Journal of the International Neuropsychological Society (2016), **22**, 1016–1025.

doi:10.1017/S135561771600093X

Intraindividual Cognitive Variability in Middle Age Predicts Cognitive Impairment 8–10 Years Later: Results from the Wisconsin Registry for Alzheimer's Prevention

Rebecca L. Koscik,[1] Sara E. Berman,[2] Lindsay R. Clark,[1,2] Kimberly D. Mueller,[1] Ozioma C. Okonkwo,[1] Carey E. Gleason,[3,1,2] Bruce P. Hermann,[1,4] Mark A. Sager,[1] AND Sterling C. Johnson[3,1,2]

[1]Wisconsin Alzheimer's Institute, University of Wisconsin School of Medicine and Public Health, Madison, Wisconsin
[2]Alzheimer's Disease Research Center, University of Wisconsin School of Medicine and Public Health, Madison, Wisconsin
[3]Geriatric Research Education and Clinical Center, Wm. S. Middleton Veterans Hospital, Madison Wisconsin
[4]Department of Neurology, University of Wisconsin School of Medicine and Public Health, Madison, Wisconsin

(RECEIVED February 29, 2016; FINAL REVISION September 15, 2016; ACCEPTED October 12, 2016)

Abstract

Objectives: Intraindividual cognitive variability (IICV) has been shown to differentiate between groups with normal cognition, mild cognitive impairment (MCI), and dementia. This study examined whether baseline IICV predicted subsequent mild to moderate cognitive impairment in a cognitively normal baseline sample. **Methods:** Participants with 4 waves of cognitive assessment were drawn from the Wisconsin Registry for Alzheimer's Prevention (WRAP; $n = 684$; 53.6(6.6) baseline age; 9.1(1.0) years follow-up; 70% female; 74.6% parental history of Alzheimer's disease). The primary outcome was Wave 4 cognitive status ("cognitively normal" *vs*. "impaired") determined by consensus conference; "impaired" included early MCI ($n = 109$), clinical MCI ($n = 11$), or dementia ($n = 1$). Primary predictors included two IICV variables, each based on the standard deviation of a set of scores: "6 Factor IICV" and "4 Test IICV". Each IICV variable was tested in a series of logistic regression models to determine whether IICV predicted cognitive status. In exploratory analyses, distribution-based cutoffs incorporating memory, executive function, and IICV patterns were used to create and test an MCI risk variable. **Results:** Results were similar for the IICV variables: higher IICV was associated with greater risk of subsequent impairment after covariate adjustment. After adjusting for memory and executive functioning scores contributing to IICV, IICV was not significant. The MCI risk variable also predicted risk of impairment. **Conclusions:** While IICV in middle-age predicts subsequent impairment, it is a weaker risk indicator than the memory and executive function scores contributing to its calculation. Exploratory analyses suggest potential to incorporate IICV patterns into risk assessment in clinical settings. (*JINS*, 2016, *22*, 1016–1025)

Keywords: Prodromal period, Mild cognitive impairment, Alzheimer's disease, Neuropsychological tests, Risk screening, Psychometrics

INTRODUCTION

Efforts to prevent or delay the onset of Alzheimer's disease (AD) will be more effective if started as early in the disease process as possible necessitating sensitive methods for early detection and risk assessment. Recently, Imtiaz and colleagues identified seven mid-life risk scoring systems for predicting subsequent dementia (Imtiaz, Tolppanen, Kivipelto, & Soininen, 2014). These systems were based on 7 to 15 variables each, including multiple demographic variables (e.g., age, education, or gender) and multiple health status variables (e.g., diabetes status, blood pressure, body mass index, or cholesterol). While higher scores were associated with higher risk of dementia across all systems reviewed, predictive accuracy was low while obtaining all risk factor information was resource intensive.

Methods to detect preclinical cognitive changes associated with prodromal AD exist which primarily focus on individual cognitive test scores and published norms; these methods have also been shown to have limited sensitivity in identifying early changes before the syndromes of mild cognitive

Correspondence and reprint requests to: Rebecca Koscik, Wisconsin Alzheimer's Institute, 7818 Big Sky Drive, Suite 215, University of Wisconsin School of Medicine and Public Health, Madison, WI 53719. E-mail: rekoscik@wisc.edu

impairment (MCI) and dementia (Koepsell & Monsell, 2012; Sperling, Karlawish, & Johnson, 2013).

The standard deviation (*SD*) of multiple cognitive performance scores at a given assessment, also known as intraindividual cognitive variability (IICV) or cognitive dispersion, has been shown to differentiate between cognitively healthy and clinical groups such as MCI and AD (Kälin et al., 2014; Salthouse & Soubelet, 2014) and may potentially identify persons at risk of declining to MCI or dementia. Research on the association between IICV and cognitive decline has been extended to determine its predictive value in older populations that are cognitively normal at baseline. For example, in a study of 897 older individuals (mean age of 78.6 years) who underwent follow-up examinations every 12–18 months, the inclusion of IICV in the model improved prediction of subsequent dementia (Holtzer, Verghese, Wang, Hall, & Lipton, 2008). Similarly, within subject variability across multiple cognitive domains was examined in 204 community-dwelling adults (mean age 74). Increasing age and declining cognitive status, relative to premorbid functioning, were positively associated with the degree of IICV (Hilborn, Strauss, Hultsch, & Hunter, 2009). Analyses of over 2000 older women found that after adjusting for differences in overall cognitive performance, increased variability was associated with a greater risk of developing dementia (Vaughan et al., 2013).

While much of the IICV literature has focused on IICV as a predictor of subsequent functioning in older baseline samples (mid-sixties and above), IICV may be a useful method for identifying risk of cognitive decline among cognitively healthy middle-aged individuals. Thus, the primary aims of this study were to: (1) examine whether baseline IICV measured in early to late middle-age predicted mild to moderate cognitive impairment approximately nine years later in a sample that was free of clinical impairment and middle-aged at baseline; and (2) determine whether baseline IICV improved predictions of subsequent MCI status over predictions based on traditional measures of cognitive functioning (e.g., continuous neuropsychological test scores). To encourage consideration of how IICV might be incorporated into clinical practice, the exploratory aim examined whether a set of cut-points that incorporated IICV patterns across a small set of neuropsychological tests might allow clinicians to incorporate IICV into their assessments of patients' risk of meeting criteria for impairment.

