Within-Person Across-Neuropsychological Test Variability and Incident Dementia

Roee Holtzer, PhD
Joe Verghese, MD
Cuiling Wang, PhD
Charles B. Hall, PhD
Richard B. Lipton, MD

Context Neuropsychological tests are used to predict and diagnose dementia. However, to our knowledge, no studies to date have examined whether within-person across-neuropsychological test variability predicts dementia.

Objective To examine whether within-person across-neuropsychological test variability predicts future dementia.

Design The Einstein Aging Study (EAS) is a population-based longitudinal study of aging and dementia located in Bronx County, New York. We used Cox proportional hazards models using age as the time scale to estimate hazard ratios (HRs) for performance on individual neuropsychological tests (Free and Cued Selective Reminding Test, Digit Symbol Substitution subtest of the Wechsler Adult Intelligence Scale Revised, and the Vocabulary subtest of the Wechsler Adult Intelligence Scale Revised) and for within-person across-neuropsychological test variability as predictors of incident dementia. Analyses were stratified by sex, and controlled for education and medical illness.

Setting and Participants A total of 1797 participants (age ≥70 years) enrolled in the EAS between October 1993 and December 2007. Participants seen for the baseline visit only (n=750), prevalent dementia cases (n=72), and those with missing follow-up information (n=78) were excluded. A total of 897 individuals were included in this investigation. Participants had follow-up visits every 12 to 18 months.

Main Outcome Measure Incident dementia.

Results Sixty-one cases of incident dementia were identified during follow-up (mean [SD], 3.3 [2.4] years), of which 26 were in the highest quartile of within-person across-neuropsychological test variability. Adjusting for sex, education, and medical illness, variability was associated with incident dementia (HR for 1-point difference in variability, 3.93 [95% confidence interval {CI}, 2.04-7.56]). The association persisted even after adjusting for level of performance on individual neuropsychological tests (HR for 1-point difference in variability, 2.10 [95% CI, 1.04-4.23]). Comparing Cox models using neuropsychological tests with and without within-person across-neuropsychological test variability showed that the former improved the prediction of dementia. Sensitivity in a model predicting dementia at 1 year also improved when neuropsychological test variability was included.

Conclusions In this population, within-person across-neuropsychological test variability was associated with development of incident dementia independent of neuropsychological test performance. This finding needs to be confirmed in future studies.

JAMA. 2008;300(7):823-830

www.jama.com
variability in performance across neuropsychological tests administered in a single session (such variability is expected in the normal population, including older adults without dementia, reflecting the individual’s relative cognitive strengths and weaknesses). This study focused on within-person across-neuropsychological test variability because it can be estimated directly from standardized clinical neuropsychological procedures.

Increased within-person across-neuropsychological test variability has been related to poor cognitive function in normal older adults in cross-sectional studies. Accordingly, we examined whether increased within-person across-neuropsychological test variability was associated with future dementia after controlling for level of performance on individual neuropsychological tests. We also assessed whether variability improved the sensitivity and specificity of a model predicting development of dementia 1 year later.

METHODS

Study Population
Participants were enrolled in the Einstein Aging Study (EAS), a longitudinal study of aging and dementia located at the Albert Einstein College of Medicine in Bronx County, New York. The study design, recruitment, and follow-up methods have been previously described. Briefly, the EAS used telephone-based screening procedures to recruit and follow-up a community-based cohort since 1993. The primary aim of the EAS was to identify risk factors for dementia. Eligibility criteria required that participants be aged 70 years or older, reside in the Bronx, and speak English. Exclusion criteria included severe audiovisual disturbances that would interfere with completion of neuropsychological tests, inability to ambulate even with a walking aid or in a wheelchair, and institutionalization. Potential participants older than 70 years from the Centers for Medicare & Medicaid Services population lists of Medi-
mixed were considered in the exploratory analyses.

