memory, attention, calculation, language, and visuospatial skills. A total correct response (out of 30) was used in analyses.

Health measures, such as cerebrovascular disease and stroke, were measured by a routine physical exam at each BLSA visit as was self-reported assessment of health status. These measures were compared between the two groups to determine if they should be included in the analyses as potential covariates.

Dementia Classification

Dementia diagnosis was determined from a clinical consensus agreement among physicians, neuropsychologists, cognitive psychologists, and gerontologists using the Diagnostic and Statistical Manual of Mental Disorders Criteria, Third Edition, Revised Criteria (American Psychiatric Association, 1987) and National Institute of Neurological and Communication Disorders-Alzheimer's Disease and Related Disorders Association Criteria (McKhann, Drachman, Folstein, Katzman, Price, & Stadlan, 1984). Dementia cases were further classified in diagnostic categories (i.e. probable AD, possible AD, vascular dementia, PD, mixed dementia, unspecified dementia) using the Hachinski ischemic scores, history of cardiovascular disease, examination, and neuroimaging. A more detail description of diagnostic category protocol has been previously published (Kawas, Gray, Brookmeyer, Fozard, & Zonderman, 2000). Within our sample, 77% of the dementia cases were classified as AD (probable AD, possible AD), 7% Vascular dementia, 4% PD, 11% Unspecified, and 1% other. For each dementia case, the clinical consensus panel estimated date of initial cognitive impairment onset and date of dementia diagnosis. The onset date was identified as a marker of initial cognitive functioning changes based upon objective cognitive testing as well as functional changes. For the purposes of the current study's aims, only cognitive testing that occurred prior to the onset date was examined to explore for potential fluctuations in cognitive performance even though the participant's cognitive performance during this time did not show signs of cognitive impairment. Thus, cognitive testing occurred after the onset date was excluded for all dementia cases (see Figure 1). Task performance was examined across a time span window, which accounted for any cognitive assessments that occurred from initial testing to the testing session at least 5 years before the estimated onset date of cognitive impairment. Within this time span, the impaired group had roughly 3.03 assessments ($SD = 2.80$) conducted over an average of 5.59 years ($SD = 6.15$). The matched unimpaired group had roughly 4.13 assessments ($SD = 3.71$). This time span window included visits from the impaired group that were on average 7.11 ($SD = 2.38$, range = 5-17) years before the estimated onset of cognitive impairment and 9.37 years ($SD = 3.25$, range = 5-20) before the estimated date of dementia diagnosis.

Analyses

We used linear mixed-effects models (Hox, 2002; Laird & Ware, 1982), also known as multilevel models (MLM), for each cognitive measure using SAS System software (version 9.2; SAS Institute). These models allow us to make inference on population average parameters (fixed effects) as well as to separate the variability in the data into between subjects (random effects) and within subjects (residual variance) and make inference on these variability parameters.

In detail, we applied the following multilevel models (MLM) to each of the cognitive measures:

$$y_{ij} = \beta_0 + \beta_1 \text{group}_i + \beta_2 \text{time}_{ij} + \beta_3 \text{group}_i \times \text{time}_{ij} + b_{0i} + b_{1i} \text{time}_{ij} + \varepsilon_{ij}$$

Where, y_{ij} is the cognitive score for the i th subject on the j th follow-up test visit. Group is a binary variable that indicates whether the i th participant is either impaired or unimpaired. The variable time is the longitudinal predictor coded as the time (yrs) from baseline age (1st time is 0). β . denote population specific (fixed effects) estimates, for example β_1 represents the average group difference at baseline and β_3 represents the average group difference in the longitudinal rates of change. $b_{..}$ denote subject specific random effects estimates, b_{0i} is random intercept and is b_{1i} random slopes. We allowed the initial values and slopes to vary among the participants. $b_{..} \sim N(0, D)$, where D here is the 2×2 covariance matrix, and it represents the between subject variability. $\varepsilon_{..}$ is the residual error and we assumed that conditional on random effects is independent and $\varepsilon_{..} \sim N(0, \sigma^2)$. It represents the within subject variability.