METHODS

Participants

Participants were drawn from the Wisconsin Registry for Alzheimer's Prevention (WRAP), an ongoing longitudinal cohort study examining cognitive trajectories and associated risk factors in a sample that was middle-aged and enriched for AD risk at baseline ($n = 1540$ with baseline assessment; mean (*SD*) baseline age = 53.6(6.7) years; 72.4% with a parental family history of AD) (Sager, Hermann, & La Rue, 2005). To ensure that individuals were not cognitively impaired when they enrolled in WRAP, baseline performances were screened by a physician and neuropsychologist throughout the study and participants were excluded post consent if their baseline performance indicated clinical MCI or dementia. The first follow-up visit occurs 4 years after the baseline visit; all subsequent visits occur at 2-year intervals thereafter. Participants for this analysis were selected on the basis of having completed four waves of cognitive assessment, were free at baseline of neurological diagnoses (stroke, Parkinson's disease, MS, or epilepsy/seizure disorder), speak English as their native language, and have had their fourth wave cognitive status reviewed *via* consensus conference ($n = 684$). All activities for this study were approved by the Institutional Review Board and completed in accordance with the Helsinki Declaration.

Cognitive Variables

At each study visit, participants complete a multi-domain neuropsychological battery and questionnaires covering demographic, health, and lifestyle characteristics (Sager et al., 2005). Factor analyses of the baseline test battery identified 6 factors, two representing broad domains of visuospatial and verbal abilities, two learning and memory factors (Immediate Memory, Verbal Learning, and Memory), and two factors pertaining to attention and executive function [Speed and Flexibility, Working Memory; see (Dowling, Hermann, La Rue, & Sager, 2010; Koscik et al., 2014) for details]. Residualized *Z*-scores were created for each factor based on regression parameters for age, gender, and literacy derived from a robust normative subsample that was at lower genetic risk for AD and cognitively normal through at least 4 years of follow-up (Clark et al., 2016; Koscik et al., 2014). Additional memory and executive function tests were added to follow-up assessment waves and were, therefore, available for use in determining Wave 4 cognitive outcomes but not for determining baseline IICV.

Calculation of baseline IICV (primary predictors)

One common approach to estimating IICV is to calculate the *SD* across factors or test scores for a given testing session using either *Z*-scores based on the mean (*SD*) of each score that is contributing to the IICV estimate or residualized *Z*-scores (after adjusting for covariates). Scores used in the literature include individual neuropsychological test scores (Hilborn et al., 2009; Holtzer et al., 2008) or cognitive domain scores derived from factor analysis or other clustering procedures (Salthouse & Soubelet, 2014). The use of a small number of test scores is appealing due to its potential for easy translation to clinical settings while the use of factor scores is appealing because they condense related variables into one summary score per cognitive domain and are thought to yield more stable estimates of functioning within domains (Grice, 2001). For this study, we examined the predictive properties of both a complex and a simple calculation of IICV.

First, a complex IICV variable was developed by calculating the variability across the six baseline cognitive factor *Z*-scores described above. Specifically, we calculated the standard deviation for each person across the six *Z*-scores and converted this set of *SD*s to a normal distribution (e.g., ~N(0,1)) to create the variable, "6 Factor IICV."

Second, to estimate whether IICV might be useful in a clinic setting based on a simple cognitive screen representing multiple cognitive domains, we calculated a 4-test IICV using the baseline Rey Auditory Verbal Learning Task (AVLT) (Schmidt, 1996) sum of learning trials score , Trail Making Test (Heaton, Miller, Taylor, & Grant, 2004) A and B completion times (Trails A and B, seconds), and the Wide Range Achievement Test-3rd edition (Wilkinson, 1993) (WRAT-III) reading subtest standard score. The tests were selected to represent multiple cognitive domains known to change earlier (AVLT and Trails B; verbal memory, executive functioning) *versus* later (Trails A and WRAT-III; attention, verbal ability) in the progression from prodromal to clinical dementia due to AD. Trails A and B were first log-transformed and multiplied by −1 so that higher scores indicated better performance across all 4 variables. Raw scores for each of the 4 variables were standardized to *Z*-scores (e.g., ~N(0,1)) using the sample mean and *SD* for each test. The *SD* across the four *Z*-scores was then calculated for each person. Last, this set of *SD*s was also converted to a set of *Z*-scores called "4 Test IICV."

Cognitive outcome

The primary outcome was participant cognitive status at the fourth wave of assessment as determined by consensus review by a panel of dementia experts (clinical neuropsychologists, physicians, and clinical nurse practitioners). The consensus conference review process was initiated in 2012 to classify participants based on their most recent WRAP assessment data into outcomes on the trajectory from cognitively normal to dementia endpoints. All data collected in the month before a given conference are analyzed first *via* an algorithm which selects cases for discussion at the consensus conference if they meet *one or more* of the following criteria: (1) cognitive abnormalities (i.e., at least 1.5 *SD*s below expected relative to robust internal norms adjusting for age, gender, and literacy-level) on: (a) the most recent assessment for factor scores or individual measures of memory, executive function, language, working memory, or attention (Clark et al., 2016), or (b) any two assessments for factor scores representing these domains (Koscik et al., 2014); (2) cognitive performance on one or more tests fell below values used in other studies as cutpoints for mild cognitive impairment (MCI) diagnoses [e.g., WMS-R Logical Memory II (Wechsler, 1987) story A score < 9, Alzheimer's Disease Neuroimaging Initiative (Petersen et al., 2010)]; or (3) an abnormal informant report indicating subjective cognitive or functional decline. The algorithm is set to over-identify cases for consensus review, with approximately half of those flagged for review being coded as cognitively normal by the consensus review panel. Cases not flagged for consensus review are classified as cognitively normal.

Cognitive status for each participant's most recent visit is determined by the consensus panel based on review of cognitive performance at all waves and review of additional information in the participant's chart (e.g., summary report of neurological and physical exam conducted by MD or NP; self-reported medical history, depressive symptoms, self-reported memory functioning, social history; and informant reports of cognitive and functional status; detailed list in Supplementary Table S1).

Consensus conference review has also been applied retrospectively to all baseline visits to confirm that the sample was free of clinical impairment at baseline. Possible cognitive status outcomes include: (1) cognitively normal, (2) early MCI (with delineation of subtypes), (3) clinical MCI (with delineation of subtypes), (4) impairment-not MCI, or (5) dementia. The diagnosis of clinical MCI is based on NIA-AA criteria (Albert et al., 2011) and includes (a) concern regarding change in cognition, (b) impairment in one or more cognitive domains, (c) preservation of functional abilities, and (d) does not meet criteria for dementia. Because the cohort was enrolled in late middle-age (mean age ~54 years), few participants have progressed to diagnoses of dementia or clinical MCI.