**Neuropsychological Assessment**
The tests included in this battery have been validated for use in the aging population in our and other aging studies. Our recent studies reveal that factor analysis of the neuropsychological test battery consistently yielded 3 empirically derived and statistically orthogonal cognitive domains including verbal IQ, attention/executive function, and memory. For the purpose of this study, we identified 3 tests that represented the 3 cognitive domains mentioned above, and were available for all 897 participants included in this study. The Vocabulary (total score) subtest of the Wechsler Adult Intelligence Scale Revised (WAIS-R) considered a hold test in that it is not sensitive to the effect of aging and age-related diseases, represented the verbal IQ domain. The Digit Symbol Substitution subtest of the WAIS-R is commonly used to assess attention and executive function. The Free and Cued Selective Reminding Test (FCSRT; free recall) is commonly used to assess verbal memory and is sensitive to dementia.

**Estimate of Within-Person Variability Across Neuropsychological Tests**
The method used herein to estimate within-person across-test variability was described in other studies as well. Assumptions of normality were met for the distribution of scores on the FCSRT, WAIS-R Vocabulary, and WAIS-R Digit Symbol Substitution tests. The raw scores of each test were transformed on the basis of the distribution of the entire sample (n = 897). Then, the transformed test scores were used to calculate within-person variability across the 3 tests using this equation. Variability = \[ \sqrt{\frac{(Z_{k} - A_{k})^{2}}{(K - 1)}} \]

where \(Z_{k}\) is the kth cognitive test score for ith individual. Here, \(k = 1, \ldots, K\), \(K = 3\) (FCSRT, WAIS-R Vocabulary, and WAIS-R Digit Symbol Substitution) and

\[ A_{k} = \sum_{i=1}^{K} Z_{i} \]

is the individual’s mean z-transformed score based on the 3 tests.

**Illness Index**
Consistent with our previous studies, dichotomous rating (present or absent) of diabetes, chronic heart failure, arthritis, hypertension, depression, stroke, Parkinson disease, chronic obstructive pulmonary disease, angina, and myocardial infarction was used to calculate an illness index summary score (range, 0-10). Medical history was obtained from multiple sources including significant others and family physicians when available. The study physician obtained the medical history during the neurological examination and trained research assistants conducted the structured clinical interview. Smoking was ascertained but did not contribute to variability in incident dementia in this study and was not included in the analyses reported herein.

**Statistical Analyses**
We performed 3 Cox proportional hazards regression analyses to assess the relationship between neuropsychological tests and dementia. First, we examined whether level of performance on the neuropsychological tests at baseline predicted the development of dementia. Second, we examined the association between within-person across-neuropsychological test variability and the development of dementia. Third, we assessed whether within-person across-neuropsychological variability provided incremental prediction of incident dementia above and beyond what was predicted by the absolute level of performance on the neuropsychological tests by controlling for performance differences on the WAIS-R Vocabulary, FCSRT, and WAIS-R Digit Symbol Substitution. Individuals who died before developing dementia were censored at the time of their last study visit when their status as not having dementia was determined.

To assess the amount of variability of development of incident dementia explained by the models, we compared the goodness of fit of nested Cox proportional hazards models using the partial likelihood ratio test. We compared a model with the 3 individual neuropsychological tests with a model that added within-person across-neuropsychological test variability. Finally, using the McNemar test we examined whether within-person across-neuropsychological test variability significantly improved sensitivity for the prediction of new cases of dementia within 1 year.

We used age as the time scale because it is considered more appropriate than follow-up time in cohort studies. When age serves as the time scale, the hazard function can be directly interpreted as the age-specific incidence function and age is accounted for in the nonparametric term of the hazard function, providing a more flexible and effective control of age. Time to event was from age at the baseline assessment accounting for the left truncation at study inclusion to age at which diagnosis of dementia was ascertained or to final study contact for participants without dementia. In addition, all multivariate analyses were stratified by sex and controlled for education and medical illness. For all primary analyses, incident dementia was defined using the algorithmic procedures described earlier. Proportional hazards assumptions of the models were examined analytically and graphically and were met.