To test for the equality of the within-person variability, we ran two models, which differ only in the residual error term, which estimates the within-person/group variance of performance over time. The first model (null model) assumed that the within-person variability was the same between unimpaired and impaired. In this model, the residual term is distributed $\varepsilon_{..} \sim N(0, \sigma^2)$ and we used one σ^2 term parameter for both groups. The second model (hypothesized model) assumed that within-person variability was different between unimpaired and impaired and we used two σ^2 term parameters to be estimated for each group. The first model is nested in the second model. This allows us to statistically test our primary aim of whether within-person cognitive performance variability was different between the clinical status groups using likelihood ratio tests. Likelihood based Akaike information criterion (AIC) was also used to determine which model provides a better fit to the data. Restricted maximum likelihood (REML) estimation was used for all the models because it generates the unbiased estimates of variance components.

Results

None of the participants had a reported history of a stroke. In addition, there was no significant difference between the groups on cardiovascular disease ($\chi^2(1) = 3.08, p = .078$); 8.9% of the impaired participants and 3.7% of the unimpaired participants had a diagnosis of cardiovascular disease. As such, these health factors were not included in the subsequent analyses as covariates.

Model Fit

The hypothesized model fit the data better than the null model for BVRT ($\chi^2(1) = 6, p < .05$), Trails A ($\chi^2(1) = 29, p < .001$), Trails B ($\chi^2(1) = 17, p < .001$), Category Fluency ($\chi^2(1) = 14, p < .001$), and BNT ($\chi^2(1) = 4, p < .05$), suggesting that within-person performance variability was significantly different between the clinical status groups. However, the hypothesized model did not fit the data better than the null model for CVLT-LA ($\chi^2(1) = 0, p > .05$), CVLT-SD ($\chi^2(1) = 0, p > .05$), CVLT-LD ($\chi^2(1) = 1, p > .05$), Letter Fluency ($\chi^2(1) = 3, p > .05$), and MMSE ($\chi^2(1) = 0, p > .05$), suggesting that within-person performance variability was not significantly different between the clinical status groups.

Variability in Performance

Table 1 illustrates the random and residual effect estimates for the impaired and unimpaired groups. The demented group had roughly twice as much within-person performance variability (σ^2) as the cognitively unimpaired group for Trails A, Trails B, and Category

Fluency. However, the impaired group had less within-person performance variability on the BVRT than the cognitively unimpaired group. We did not observe significant fixed effects (Table 2) or baseline cognitive performance differences between impaired and unimpaired groups on any of the measures. In addition, we did not observe significant differences in performance over time between the groups on any of the measures (Table 2).

Discussion

Fluctuating performance on the Trail Making Test, Category Fluency, Boston Naming Test and BVRT distinguishes cognitively unimpaired individuals from individuals eventually diagnosed with dementia as early as 5 years before the onset of their impairment. Our findings suggest cognitively impaired individuals had, on average, greater performance variability on the Trail Making Test, Category Fluency, and Boston Naming, which are measures of attention, executive function, semantic memory and language over a time span of several years before the onset of their cognitive impairment than cognitively unimpaired individuals. Given that these cognitive measures tap various cognitive abilities, it is difficult to disentangle the specific abilities that are particularly sensitive to performance variability. In contrast to these findings, the cognitively impaired individuals had less performance variability on the BVRT, a measure of visual memory, than cognitively unimpaired individuals. Our findings for the Trails Making Test, Category Fluency, and Boston Naming Test support previous literature suggesting that performance variability may indicate maladaptive cognitive functioning (MacDonald et al., 2009). Our contradictory finding for the BVRT was unexpected, and the explanation for this result is unclear at this point. However, a possible explanation is that variability on some tasks may indicate adaptive functioning (Allaire & Marsiske, 2005; Li, Aggen, Nesselrode, & Baltes, 2001).