Therefore, the additional outcome of early MCI was developed to identify individuals in the cohort who exhibit lower than expected objective performance in one or more cognitive domains (at least 1.5 *SD*s below expected relative to internal robust norms), but may not yet report subjective cognitive complaints. This experimental construct is thought to represent a phenotype of early cognitive decline expected to precede a clinical diagnosis of MCI (Aisen et al., 2010; Duara et al., 2011; Jessen et al., 2014) and could signal an opportunity to identify potential mid-life causes of cognitive decline. Based on prior literature, we hypothesize that, in the WRAP sample, participants classified as early MCI are at increased risk of progressing to MCI and participants with MCI are at increased risk of progressing to dementia (Bondi et al., 2014).

Consensus conference diagnoses given at Wave 4 included cognitively normal ($n = 563$; 82.3%), early MCI ($n = 109$; 15.9%) clinical MCI ($n = 11$; 1.6%) and dementia ($n = 1$; 0.2%). Due to the small size of the clinical groups, early MCI and clinical statuses were combined to form a binary cognitive status outcome (0 = cognitively normal; 1 = impaired).

Statistical Analyses

Sample descriptives

Sample characteristics of the cognitively normal and impaired groups were compared using *t* tests for continuous data and chi-square tests for categorical data. For descriptive purposes, we also examined baseline and change from baseline to Wave 4 values on several neuropsychological

measures associated with development of AD. Change scores of the paired differences between Waves 1 and 4 were calculated such that negative numbers indicated worse performance at Wave 4 than at initial assessment. The cognitively normal and impaired groups were compared using general linear models after adjusting for covariates; covariates included any general sample characteristics that differed between cognitive status groups; change score covariates also included years of follow-up and corresponding baseline performance on the variable for which change was being examined.

Adjusted means (*SE*) and Cohen's f effect size estimates were reported; Cohen's f values of .10, .25, and .40 are considered small, medium, and large effect sizes, respectively (Cohen, 1992). Relationships among the IICV variables and their contributing neuropsychological factors or test scores were examined using Pearson correlations before running primary analyses to assess for collinearity in predictor variables.

Primary aims

To address our primary aims, we compared logistic regression model fit statistics for the primary outcome, cognitive status at Wave 4, across four models for each of the two IICV variables: Model 1 included years of follow-up between baseline and Wave 4 plus a set of variables that are typically associated with increased risk of MCI or AD [gender, literacy (estimated as WRAT-III reading score), family history of AD, *APOE* $\varepsilon 4$ carrier status, baseline age]—the Model 1 variables constituted the covariates for subsequent models. Model 2 included the covariates plus the linear and quadratic IICV terms (if the quadratic term was not significant (i.e., $p > .05$), it was dropped from the model). The quadratic term was included to test for a non-linear relationship between baseline IICV and subsequent cognitive status. Model 3 included covariates plus the memory and executive function scores that comprised the corresponding IICV term (non-significant memory and executive function variables were removed sequentially to make the most parsimonious model using these variables); and Model 4 included covariates plus both the IICV variable and corresponding memory and executive function terms from Model 3.

For each IICV method, nested models (Model 2 *vs*. 4, Model 3 *vs*. 4) were compared using the Likelihood Ratio Test (Lewis, Butler, & Gilbert, 2011). The Akaike Information Criteria (AICs) of Models 2–4 were compared across the 6 Factor IICV and 4 Test IICV variables to characterize relative model fits of the two IICV approaches; lower AIC values indicate better fit.

Secondary analyses for primary aims

In secondary analyses, we examined the consistency of the results for Models 2 and 4 after removing four participants identified as being highly influential during model diagnostic analyses and after excluding the 12 participants with clinical diagnoses at Wave 4. In supplementary analyses, we also reran the primary analyses (a) using baseline cognitive status as the primary outcome, and (b) using Wave 4 cognitive status as the outcome after excluding the 57 participants who met criteria for early MCI at baseline.

Exploratory aim

This analysis explored whether replacing continuous test and IICV scores with a small set of indicator variables created from distribution-based cutoffs from the WRAP baseline sample could generate an MCI risk group variable to predict risk of subsequent impairment.

MCI risk group

We created an MCI Risk Group variable by combining risk statuses across AVLT, Trails B, and IICV patterns among these scores and the WRAT-III reading score (Trails A was omitted from this analysis since it was not a significant predictor in Model 3). High-risk AVLT and Trails B scores were defined using the value that indicated the lowest performing tertile in the WRAP baseline sample: baseline AVLT at or below 47 and Trails B greater than or equal to 68 s. High-risk IICV was defined by: (1) identifying the inter-quartile range (IQR) for AVLT (45 to 56), Trails B (75 to 47), and WRAT-III standard score (98 to 112); (2) coding whether each score was below, within (inclusive) or above the IQR for each test; and (3) identifying low- *versus* high-risk IICV based on the directionally-informed pattern across the three variables. Low-risk IICV was defined as: (a) AVLT, Trails B, and WRAT-III scores all within or above the IQR; in addition, when WRAT-III (i.e., global ability) was in the highest quartile, either AVLT or Trails B also had to be in the highest quartile), or (b) only WRAT-III was in the lowest quartile (e.g., high IICV *Z*-scores due to strengths in memory and executive function relative to literacy were considered low IICV risk in these analyses). All other patterns were coded as high-risk IICV.

The statuses across the AVLT, Trails B, and IICV risk indicator variables were then combined into the following eight MCI Risk groups: 0 = none of the scores in a risk zone ($n = 251$); 1 = only IICV in the risk zone ($n = 127$); 2 = only Trails B in the risk zone ($n = 17$); 3 = Trails B and IICV in risk zones ($n = 112$); 4 = only AVLT in risk zone ($n = 12$); 5 = AVLT and IICV in risk zones ($n = 88$); 6 = AVLT and Trails B in risk zones ($n = 0$); 7 = AVLT, Trails B, and IICV in risk zones ($n = 77$). This variable replaced IICV in Model 2 and we estimated subsequent risk of impairment using the MCI Risk group = 0 as the reference group. Supplementary Figure S2 depicts how these criteria could be operationalized for use in clinic settings. We also repeated the exploratory analysis after excluding those with early MCI at baseline or clinical MCI or dementia at Wave 4.