Finally, Cox proportional hazards regression analysis with age as the time scale was used to examine whether within-person across-neuropsychological test variability predicted incident dementia subtypes (e.g., vascular and Alzheimer) defined by consensus diagnosis as previously described. Given the low number of incident cases of Alzheimer dementia and vascular dementia, as well as...
concerns for potential diagnostic circularity, these analyses were considered exploratory. All analyses were performed using SAS statistical software version 9.1 (SAS Institute Inc, Cary, North Carolina) and S-Plus version 8.0 (Insightful Corp, Seattle, Washington). The significance level was set at .05 and all tests were 2-sided.

**RESULTS**

**Demographics**

Participants were community residents who were relatively healthy and had normal global mental functioning based on BIMC scores that were within the normal range (Table). Of the 897 participants, there were 61 cases of incident dementia (6.8%) (defined by the algorithmic procedure described earlier) identified during the follow-up period (mean [SD], 3.3 [2.4] years). On the basis of the consensus clinical diagnostic procedures, 47 participants developed incident dementia of the Alzheimer type and 18 participants developed incident vascular dementia. During the study, 128 individuals died, as expected for the age of this cohort. Of those, 18 had developed incident dementia.

**Prediction of Dementia**

All Cox proportional hazards regression models used age as the time scale and were adjusted for sex, education, and medical illness. In the first model, higher scores on the FCSRT (hazard ratio [HR] for 1-point difference on the test, 0.87 [95% confidence interval [CI], 0.84-0.91]; P < .001) and WAIS-R Digit Symbol Substitution (HR for 1-point difference on the test, 0.97 [95% CI, 0.94-1.00]; P = .04) were associated with a lower risk of developing dementia. Level of performance on the WAIS-R Vocabulary test was not associated with future dementia (HR for 1-point difference on the test, 1.01 [95% CI, 0.98-1.03]; P = .67).

The second Cox proportional hazards regression analysis also used age as the time scale and adjusted for sex, education, and medical illness. Within-person across-neuropsychological test variability was significantly associated with incident dementia defined algorithmically (HR for 1-point difference in variability, 3.93 [95% CI, 2.04-7.56]; P < .001). Risk of dementia increased as a function of greater within-person variability.

The third Cox proportional hazards regression analysis controlled for each of the neuropsychological test scores to determine the independent association of variability with incident dementia. The model used age as the time scale and adjusted for sex, education, and medical illness. Similar to the previous analysis, the results revealed that higher scores on the FCSRT (HR for 1-point difference on the test, 0.89 [95% CI, 0.85-0.92]; P < .001) and WAIS-R Digit Symbol Substitution (HR for 1-point difference on the test, 0.97 [95% CI, 0.94-1.00]; P = .04) were associated with lower risk of incident dementia, and the WAIS-R Vocabulary test was not associated with it (HR for 1-point difference on the test, 1.00 [95% CI, 0.98-1.03]; P = .81). Importantly, even after controlling for individual neuropsychological test scores, within-person across-neuropsychological test variability remained significantly associated with incident dementia (HR for 1-point difference in variability, 2.10 [95% CI, 1.04-4.23]; P = .03). The HR of variability was lower after adjusting

---

**Table.** Summary of Participant Characteristics per Quartile of Within-Person Across-Neuropsychological Test Variability at Baseline

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Entire Sample (N = 897)</th>
<th>Variability by Quartilea</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1 (n = 225)</td>
<td>2 (n = 224)</td>
</tr>
<tr>
<td>Age, y</td>
<td>78.6 (5.3)</td>
<td>78.3 (5.1)</td>
</tr>
<tr>
<td>Education, y</td>
<td>13.3 (3.6)</td>
<td>13.6 (3.2)</td>
</tr>
<tr>
<td>Blessed Information Memory Concentration Testb</td>
<td>2.6 (2.5)</td>
<td>2.2 (2.1)</td>
</tr>
<tr>
<td>Free and Cued Selective Reminding Testc</td>
<td>30.0 (6.8)</td>
<td>30.8 (4.9)</td>
</tr>
<tr>
<td>Vocabulary subtest of WAIS-Rd</td>
<td>46.9 (13.9)</td>
<td>49.9 (10.9)</td>
</tr>
<tr>
<td>Digit Symbol Substitution subtest of WAIS-Rd</td>
<td>34.5 (11.4)</td>
<td>35.9 (8.6)</td>
</tr>
<tr>
<td>Variability</td>
<td>0.74 (0.40)</td>
<td>0.29 (0.11)</td>
</tr>
<tr>
<td>Illness index</td>
<td>1.2 (1.0)</td>
<td>1.2 (1.0)</td>
</tr>
</tbody>
</table>