The adaptive function of performance variability has been observed during the administration of a novel measure, particularly a measure that allows for practice-related gains, such as an accuracy-based task (Allaire & Marsiske, 2005; Li et al., 2001). For children and older adults, adaptive variability reflects strategy use and learning (Allaire & Marsiske, 2005; Siegler, 1994). Considering the novelty and accuracy-based component of the BVRT, greater performance variability for the unimpaired group indicate strategy use, but reduced performance variability for the impaired group may indicate reduced strategy use. It has been proposed that once an individual has adapted and/or revised their strategies to perform a task successfully, further performance variability may suggest maladaptive functioning (MacDonald et al., 2009). However, given the limited number of assessments that we had in our study, we could not test the possibility that performance variability for our unimpaired group may be less apparent in subsequent testing occasions because they have learned the task. Thus, further research is needed to support the proposal that decreased within-person variability for pre-dementia cases may reflect maladaptive functioning.

In contrast to adaptive variability, maladaptive variability has been typically observed for measures of latency or reaction time, which are unlikely to have substantial practice-related gains with repeated testing occasions (Allaire & Marsiske, 2005). Given our measures of attention (Trail Making Testing and Category Fluency) have a latency component, the observed enhanced variability for the impaired group may indicate maladaptive functioning. Indeed, these findings support previous literature that suggests that more severe neurological dysfunction is associated with greater variability in performance, particularly on latency-based measures (Hultsch, Strauss, Hunter, & MacDonald, 2008).

Maladaptive variability in cognitive performance may be influenced by endogenous sources (i.e. central nervous system compromise) or exogenous sources (i.e. stress and fatigue; MacDonald et al., 2009). The literature is inconclusive regarding the underlying

mechanisms that explain the endogenous sources of performance variability, but some research suggests that variability is a proxy of general deficits in central nervous system processing, such as errors in the transmission of neural signals or inefficient neurotransmitter systems (MacDonald et al., 2009). Furthermore, the association between variability and structural brain correlates, such as white matter volume reduction (Walhovd & Fjell, 2007), white matter hyperintensities (Bunce, Anstey, Christensen, Dear, Wen, & Sachdev, 2007), white matter demyelination (Britton, Meyer, & Benecke, 1991), and lesions in frontal gray matter (Stuss, Murphy, Binns, & Alexander, 2003), also suggest that increased variability may specifically reflect impaired attentional and executive control processes (Hultsch et al., 2008). Interestingly, we consistently observed significant differences in variability between our groups on tasks of attention and executive control over time, which may suggest that the variability parameter is particularly sensitive in detecting oscillations in these abilities and that these abilities may show early fluctuations in performance (Dixon, Garrett, Lentz, MacDonald, Strauss, & Hultsch, 2007).

A number of studies have suggested that factors, such as baseline or average performance levels, on measures of memory (Albert, Moss, Tanzi, & Jones, 2001; Amieva et al., 2008; Elias, Beiser, Wolf, Au, White, D'Agostino, 2000; Kawas, Corrada, Brookmeyer, Morrison, Resnick, Zonderman, & Avenberg, 2003; Small, Fratiglioni, Viitanen, Winbald, & Bäckman, 2000; Tierney, Yao, Kiss, & McDowell, 2005), abstract reasoning (Elias et al., 2000; Fabrigoule et al., 1996), visuospatial ability and executive function (Albert et al., 2001; Fabrigoule et al., 1996), language (Albert et al., 2001; Amieva et al., 2008; Tierney et al., 2005), attention (Fabrigoule et al., 1996), and global mental status (Small et al., 2000) are associated with the risk for developing dementia, particularly Alzheimer's disease. Although we observed performance variability differences between the unimpaired and demented group on some of these domains, we did not observe significant group differences in initial performance and rates of cognitive change, particularly on memory tasks which are typically the earliest sign of Alzheimer's disease (Albert et al., 2001; Amieva et al., 2008; Elias et al., 2000; Kawas et al., 2003; Small et al., 2000; & Tierney et al., 2005). This may partially be because our assessments occurred before the onset of clinical impairment. However, our smaller sample size relative to some of these previous studies may also have limited our power to detecting significant differences between the groups.