All analyses were performed in SAS 9.3 and tests of significance were set at $p < .05$ unless otherwise noted.

RESULTS

Sample Descriptives

Approximately 18% of the 684 (n = 121) participants met criteria for the "impaired" Wave 4 cognitive status group. The impaired group was older, had more men, lower full scale IQ at baseline and reported greater depressive symptomatology at baseline than the Cognitively Normal group. The impaired group also showed higher mean IICV for both 6 Factor and 4 Test IICV (Table 1).

After adjusting for baseline age, gender, and depression scores, the two cognitive status groups also differed on numerous baseline neuropsychological test scores (Table 2). While both groups showed average scores within published normal cognitive ranges, the impaired group consistently performed lower at baseline, with the greatest deficits in memory and executive function domains. Despite starting lower, the impaired group also demonstrated greater declines on all cognitive variables except for Digit-Span forward and backward. The largest effect sizes for change were again observed for memory and executive function scores (Table 2), providing evidence supporting the consensus diagnoses of pre-clinical or clinical decline in the impaired cognitive status group.

The correlogram (Figure 1) depicts the correlations between 6 Factor IICV, 4 Test IICV, baseline age, literacy, and the memory and executive function variables contributing to each IICV variable. There is a moderate correlation (.48) between 6 Factor IICV and 4 Test IICV. Other correlations between 6 Factor IICV are small, with weak significant associations with Verbal Learning and Memory (VLM), Speed and Flexibility, and Trails A. Correlations between 4 Test IICV and other variables showed significant correlations only with Trails A and B (moderate magnitude). Not surprisingly, factor scores also correlate significantly (moderate to strong) with the neuropsychological tests that contribute to them (e.g., VLM and AVLT).

Primary Aims

Results of the logistic regressions for Models 1–4 are shown in Tables 3 (Model fit statistics) and 4 (Parameter estimates) for both IICV variables and corresponding cognitive scores.

6 Factor IICV model comparisons

Models 2–4 corresponding to the 6 Factor IICV set showed better fit than Model 1 (covariates only). In Model 2, higher baseline 6 Factor IICV *Z*-scores were associated with higher odds of being in the cognitively impaired group at Wave 4. In Model 3, the Immediate Memory factor *Z*-score was not a significant predictor of impaired status, and it was removed from the model before obtaining the values shown in Tables 3 and 4. Each of the other three factors contributed significantly to predicting risk of cognitive impairment, with better baseline performance associated with lower risk for all three factors. In Model 4, while all three factor scores from Model 3 were significant predictors of cognitive impairment following the same patterns seen in Model 3, the effects of IICV were attenuated (p = .11). The Likelihood Ratio Test (LRT) comparing Models 2 and 4 showed that Model 4 yielded a significantly better fit than Model 2, while the LRT comparing Models 3 and 4 showed that Model 4 did not improve the fit over Model 3.

4 Test IICV model comparisons

Results for 4 Test IICV showed similar patterns to 6 Factor IICV. Models 2–4 for 4 Test IICV showed better fits than Model 1. In Model 2, higher baseline 4 Test IICV was associated with higher odds of cognitive impairment at Wave 4. In Model 3, AVLT Total and Trails B *Z*-scores contributed significantly to predicting cognitive impairment, with better performance associated with lower risk. In Model 4, the two baseline neuropsychological test scores from Model 3 remained significant predictors of subsequent cognitive

Table 1. Sample characteristics: Cognitively normal vs impaired at Wave 4

	Cognitively normal	Cognitively impaired	*p*-Value*
Sample characteristics	(n = 563)	(n = 121)	
Age at baseline, mean (*SD*) years	53.1 (6.6)	55.4 (6.4)	.0005
Follow-up, mean (*SD*) years	9.1 (1.0)	9.2 (1.0)	.45
Female, n (%)	406 (72.1)	70 (57.9)	.002
Education >=BA, n (%)	361 (64.1)	75 (62.0)	.40
non-Hispanic Caucasian, n (%)	558 (99.1)	121 (100)	.30
Apoe e4 carrier, n (%)	218 (38.7)	43 (35.5)	.51
Family history of AD, n (%)	427 (75.8)	83 (68.6)	.10
CESD total score, mean (*SD*)	5.6 (6.1)	7.2 (7.6)	.026
FSIQ, mean (*SD*)	114.6 (8.7)	112.0 (9.0)	.003
WRAT standard reading score, mean (*SD*)	106.3 (9.1)	106.6 (9.3)	.75
Intraindividual variability measures			
Baseline 6 Factor IICV, mean (*SD*)	−.052 (1.0)	.28 (1.1)	.0008
Baseline 4 Test IICV, mean (*SD*)	−.30 (.70)	−.052 (.75)	.0005

*p-Values are from t-tests or chi-square tests for the sample characteristics and baseline IICV, depending on whether data are continuous or categorical.

Table 2. Cognition at baseline and change to Wave 4 by Wave 4 cognitive status group