**No. (%)**

| Women | 538 (60) | 134 (60) | 146 (65) | 129 (58) | 129 (58) |
| Incident dementia cases | 61 (7) | 3 (1) | 19 (8) | 13 (6) | 26 (12) |
| Depression | 83 (9) | 28 (12) | 23 (10) | 14 (6) | 18 (8) |
| Angina | 90 (10) | 22 (10) | 22 (10) | 25 (11) | 21 (9) |
| Diabetes | 51 (6) | 13 (6) | 12 (6) | 15 (7) | 11 (5) |
| Chronic obstructive pulmonary disease | 35 (4) | 11 (5) | 9 (4) | 8 (4) | 7 (3) |
| Hypertension | 464 (52) | 114 (51) | 117 (52) | 121 (54) | 112 (50) |
| Myocardial infarction | 90 (10) | 25 (11) | 18 (8) | 26 (12) | 21 (9) |
| Chronic heart failure | 12 (1) | 2 (1) | 4 (2) | 4 (2) | 2 (1) |
| Parkinson disease | 12 (1) | 4 (2) | 2 (1) | 2 (1) | 4 (2) |
| Stroke | 86 (10) | 22 (10) | 17 (8) | 25 (11) | 22 (10) |

Abbreviation: WAIS-R, Wechsler Adult Intelligence Scale Revised.

a Variability denotes within-person across-neuropsychological test variability. Increased variability denotes worse cognitive function.

b Possible score range is 0 to 32.

c Possible score range is 0 to 48.

©2008 American Medical Association. All rights reserved.
INCIDENT DEMENTIA AND NEUROPSYCHOLOGICAL TEST VARIABILITY

for performance level on the individual tests, as expected, but the correlations between variability and performance on the neuropsychological tests were relatively low (FCSRT, r = −0.21, P < .001; WAIS-R Digit Symbol Substitution, r = −0.12, P < .001; WAIS-R Vocabulary, r = −0.18, P < .001), suggesting that these measures were not collinear. The lower, but significant HR for within-person across-neuropsychological test variability in the third compared with the second Cox analysis, along with HRs for level of performance on the individual neuropsychological tests that were similar in the 2 analyses, are evidence that part, but not all, of the effect of variability is mediated through changes in levels of cognitive performance. We used the partial likelihood ratio test to examine whether adding within-person across-neuropsychological test variability to neuropsychological test performance improved prediction of dementia compared with using neuropsychological test performance alone. The results were significant (log partial likelihood ratio with and without variability, respectively, −200.5 and −202.6; χ² = 4.19, P = .04), indicating that prediction of incident dementia was improved by adding within-person across-neuropsychological test variability. We also examined whether within-person across-neuropsychological test variability improved the sensitivity for the prediction of dementia within 1 year. In a model that included the 3 individual neuropsychological tests and controlled for age, sex, education, and disease illness, a score combining the coefficients of all covariates in the model that resulted in 80% specificity for detecting dementia yielded 83% sensitivity for predicting dementia within 1 year. Including within-person across-neuropsychological test variability in this model, maintaining specificity at 80%, significantly increased the sensitivity for predicting dementia within 1 year to 88% (McNemar test, P = .01).