It is difficult to directly compare our results with earlier studies of cognitive performance because few of the measures among studies overlap. With some exceptions (Elias et al., 2000; Kawas et al., 2003; Amieva et al., 2008), these studies have not explored cognitive performance more than 5 years before dementia diagnosis. Additionally, as our study included assessments before the onset date of initial cognitive impairment rather than date of dementia diagnosis, the time span we were exploring was several years earlier than that commonly used in previous studies making direct comparisons more difficult. Interestingly, we were able to observe significant clinical group differences in within-person variability on several of the abilities (i.e., attention, executive function, language, and memory) that have been previously shown to decline in pre-dementia patients (Albert et al., 2001; Amieva et al., 2008; Fabrigoule et al., 1996). For example, in a sample of AD patients, Amieva and colleagues (2008) observed declines in language performance (i.e., semantic naming and word retrieval) at least 12 years prior to dementia diagnosis, suggesting that temporal brain regions are particularly vulnerable to disease pathology. Although our findings did not support average group differences in the longitudinal rates of change for language, we did observe that for these measures our impaired group had greater within-person variability than the unimpaired group. Thus, our findings may further validate the use of models exploring fluctuations in performance to detect those at risk of future dementia.

Although this study produced intriguing findings, there are limitations. First, the sample size was relatively small, which may partially explain why we did not observe significant differences in the residual component parameters between the demented and non-demented groups for many cognitive measures. However, relative to previous studies that have examined cognitive performance within-person variability in clinical demented samples (Burton et al., 2006; Hultsch et al., 2000; Murtha et al., 2002), our study has included one of the largest samples of clinical diagnosed demented patients. Second, the sample was highly educated and primarily white. It should also be noted that the BLSA is a convenience sample, and not a representative sample of the population. Thus, our preliminary findings should be further explored in a larger, more representative sample of the population. Third, our study is limited by our use of the untenable assumption of missing at random (MAR) to facilitate our analyses. It is possible that participants who had not completed the measures due to poor cognitive functioning may have been excluded from the analyses due to missing data. As a result, our sample would be positively bias. However, the true mechanism of missing values may never be known and may vary on a case by case basis. Future investigations can explore the use of multiple imputations (MI) to account for missing values. Alternatively, assumptions regarding the underlying cause of missing data can be specified and incorporated into the statistical models. An example is the pattern-mixture model in longitudinal analyses. Despite these limitations common in most observational studies, the BLSA sample provides a rich source of prospectively collected cognitive data in a well-characterized cohort with valuable opportunities to investigate cognitive aging in relation to changes in physical and psychological health. Fourth, previous literature has typically measured within-person/group variability on one occasion (i.e. multiple trials within a session) or multiple occasions within a short span of time (i.e. daily measurements). The current study measured within-group variability over annual and biennial intervals, which may not be representative of reversible and fast changes observed with shorter intervals (i.e. trial-to-trial assessments; MacDonald et al., 2009). However, our results appear to be consistent with some of the previous literature that have included these shorter intervals, which may suggest that variability over extended time intervals is detectable early in the course of disease. Furthermore, examining variability across longer time intervals is applicable to general practice. Previous studies that have observed cognitive performance variability differences between cognitively impaired participants and unimpaired participants have tended to assess performance over daily or weekly spaced intervals (Burton et al., 2006; Knotek et al., 1990; Murtha et al., 2002). Annual observations are more consistent with the time frame when individuals visit a general practitioner for a routine physical. Our results cannot provide evidence that variability indexed over years is reflective of changes in endogenous sources. More research is needed to explore whether variability indexed over years can be linked to endogenous sources, particularly with brain-imaging techniques.