	Wave 4 cognitive status			
	Cognitively normal	Cognitively impaired	*p*-Value**	Cohen's f
Baseline[a] neuropsychological performance	(*n* = 563)	(*n* = 121)		
AVLT Total of 5 Learning Trials, lsmean (*SE*)	53.3 (.29)	46.5 (.63)	<.0001	0.38
AVLT Delayed Recall, lsmean (*SE*)	11.2 (.11)	8.9 (.23)	<.0001	0.34
Trails A seconds to complete, lsmean (*SD*)	24.58 (1.02)	29.24 (1.03)	<.0001	0.24
Trails B seconds to complete, lsmean (*SD*)	57.12 (1.02)	68.01 (1.03)	<.0001	0.26
Digit Span Forward, lsmean (*SE*)	10.7 (.09)	10.1 (.19)	0.004	0.11
Digit Span Backward, lsmean (*SE*)	7.4 (.09)	6.4 (.21)	<.0001	0.16
Letter Number Sequencing, lsmean (*SE*)	11.1 (.10)	10.1 (.22)	<.0001	0.15
COWAT CFL, lsmean (*SE*)	44.6 (.47)	39.4 (1.0)	<.0001	0.17
Logical Memory I, lsmean (*SE*)	30.7 (.25)	26.2 (.54)	<.0001	0.28
Logical Memory II, lsmean (*SE*)	27.6 (.27)	22.3 (.60)	<.0001	0.31
BVMT Immediate, lsmean (*SE*)	25.3 (.21)	21.4 (.46)	<.0001	0.29
BVMT Delayed, lsmean (*SE*)	9.8 (.07)	8.8 (.16)	<.0001	0.23
Digit-Symbol, lsmean (*SE*)	58.8 (.38)	54.5 (.84)	<.0001	0.18
Change to Wave 4 (negative indicates worse)				
AVLT Total of 5 Learning Trials, lsmean (*SE*)	.30 (.25)	−5.24 (.57)	<.0001	0.33
AVLT Delayed Recall, lsmean (*SE*)	.32 (.09)	−1.39 (.20)	<.0001	0.29
Trails A seconds to complete, lsmean (*SE*)	1.45 (.26)	−1.07 (.58)	0.0001	0.15
Trails B seconds to complete, lsmean (*SE*)	2.07 (.84)	−14.7 (1.86)	<.0001	0.31
Digit Span Forward, lsmean (*SE*)	.047 (.07)	−.20 (.16)	0.15	0.06
Digit Span Backward, lsmean (*SE*)	.13 (.07)	−.15 (.16)	0.10	0.06
Letter Number Sequencing, lsmean (*SE*)	.06 (.08)	−.69 (.17)	<.0001	0.15
COWAT CFL, lsmean (*SE*)	3.43 (.37)	.41 (.81)	0.0009	0.13
Logical Memory I, lsmean (*SE*)	.005 (.19)	−3.5 (.43)	<.0001	0.29
Logical Memory II, lsmean (*SE*)	.73 (.21)	−3.2 (.48)	<.0001	0.28
BVMT Immediate, lsmean (*SE*)	1.34 (.18)	−1.48 (.4)	<.0001	0.24
BVMT Delayed, lsmean (*SE*)	.46 (.06)	−.57 (.14)	<.0001	0.25
Digit-Symbol, lsmean (*SE*)	−2.20 (.22)	−4.10 (.48)	0.0004	0.14

[a]The first eight neuropsychological tests (AVLT - CFL) are part of the test battery for every wave of assessment. The last five tests (Logical Memory - Digit Symbol) were added to the test battery at Waves 2 and on.
**P-values are from generalized linear models, adjusting for significant sample characteristics of age, gender, and CES-D depression score; follow-up years included as a covariate for analysis of changes Baseline to Wave 4. Trails A and B were transformed using Log 10; values reported were backtransformed from the lsmeans estimates for the baseline scores. All difference scores from raw data were distributed approximately normally. Cohen's f effect sizes are reported to facilitate comparison of effect sizes across variables.

status while 4 Test IICV was no longer significant. The LRTs comparing Models 2 and 4 and Models 3 and 4 showed that Model 4 yielded a significantly better fit than Model 2, but not Model 3.

Comparison of the two IICV approaches

Comparing results of the models that included covariates + IICV as predictors (Model 2) for the 6 Factor IICV and 4 Test IICV indicated that the two IICV variables yielded nearly identical fit statistics (AIC difference <0.5). A comparison of fit statistics from Model 3 across the two approaches indicated that inclusion of the factor scores that comprised the 6 Factor IICV exhibited a better fit (AIC 10 points lower) than the model including the neuropsychological test scores that comprised the 4 Test IICV. The same pattern emerged when Model 4 results were compared across the two IICV methods. While the AICs were lower in Models 3 and 4 for the 6 Factor IICV, the AUC and model R-squared values differed only slightly between 6 Factor IICV and 4 Test IICV (Table 3), indicating that a relatively small set of neuropsychological test scores may be sufficient for screening for risk of subsequent cognitive impairment.

Secondary analyses

In secondary analyses, the removal of four participants identified as influential *via* diagnostic modeling did not result in a change in significance relative to $\alpha = .05$ for Models 2 and 4; patterns remained similar when the 12 participants with a clinical diagnosis at Wave 4 were excluded. Results incorporating the baseline consensus conference statuses (normal *vs.* early MCI) are presented in the Supplementary Materials.

Exploratory Aim

In exploratory analyses after adjusting for Model 1 covariates, baseline MCI Risk Group was a significant predictor of

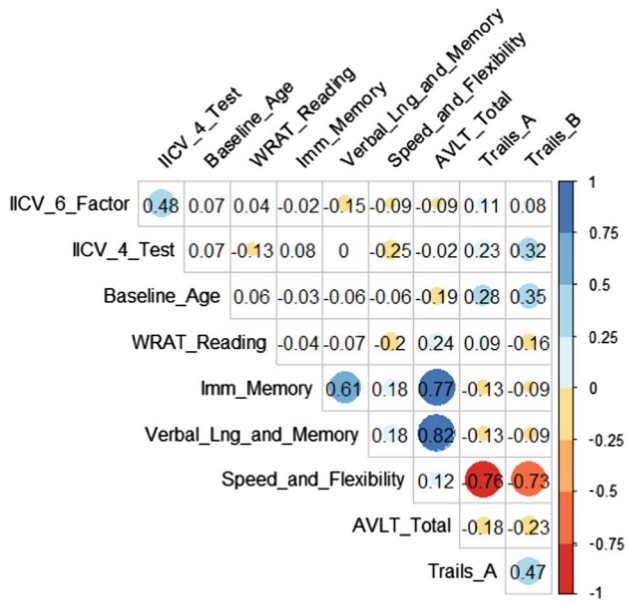

Fig. 1. Correlogram of 6 Factor IICV and 4 Test IICV with variables that contribute to each test and with baseline age. Trails A and B were log-transformed. Correlations with $p < .05$ also have a colored dot overlapping with the correlation coefficient to provide an easy means to visualize the magnitude and direction of correlations: larger dots indicate larger coefficients; blue shades indicate positive correlations while yellow to red shades indicate negative correlations.

subsequent cognitive status (MCI Risk Group Wald chi-square 82.5, df = 7; $p < .0001$; model AIC = 539.28 and −2 Log L = 513.28). Figure 2 depicts the proportion (± *SE*) of participants who were impaired within each of the MCI Risk groups. The odds ratios (and 95% Wald confidence intervals) comparing MCI Risk = 1–7 with MCI Risk = 0 were as follows: IICV risk (MCI Risk = 1; $n = 127$), 2.86 (1.25–6.50); Trails B risk (MCI Risk = 2; $n = 17$), 1.25 (0.15–10.44); Trails B and IICV risk (MCI Risk = 3; $n = 112$), 4.45 (1.99–9.96); AVLT risk (MCI Risk = 4; $n = 12$), 12.29 (3.13, 48.30); AVLT and IICV risk (MCI Risk = 5; $n = 88$), 12.10 (5.49–26.68); Trails B and AVLT risk (MCI Risk, $n = 0$, OR not available); All 3 risk (MCI Risk = 7; $n = 77$), 39.49 (13.01–66.83). All odds ratios were significant ($p < .05$) after adjusting for Model 1 covariates, except for MCI Risk = 2 (only Trails B in the risk range) and MCI Risk = 6 (AVLT and Trails B risk without IICV risk).