Kaplan-Meier curves were used to delineate the cumulative risk of incident dementia using age as the time scale. Within-person across-neuropsychological test variability dichotomized into the highest (ie, worst) quartile vs the rest revealed that (consistent with the previous results) higher variability was associated with increased risk of incident dementia (HR, 2.25; 95% CI, 1.32-3.86) (Figure, A). Additionally, Kaplan-Meier curves for the FCSRT (HR, 4.77; 95% CI, 2.80-8.10), WAIS-R Digit Symbol Substitution (HR, 2.43; 95% CI, 1.44-4.09), and WAIS-R Vocabulary tests (HR, 1.51; 95% CI, 0.88-2.57) dichotomized into the lowest (ie, worst) quartile vs the rest are illustrated in the Figure, B-D, respectively.

Exploratory Analyses Using Clinical Consensus Diagnosis of Dementia Subtypes

Cox proportional hazards regression analyses with age serving as the time scale, adjusting for sex, education, and illness index, were used to examine whether associations between variability and clinical diagnosis of dementia using consensus case conference procedures varied as a function of dementia subtype. Variability predicted development of both incident dementia of the Alzheimer type (n = 47; HR, 3.63 [95% CI, 1.79-7.37]), and incident vascular dementia (n = 18; HR, 5.26 [95% CI, 1.70-16.30]). The HR for Alzheimer disease after adjusting for neuropsychological test results was 1.68 (95% CI, 0.80-3.55; P = .17); these results are comparable with dementia overall but power was limited to demonstrate a significant difference. The disease processes underlying pure subtypes of Alzheimer dementia and vascular dementias are different. Thus, it appears that variability is sensitive to dementia irrespective of disease subtype.

COMMENT

Our assessment of intrindividual variability addresses within-person differences in cognitive function and its use as a marker of pathology. The taxonomy of intrindividual variability offers several operational definitions for this construct. To our knowledge, this is the first study to demonstrate that within-person across-neuropsychological test variability is associated with incident dementia in a population-based cohort aged 70 years or older, even after adjusting for level of performance on each individual neuropsychological test.

The premise guiding the study was that variability or inconsistency in performance across neuropsychological tests would increase in the preclinical stages of dementia compared with normal aging. This, in part, is attributed to the decline in memory function years before the diagnosis of dementia. As opposed to summary scores or individual indices that estimate function within 1 cognitive domain, within-person across-neuropsychological test variability may be conceptualized as a single representation of variability across multiple domains subserved by several cortical regions and networks. In this context, this form of cognitive variability may be considered a signature of decline in cerebral integrity in the early stages of dementing illnesses.

In choosing the individual tests used to calculate within-person across-neuropsychological test variability, we aimed to maximize the number of participants, represent different cognitive domains, and increase generalizability by including a few easily administered tests that are commonly used in other studies and by clinicians assessing cognitive impairment in elderly individuals. Within-person across-neuropsychological test variability can be computed using tests other than those used in this study. Further, this form of cognitive variability can be estimated using standard and widely used clinical and neuropsychological assessment procedures that are typically given in 1 testing session. Hence, the potential clinical utility of this aspect of cognitive function is quite appealing because it requires no changes to standard assessment procedures in aging studies or assessment of cognitive disorder.
in clinic settings. It is noteworthy that variability on subdomains of the Mini-Mental State Examination\(^4\) was related to short-term decline on a global measure of cognitive function in a small cohort of centenarians.\(^4\) Hence, examining across-cognitive domain variability in screening measures, including those currently used to support the diagnosis of mild cognitive impairments,\(^4\) is of interest.

Research and theories concerning within-person variability on single measures, although relatively recent, suggest that increased inconsistency represents impaired top-down executive control processes.\(^5\) However, at present, little is known about within-person across-cognitive test variability, the theoretical and neuroanatomical basis for this putative construct. Our findings suggest that even though substantial in older adults without dementia,\(^2\) within-person across-neuropsychological test variability may be pathological when a certain threshold is exceeded. It is noteworthy that variability was sensitive to both Alzheimer and vascular dementia subtypes, which vary in terms of their etiology and cognitive profile.\(^3\) Therefore, as a signature of early decline in global cerebral integrity, variability may capture the summation of differing sensitivities of brain regions and networks to various disease processes as opposed to estimating the effect of disease on a single brain region. The relatively low correlations between variability and the level of performance on the individual neuropsychological tests appear to support this notion.