Variability in performance appears to be a useful measure for the early identification of individuals who will eventually develop cognitive impairment. Our results suggest that variability assessed over years may be useful in detecting early signs of impairment, and that this variability occurs at least 5 years prior clinical impairment. It is important; however, to next explore whether our results are consistent using micro-longitudinal assessments with shorter testing intervals (i.e. trails, hours, days) at least 5-10 years before clinical impairment. More research is also needed to determine if this approach is associated with neuropathology as well as structural and functional brain changes as assessed by brain-imaging techniques.

Acknowledgments

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

References

- Albert MS, Moss MB, Tanzi R, Jones K. Preclinical prediction of AD using neuropsychological tests. *Journal of the International Neuropsychological Society*. 2001; 7:631–639. [PubMed: 11459114]
- Allaire JC, Marsiske M. Intraindividual variability may not always indicate vulnerability in elders' cognitive performance. *Psychology and Aging*. 2005; 20:390–401. [PubMed: 16248699]
- Amieva H, Le Goff M, Millet X, Orgogozo JM, Pérès K, Barberger-Gateau P, Jacqmin-Gadda H, Dartigues JF. Prodromal Alzheimer's Disease: Successive emergence of clinical symptoms. *Annals of Neurology*. 2008; 64:492–498. [PubMed: 19067364]
- American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders*. 3rd ed. American Psychiatric Association; Washington, DC: 1987.
- Benton, AL.; Hamsher, K.; de, S. *Multilingual Aphasia Examination*. University of Iowa; Iowa City: 1976.
- Britton TC, Meyer BU, Benecke R. Variability of cortically evoked motor responses in multiple sclerosis. *Electroencephalography and Clinical Neurophysiology*. 1991; 81:186–194.
- Bunce DJ, Anstey KJ, Christensen H, Dear K, Wen W, Sachdev P. White matter hyper intensities and within-person variability in community-dwelling adults 60 to 64 years. *Neuropsychologia*. 2007; 45:2009–2015. [PubMed: 17382358]
- Burton CL, Strauss E, Hultsch DF, Hunter MA. Intraindividual variability as a marker of neurological dysfunction: A comparison of Alzheimer's disease and Parkinson's disease. *Journal of Clinical and Experimental Neuropsychology*. 2006; 28:67–83. [PubMed: 16448976]
- Delis, DC.; Kramer, JH.; Kaplan, E.; Ober, BA. *California Verbal Learning Test (CVLT)*. Adult Version. Psychological Corporation; San Antonio: 1987.
- Dixon RA, Garrett DD, Lentz TL, MacDonald SWS, Strauss E, Hultsch DF. Neurocognitive markers of cognitive impairment: Exploring the roles of speed and inconsistency. *Neuropsychology*. 2007; 21:381–399. [PubMed: 17484601]
- Elias MF, Beiser A, Wolf PA, Au R, White RF, D'Agostino RB. The preclinical phase of Alzheimer Disease. *Archives of Neurology*. 2000; 57:808–813. [PubMed: 10867777]
- Fabrigoule C, Lafont S, Letenneur L, Rouch I, Dartigues JF. WAIS similarities subtest performances as predictors of dementia in elderly community residents. *Brain Cognition*. 1996; 30:323–326.
- Folstein MF, Folstein SE, McHugh PR. "Mini-Mental State": A practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research*. 1975; 12:189–198. [PubMed: 1202204]
- Goodglass, H.; Kaplan, E. *Boston Diagnostic Aphasia Examination (BDAE)*. Lea and Febiger; Philadelphia, Pennsylvania: 1983.
- Hox, J. *Multilevel analysis: Techniques and applications*. Erlbaum; Mahwah: 2002.
- Hultsch DF, MacDonald SWS, Hunter MA, Levy-Bencheton J, Strauss E. Intraindividual variability in cognitive performance in the elderly: Comparison of adults with mild dementia, adults with arthritis, and healthy adults. *Neuropsychology*. 2000; 14:588–598. [PubMed: 11055261]
- Hultsch, DF.; Strauss, E.; Hunter, MA.; MacDonald, SWS. Intraindividual variability, cognition, and aging. In: Craik, FIM.; Salthouse, TA., editors. *The Handbook of Aging and Cognition*. 3rd ed. Psychology Press; New York: 2008. p. 491-556.
- Kawas C, Gray S, Brookmeyer R, Fozard J, Zonderman A. Age-specific incidence rates of Alzheimer's disease: The Baltimore longitudinal study of aging. *Neurology*. 2000; 54:2072–2077. [PubMed: 10851365]
- Kawas CH, Corrada MM, Brookmeyer R, Morrison A, Resnick SM, Zonderman AB, Avenberg D. Visual memory predicts Alzheimer's disease more than a decade before diagnosis. *Neurology*. 2003; 60:1069–1093.
- Knotek PC, Bayles KA, Kaszniak AW. Response consistency on a semantic memory task in persons with dementia of the Alzheimer type. *Brain and Language*. 1990; 38:465–475. [PubMed: 2375976]
- Laird NM, Ware JH. Random-Effects Models for Longitudinal Data. *Biometrics*. 1982; 38:963–974. [PubMed: 7168798]