The smallest increase in risk of impairment was associated with only having an IICV pattern in the high risk zone while the largest increase in risk was associated with AVLT, Trails B, and IICV pattern all falling in the high risk zone. Rerunning the analyses after excluding those with early MCI at baseline or those with a clinical status at Wave 4 yielded consistent results (Supplementary Figures S3 and S4).

DISCUSSION

The aims of this study were to investigate whether IICV obtained at baseline in an ostensibly healthy late-middle-aged sample was predictive of meeting criteria for mild or more severe impairment approximately 8–10 years later and whether IICV effects persisted after accounting for performance on measures of memory and executive function which contributed to the IICV calculations.

In our sample with mean baseline age of 54, baseline IICV was predictive of subsequent cognitive impairment for both IICV calculations (i.e., *via* a complex method using factor scores adjusted for age, gender, and literacy level and *via* a simpler method based on the Rey AVLT sum of learning trials, Trails A and B, and the WRAT-III reading). When considered only with covariates, the model fits of the

Table 3. Model fit statistics for logistic regression Models 1–4 for 6 Factor IICV and 4 Test IICV

Model fit statistics	Model 1 (Covariates[a] only)	Model 2 (Covariates + IICV)	Model 3 (Covariates + Cognitive Scores)	Model 4 (Covariates + IICV + Cognitive Scores)
6 Factor IICV set of models				
AIC	627.43	620.06	498.70	498.09
−2LogL	613.43	604.06	478.70	476.09
ROC-AUC	0.64	0.67	0.83	0.83
Max-rescaled R^2	0.059	0.081	0.343	0.342
4 Test IICV set of models				
AIC	627.43	620.40	508.90	510.64
−2LogL	613.43	604.40	490.90	490.64
ROC-AUC	0.64	0.66	0.83	0.83
Max-rescaled R^2	0.059	0.080	0.320	0.320

Note. Variables are Z-scores all with positive values indicating better performance.
[a]Covariates included in Model 1: gender, literacy (estimated as WRAT-3 reading score), family history of AD, APOE ε4 carrier, baseline age, and years of follow-up at the time the cognitive status was determined.

Table 4. Model parameter estimates for Models 2–4 for 6 Factor IICV and 4 Test IICV

	Model 2		Model 3		Model 4	
	(Covariates[a] + IICV)		(Covariates[a] + Cognitive Scores)		(Covariates[a] + IICV + Cognitive Scores)	
Predicting increased risk of cognitive impairment[b]	Parameter estimate (*p*-value)	Odds ratio (Wald 95% conf limits)	Parameter estimate (*p*-value)	Odds ratio (Wald 95% conf limits)	Parameter estimate (p-value)	Odds ratio (Wald 95% conf limits)
6 Factor IICV and Factor Scores						
6 Factor IICV	.307 (.0022)	1.36 (1.12, 1.65)	—	—	.19 (.11)	1.21 (.96, 1.52)
Immediate Memory	—	—	NS	—	—	—
Verbal Learning and Memory	—	—	−.84 (<.0001)	.43 (.35, .53)	−.81 (<.0001)	.44 (.36, .55)
Speed and Flexibility	—	—	−.59 (<.0001)	.55 (.44, .70)	−.58 (<.0001)	.56 (.44, .71)
Working Memory	—	—	−.32 (.007)	.72 (.57, .91)	−.36 (.002)	.70 (.55, .88)
4 Test IICV and Test Scores						
4 Test IICV	.411 (.0025)	1.51 (1.16, 1.97)	—	—	.089 (.61)	1.09 (.78, 1.53)
AVLT Total	—	—	−1.31 (<.0001)	.27 (.20, .37)	−1.30 (<.0001)	.27 (.20, .37)
Log 10 Trails A	—	—	NS	—	—	—
Log 10 Trails B	—	—	−.66 (<.0001)	.52 (.40, .67)	−.63 (<.0001)	.53 (.41, .70)

[a]Covariates included in Model 1: gender, literacy (estimated as WRAT-3 reading score), family history of AD, APOE ε4 carrier, baseline age, and years of follow-up at the time the cognitive status was determined
[b]Variables are *Z*-scores all with positive values indicating better performance.

complex and simple IICV methods (Model 2 AICs) were nearly identical, suggesting that the intraindividual variability in a small number of neuropsychological test scores could be useful in predicting long-term risk of mild to moderate cognitive impairment.

When the continuous baseline memory and executive function scores were included in the corresponding models (Model 4), the effects of IICV were attenuated to a non-significant level, suggesting that the calculation of continuous IICV may not provide any additional information regarding subsequent risk above and beyond the individual test scores in a late middle-age cognitively normal sample. These findings contrast with those observed by Holtzer et al. (2008). In their study, IICV was calculated using tests from the same domains used for our 4 Test IICV in an older sample (mean (*SD*) age = 78.6 (5.3), *n* = 897). After controlling for the neuropsychological test scores comprising their IICV variable, higher IICV was associated with significantly higher risk of subsequent dementia over a shorter follow-up window (mean = 3.3 (2.4) years). While our results suggest that IICV may not be as sensitive to later decline as other studies have shown, it may also be that IICV functions differently in middle-age when cognitive declines tend to be minimal; specifically, the younger age of our sample and the less severe status of our impaired group are possible explanations for the attenuated effects of IICV in our analyses.

Using the tests comprising 4-Test IICV, our exploratory analyses suggest that using distribution-based cutoffs to identify higher risk AVLT, Trails B, *and* directionally informed IICV patterns could provide a simple method to identify people at increased risk of decline to a cognitively impaired status (Supplementary Figure S2). This approach has practical appeal and is potentially easier to use, less costly, and more directly linked to clinical cognitive outcomes than those that rely on demographics, vascular risk factors, and other health status variables (see Imtiaz et al., 2014) for variables used in risk estimators). Future research could test these or other cutoffs and incorporate additional neuropsychological tests or IICV patterns to determine whether there are operationalizations of test score cutoffs and IICV patterns that are optimally sensitive to risk of early decline.