It is of further interest to examine whether and the extent to which different aspects of intraindividual variability in cognitive functions are related in terms of the theory, underlying mechanism, and utility in predicting outcomes of interest. For instance, variability denotes within-person across-neuropsychological test variability. Increased variability denotes worse cognitive function. CI indicates confidence interval; HR, hazard ratio; WAIS-R, Wechsler Adult Intelligence Scale Revised.
research suggests that individuals are not consistently diagnosed with mild cognitive impairment across multiple sessions. Although some fluctuations may be attributed to error in measurement and limitations of this construct, it would be of interest to examine the relationship between across-testing and within-testing session variability and whether the latter improves understanding and prediction of transitional stages in cognitive aging.

The limitations of the study should be considered. First, the sample, although representative of the Bronx, consisted of volunteers who resided in the community and who were relatively healthy and willing to travel to the medical center. Hence, the generalizability of the findings may potentially be limited. Second, the diagnosis of incident dementia was based on an algorithmic procedure that was independent of neuropsychological test scores and consensus diagnosis. Hence, diagnostic circularity is less likely to confound the findings. However, to examine longitudinal associations between this form of variability and clinical diagnosis of dementia, within-person across-neuropsychological test variability ideally should be derived from tests that are not part of the clinical diagnostic procedures. Third, assignment of subtypes of dementia is falible. Although the diagnoses were made according to standardized criteria, some misclassification is inevitable. Further, differential associations between variability and dementia subtypes may be demonstrated, perhaps when using neuropsychological tests that are more sensitive to either vascular or Alzheimer diseases. Fourth, attrition is a major issue of concern in any longitudinal study. A significant number of individuals enrolled in the EAS were not eligible to participate in this investigation because they had only the baseline evaluation and were awaiting the next yearly visit. However, the subsample included in this study was not different from the entire EAS cohort in terms of age, sex distribution, and level of education (data available from author on request). Nonetheless, the potential for selection bias should be considered. We emphasize that a single variability measure cannot replace a complete neuropsychological examination, nor do we advocate that 3 neuropsychological tests are sufficient to provide adequate assessment of an individual’s cognitive function. Instead, we propose that measures of within-person across-test variability be viewed as complementary to standard assessment procedures that are used to predict future risk of dementia. Finally, our results were identified in a single population; these results should be validated in a new population and their value in discriminating between populations that will and will not develop dementia should be defined before intraperson across-neuropsychological test variability is used routinely in clinical settings.

In summary, within-person across-neuropsychological test variability was associated with development of dementia independently of performance of the neuropsychological tests. This finding needs to be replicated in different populations before it is applied in a clinical setting.

Authors Contributions: Dr Holtzer had full access to all the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

Study concept and design: Holtzer, Verghese.

Acquisition of data: Lipton, Verghese.

Analysis and interpretation of data: Holtzer, Verghese, Hall, Wang, Lipton.

Drafting of the manuscript: Holtzer, Verghese.

Critical revision of the manuscript for important intellectual content: Holtzer, Verghese, Hall, Wang, Lipton.

Statistical analysis: Wang, Hall, Holtzer, Verghese.

Obtained funding: Lipton.

Administrative, technical, or material support: Verghese, Lipton.

Study supervision: Holtzer, Verghese.

Financial Disclosures: None reported.

Funding/Support: The Einstein Aging Study is supported by the National Institute on Aging program project grant AG03949. Dr Holtzer is supported by the National Institute on Aging Paul B. Beeson Award K23 AG03857. Dr Verghese is supported by the National Institute on Aging grant AG025119.

Role of the Sponsor: The National Institute on Aging did not have any role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; and preparation, review, or approval of the manuscript.

REFERENCES


©2008 American Medical Association. All rights reserved.
INCIDENT DEMENTIA AND NEUROPSYCHOLOGICAL TEST VARIABILITY


No being can be what he is unless he is putting his essence into action in his field.

—Arnold J. Toynbee (1889-1975)