- Li SC, Aggen SH, Nesselroade JR, Baltes PB. Short-term fluctuations in elderly people's sensorimotor functioning predict text and spatial memory performance: The MacArthur Successful Aging Studies. *Gerontology*. 2001; 47:100–116. [PubMed: 11287736]
- MacDonald SWS, Li S-C, Bäckman L. Neural underpinnings of within-person variability in cognitive functioning. *Psychology and Aging*. 2009; 24:792–808. [PubMed: 20025396]
- McKhann G, Drachman D, Folstein M, Katzman R, Price D, Stadlan EM. 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*. 1984; 34:939–944. [PubMed: 6610841]
- Murtha S, Cismaru R, Waechter R, Chertkow H. Increased variability accompanies frontal lobe damage in dementia. *Journal of the International Neuropsychological Society*. 2002; 8:360–372. [PubMed: 11939695]
- Reitan, R. Trail Making Test: Manual for Administration and Scoring. Reitan Neuropsychological Laboratory; Tucson: 1992.
- Siegler RS. Cognitive variability: A key to understanding cognitive development. *Current Directions in Psychological Science*. 1994; 3:1–5.
- Sivan, AB. Benton Visual Retention Test. 5th ed. The Psychological Corporation; San Antonio: 1982.
- Sliwinski M, Buschke H. Modeling intraindividual cognitive change in aging adults: Results from the Einstein Aging Studies. *Aging Neuropsychology and Cognition*. 2004; 11:196–211.
- Small BJ, Fratiglioni L, Viitanen M, Winbald B, Bäckman L. The course of cognitive impairment in preclinical Alzheimer Disease. *Archives of Neurology*. 2000; 57:839–844. [PubMed: 10867781]
- Stuss DT, Murphy KJ, Binns MA, Alexander MP. Staying on the job: The frontal lobes control individual performance variability. *Brain*. 2003; 126:2363–2380. [PubMed: 12876148]
- Tierney MC, Yao C, Kiss A, McDowell J. Neurological tests accurately predict incident Alzheimer disease 5 and 10 years. *Neurology*. 2005; 64:1853–1859. [PubMed: 15955933]
- Walhovd KB, Fjell AM. White matter volume predicts reaction time instability. *Neuropsychologia*. 2007; 45:2277–2284. [PubMed: 17428508]

Figure 1.
Assessed Time Span for Performance Variability within Dementia Group.