The quest to identify metrics that are sensitive to risk of cognitive decline is not new. For example, Andersson and colleagues (Andersson et al., 2006) established baseline risk groups using the AVLT sum of learning trials and delayed recall scores (Trial 7) for a sample of 224 late-middle-age individuals (mean age = 60.7 years): participants with poor delayed recall scores (<6) fell in the "Severe Impairment" group; those with delayed scores at or above 6 and sum of learning trials scores at or below 45 were in the "Moderate Impairment" group; and those with better scores on both components were in the "No Impairment" group. They found that these cutoffs were sensitive to identifying individuals at increased risk of incident dementia approximately 3 years later. Similarly, Drebing and colleagues (Drebing, Van Gorp, Stuck, Mitrushina, & Beck, 1994) published recommended

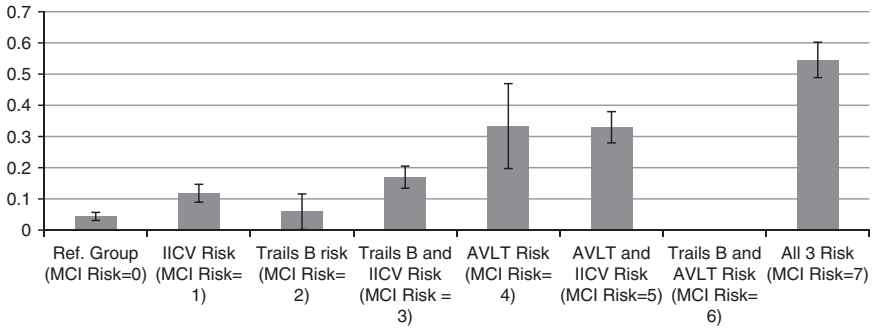

Fig. 2. Bar graphs of proportion in the cognitively impaired group at Wave 4 within each of the subgroups representing risk based on falling in the "risk tertile" for AVLT or Trails B or having an IICV risk pattern based on the IQRs of AVLT, Trails B, and WRAT-III reading. Denominators for each of the eight groups below are 251, 127, 17, 112, 12, 88, 0, and 77, respectively.

cutoffs for a screening battery that included tests of memory and executive function. Their analyses focused on a narrow age range (age 60–69 years) and normative values for more highly educated people and showed that distribution-based cutoffs could be used to distinguish individuals who were cognitively impaired *versus* normal. To our knowledge, ours is the first study to attempt to incorporate directionally informed patterns of intraindividual variability into risk estimates.

LIMITATIONS

There are multiple methods of calculating IICV; our analyses focused on the *SD* based on tests from multiple cognitive domains. The vast majority of our impaired group met criteria for an experimental construct, early MCI. We anticipate that not all of the participants with early MCI will progress to clinical MCI and dementia, thus potentially limiting the sensitivity of IICV in these analyses. The risk estimating algorithm presented in this study is presented as a "proof of concept" approach to incorporating consideration of IICV into clinical practice; as such, the cutoffs selected in the exploratory aim and even the tests selected might not represent the most sensitive selections possible. Since these analyses are based on a sample that is risk-enriched (74.6% with a parental family history of AD), caution should be used in generalizing the findings.

CONCLUSIONS

IICV, calculated as *SD* of a set of measures at a given time point, has been suggested as a novel marker that can be used to identify people at risk of later dementia-related diagnoses. It has great appeal as a potential low-cost, non-invasive risk biomarker. Our results provide mixed evidence for the potential utility of IICV during middle-age. Namely, continuous IICV does indeed predict meeting criteria for cognitive impairment approximately 9 years later, but the contribution of IICV is attenuated after accounting for the memory and executive function scores that comprise IICV. On the other hand, our exploratory, "proof of concept" analysis suggests that test scores might be used to identify IICV patterns associated with higher risk of impairment.

Future research could also characterize longitudinal trajectories of IICV in this younger age range and evaluate whether longitudinal changes in IICV might indicate risk of cognitive decline to AD or other dementia endpoints. For example, prior research showed that older adults exhibit increasing variability in cognitive performance over time, while younger adults present with stable or slightly decreasing metrics of cognitive variability (MacDonald, Hultsch, & Dixon, 2003). Future research on WRAP participants, especially those who demonstrate increased IICV at an earlier age than traditionally expected, could help to further delineate the role of cognitive variability in both pathological and normal aging.

ACKNOWLEDGMENTS

We thank the WRAP participants and WAI staff for their contributions to the WRAP study. WRAP is supported by NIA grant R01AG27161 (SCJ; Wisconsin Registry for Alzheimer Prevention: Biomarkers of Preclinical AD), Helen Bader Foundation, Northwestern Mutual Foundation, Extendicare Foundation and State of Wisconsin. WRAP is also supported by the Clinical and Translational Science Award (CTSA) program, through the NIH National Center for Advancing Translational Sciences (NCATS), grant UL1TR000427. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH. The authors state that there are no conflicts of interest. This research is also supported by the Holland Wisconsin Alzheimer's Institute Research Fund (L.R.C.) and NIA grant F30AG054115 (S.E.B.), the Medical Scientist Training Program T32 (S.E.B., A.H., T32 GM008692), the Neuroscience Training Program T32 (S.E.B., M.H., T32 GM007507) and the Rath Distinguished Graduate Research Foundation Fellowship (S.E.B.).

Supplementary material

For supplementary material/s referred to in this article, please visit https://doi.org/10.1017/S135561771600093X

REFERENCES

Aisen, P.S., Petersen, R.C., Donohue, M.C., Gamst, A., Raman, R., Thomas, R.G., … Weiner, M.W. (2010). Clinical core of the Alzheimer's disease neuroimaging initiative: Progress and plans. *Alzheimer's & Dementia*, *6*(3), 239–246. http://doi.org/10.1016/j.jalz.2010.03.006.

Albert, M.S., DeKosky, S.T., Dickson, D., Dubois, B., Feldman, H.H., Fox, N.C., … Petersen, R.C. (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*, *7*(3), 270–279.

Andersson, C., Lindau, M., Almkvist, O., Engfeldt, P., Johansson, S.-E., & Eriksdotter Jönhagen, M. (2006). Identifying patients at high and low risk of cognitive decline using Rey Auditory Verbal Learning Test among middle-aged memory clinic outpatients. *Dementia and Geriatric Cognitive Disorders*, *21*(4), 251–259.