Table 1
Random and residual effects for groups across the cognitive measures

Measures	Intercept, b_{0i} (D_{11})	Slope, b_{1i} (D_{22})	Within-person variability Impaired (σ^2)	Within-person variability Unimpaired (σ^2)
BVRT *	8.87 (1.08)	0.08 (0.04)	3.57 (0.44)	5.24 (0.43)
CVLT-LA	94.40 (20.34)	1.52 (0.68)	23.39 (9.94)	28.83 (2.95)
CVLT-SD	8.39 (1.73)	0.11 (0.05)	2.23 (0.86)	2.40 (0.25)
CVLT-LD	8.41 (1.77)	0.13 (0.07)	4.32 (1.61)	2.67 (0.28)
Trails A *	144.60 (34.53)	0.00 (0.00)	266.24 (38.35)	110.32 (10.22)
Trails B *	867.12 (236.25)	26.84 (29.41)	1309.38 (202.56)	633.22 (76.28)
BNT *	24.92 (4.53)	0.05 (0.04)	6.41 (2.51)	3.22 (0.33)
Letter	14.35 (2.23)	0.07 (0.04)	2.27 (0.38)	3.18 (0.28)
Category *	5.21 (1.04)	0.08 (0.04)	4.11 (0.65)	2.14 (0.20)
MMSE	1.08 (0.33)	0.02 (0.02)	1.33(0.20)	1.27 (0.14)

Note:

* significant within-person variability differences between the clinical status groups. Standard errors are in parentheses.

Table 2
Group average (fixed effects) estimates across the cognitive measures

Measures	Baseline				Slope			
	Impaired ($\beta_0 + \beta_1$)	Unimpaired (β_0)	Difference (β_1)	<i>p</i>	Impaired ($\beta_2 + \beta_3$)	Unimpaired (β_2)	Difference (β_3)	<i>p</i>
BVRT	6.00 (0.30)	6.30 (0.31)	0.30 (0.43)	0.48	0.40 (0.09)	0.25 (0.07)	-0.15 (0.11)	0.19
CVLT-LA	50.60 (5.34)	51.61 (1.30)	1.01 (5.50)	0.85	-0.72 (0.87)	-0.86 (0.26)	-0.13 (0.90)	0.88
CVLT-SD	10.71 (1.56)	10.01 (0.38)	-0.70 (1.61)	0.67	-0.13 (0.25)	-0.18 (0.07)	-0.04 (0.26)	0.87
CVLT-LD	9.63 (1.74)	10.50 (0.39)	0.87 (1.78)	0.63	0.11 (0.27)	-0.16 (0.08)	-0.27 (0.28)	0.34
Trails A	45.80 (3.24)	41.82 (1.75)	-3.98 (3.68)	0.28	0.80 (0.69)	-0.23 (0.27)	-1.03 (0.74)	0.17
Trails B	107.02 (7.46)	93.40 (4.21)	-13.62 (8.57)	0.11	1.72 (1.99)	2.75 (1.14)	1.03 (2.29)	0.66
BNT	51.55 (2.44)	53.59 (0.63)	2.05 (2.52)	0.42	0.36 (0.35)	-0.13 (0.07)	-0.49 (0.36)	0.18
Letter	15.71 (0.59)	15.73 (0.44)	0.02 (0.74)	0.98	-0.25 (0.11)	-0.13 (0.06)	0.12 (0.12)	0.34
Category	15.66 (0.48)	15.50 (0.29)	-0.16 (0.56)	0.77	-0.25 (0.12)	-0.28 (0.06)	-0.02 (0.13)	0.86
MMSE	28.45 (0.23)	28.45 (0.16)	0.00 (0.28)	0.99	0.03 (0.05)	0.04 (0.04)	0.01 (0.06)	0.84

Note: Standard errors are in parentheses.