Bondi, M.W., Edmonds, E.C., Jak, A.J., Clark, L.R., Delano-Wood, L., McDonald, C.R., … Galasko, D. (2014). Neuropsychological criteria for mild cognitive impairment improves diagnostic precision, biomarker associations, and progression rates. *Journal of Alzheimer's Disease*, *42*(1), 275–289.

Clark, L.R., Koscik, R.L., Nicholas, C.R., Okonkwo, O.C., Engelman, C.D., Bratzke, L.C., … Johnson, S.C. (2016). Mild cognitive impairment in late middle age in the Wisconsin Registry for Alzheimer's Prevention (WRAP) study: Prevalence and characteristics using robust and standard neuropsychological normative data. [Epub ahead of print].

Cohen, J. (1992). A power primer. *Psychological Bulletin*, *112*(1), 155.

Dowling, N.M., Hermann, B., La Rue, A., & Sager, M.A. (2010). Latent structure and factorial invariance of a neuropsychological test battery for the study of preclinical Alzheimer's disease. *Neuropsychology*, *24*(6), 742–756. http://doi.org/10.1037/a0020176.

Drebing, C.E., Van Gorp, W.G., Stuck, A.E., Mitrushina, M., & Beck, J. (1994). Early detection of cognitive decline in higher cognitively functioning older adults: Sensitivity and specificity of a neuropsychological screening battery. *Neuropsychology*, *8*(1), 31.

Duara, R., Loewenstein, D.A., Greig, M.T., Potter, E., Barker, W., Raj, A., … Potter, H. (2011). Pre-MCI and MCI: Neuropsychological, clinical, and imaging features and progression rates. *The American Journal of Geriatric Psychiatry*, *19*(11), 951–960. http://doi.org/10.1097/JGP.0b013e3182107c69

Grice, J.W. (2001). Computing and evaluating factor scores. *Psychological Methods*, *6*(4), 430.

Heaton, R., Miller, S., Taylor, J.R., & Grant, I. (2004). *Comprehensive norms for an expanded Halstead-Reitan battery: Demographically adjusted neuropsychological norms for African American and Caucasian adults (HRB). Professional Manual*. Lutz, FL: Psychological Assessment Resources Inc..

Hilborn, J.V., Strauss, E., Hultsch, D.F., & Hunter, M.A. (2009). Intraindividual variability across cognitive domains: Investigation of dispersion levels and performance profiles in older adults. *Journal of Clinical and Experimental Neuropsychology*, *31*(4), 412–424.

Holtzer, R., Verghese, J., Wang, C., Hall, CB., & Lipton, RB. (2008). WIthin-person across-neuropsychological test variability and incident dementia. *JAMA*, *300*(7), 823–830. http://doi.org/10.1001/jama.300.7.823.

Imtiaz, B., Tolppanen, A.-M., Kivipelto, M., & Soininen, H. (2014). Future directions in Alzheimer's disease from risk factors to prevention. *Biochemical Pharmacology*, *88*(4), 661–670.

Jessen, F., Wolfsgruber, S., Wiese, B., Bickel, H., Mösch, E., Kaduszkiewicz, H., … Wagner, M. (2014). AD dementia risk in late MCI, in early MCI, and in subjective memory impairment. *Alzheimer's & Dementia*, *10*(1), 76–83. http://doi.org/10.1016/j.jalz.2012.09.017.

Kälin, A.M., Pflüger, M., Gietl, A.F., Riese, F., Jäncke, L., Nitsch, R.M., & Hock, C. (2014). Intraindividual variability across cognitive tasks as a potential marker for prodromal Alzheimer's disease. *Frontiers in Aging Neuroscience*, *6*, 147.

Koepsell, T.D., & Monsell, S.E. (2012). Reversion from mild cognitive impairment to normal or near-normal cognition Risk factors and prognosis. *Neurology*, *79*(15), 1591–1598.

Koscik, R.L., La Rue, A., Jonaitis, E.M., Okonkwo, O.C., Johnson, S.C., Bendlin, B.B., … Sager, M.A. (2014). Emergence of mild cognitive impairment in late middle-aged adults in the Wisconsin registry for Alzheimer's prevention. *Dementia and Geriatric Cognitive Disorders*, *38*(1–2), 16–30. http://doi.org/10.1159/000355682.

Lewis, F., Butler, A., & Gilbert, L. (2011). A unified approach to model selection using the likelihood ratio test. *Methods in Ecology and Evolution*, *2*(2), 155–162. http://doi.org/10.1111/j.2041-210X.2010.00063.x

MacDonald, S.W., Hultsch, D.F., & Dixon, R.A. (2003). Performance variability is related to change in cognition: Evidence from the Victoria Longitudinal Study. *Psychology and Aging*, *18*(3), 510.

Petersen, R.C., Aisen, P.S., Beckett, L.A., Donohue, M.C., Gamst, A.C., Harvey, D.J., … Toga, A.W. (2010). Alzheimer's disease Neuroimaging Initiative (ADNI) clinical characterization. *Neurology*, *74*(3), 201–209.

Sager, M.A., Hermann, B., & La Rue, A. (2005). Middle-aged children of persons with Alzheimer's disease: APOE genotypes and cognitive function in the Wisconsin Registry for Alzheimer's Prevention. *Journal of Geriatric Psychiatry and Neurology*, *18*(4), 245–249. http://doi.org/10.1177/0891988705281882

Salthouse, T.A., & Soubelet, A. (2014). Heterogeneous ability profiles may be a unique indicator of impending cognitive decline. *Neuropsychology*, *28*(5), 812.

Schmidt, M. (1996). *Rey auditory verbal learning test: A handbook*. Los Angeles: Western Psychological Services.

Sperling, R.A., Karlawish, J., & Johnson, K.A. (2013). Preclinical Alzheimer disease —the challenges ahead. *Nature Reviews. Neurology*, *9*(1), 54–58. http://doi.org/10.1038/nrneurol.2012.241

Vaughan, L., Leng, I., Dagenbach, D., Resnick, S.M., Rapp, S.R., Jennings, J.M., … Coker, L.H. (2013). Intraindividual variability in domain-specific cognition and risk of mild cognitive impairment and dementia. *Current Gerontology and Geriatrics Research*, *2013*, 495793.

Wechsler, D. (1987). *WMS-R: Wechsler memory scale-revised*. San Antonio, TX: Psychological Corporation.

Wilkinson, G.S. (1993). *WRAT-3: Wide range achievement test*. San Antonio, TX: Pearson.