Journal of Alzheimer's Disease xx (20xx) x–xx
DOI 10.3233/JAD-160259
IOS Press

Within-Individual Variability: An Index for Subtle Change in Neurocognition in Mild Cognitive Impairment

David R. Roalf[a,*], Megan Quarmley[a], Dawn Mechanic-Hamilton[a], David A. Wolk[b,c], Steven E. Arnold[a,b,c,d], Paul J. Moberg[a,b,c] and for the Alzheimer's Disease Neuroimaging Initiative[1]
[a]*Departments of Psychiatry, Philadelphia, PA, USA*
[b]*Departments of Neurology, Philadelphia, PA, USA*
[c]*Departments of Alzheimer's Disease Center of the University of Pennsylvania Perelman School of Medicine, Philadelphia, PA, USA*
[d]*Department of Neurology, Massachusetts General Hospital, Harvard Medical School, Boston, MA, USA*

Handling Associate Editor: Andrea Tales

Accepted 4 June 2016

Abstract.
Background: The transition from mild cognitive impairment (MCI) to Alzheimer's disease is characterized by a decline in cognitive performance in many domains. Cognitive performance profiles in MCI are heterogeneous, however, and additional insights into markers of incipient dementia are needed. Typically, studies focus on average or mean performance, but ignore consistency of performance across domains. WIV (within-individual variability) provides an index of this consistency and is a potential marker of cognitive decline.
Objective: To use neurocognitive data from the Alzheimer's Disease Neuroimaging Initiative cohort to measure neurocognitive variability.
Methods: The utility of WIV was measured, in addition to global neurocognitive performance (GNP), for identifying AD and MCI. In addition, the association between changes in neurocognitive variability and diagnostic transition over 12 months was measured.
Results: As expected, variability was higher in AD and MCI as compared to healthy controls; GNP was lower in both groups as compared to healthy subjects. Global neurocognitive performance alone best distinguished those with dementia from healthy older adults. Yet, for individuals with MCI, including variability along with GNP improved diagnostic classification. Variability was higher at baseline in individuals transitioning from MCI to AD over a 12-month period.
Conclusion: We conclude that variability offers complementary information about neurocognitive performance in dementia, particularly in individuals with MCI, and may provide beneficial information about disease transition.

Keywords: ADNI, Alzheimer's disease, intra-individual variability, mild cognitive impairment, neurocognitive function

Uncorrected Author Proof

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

*Correspondence to: David R. Roalf, PhD, Department of Psychiatry, Neuropsychiatry Section, Brain Behavior Laboratory, 10th Floor, Gates Building, Hospital of the University of Pennsylvania, Philadelphia, PA 19104, USA; E-mail: roalf@upenn.edu.

ISSN 1387-2877/16/$35.00 © 2016 – IOS Press and the authors. All rights reserved

INTRODUCTION

Cognitive deficits are the defining features of dementia. These impairments are strong predictors of functional outcome [1, 2], and are associated with alterations in brain structure and function [3]. Hence, the focus of recent neurocognitive studies is on individuals at risk for developing dementia. This risk period, typically associated with mild cognitive impairment (MCI), is signified by a measurable deterioration in cognitive function that is greater than expected based upon an individual's age and education, but does not meaningfully affect a person's daily functioning [4]. Despite being a major research focus in recent years [5], establishing the diagnosis of MCI [6], and monitoring disease progress using neuropsychological function over time remains challenging. Moreover, there is little work in developing measures that focus on and monitor individual differences in neurocognition, despite significant evidence of heterogeneity in disease presentation and progression. Recent investigations of within-individual variability (WIV; or intraindividual variability) confer unique predictive information about cognitive functioning beyond mean performance [7–10], and suggest this measure to be a relatively stable characteristic of an individual.

Early and accurate detection of cognitive impairments that precede dementia will enhance understanding of possible individual differences in disease trajectory as well as clinical management. To this effect, recent studies of neuropsychological function in MCI confirm the utility of neuropsychological tests for early detection and prevention strategies [11, 12]. Not surprisingly, use of an efficient, but multi-dimensional neuropsychological inventory (CERAD-NB) is more accurate at distinguishing MCI or AD from healthy individuals (HC) than brief screening measures. Yet, diagnostic accuracy declines when using these instruments to distinguish between HC and MCI or MCI and AD [11]. Difficulty in differentiating MCI is likely due to several factors including, but not limited to: 1) the heterogeneity of causes of MCI diagnosis; 2) variable progression rates from MCI to AD per year [13]; 3) the historical focus of research on *cross-sectional* differentiation of MCI from AD and healthy older adults (see [6]); and 4) the relative dearth of valid screening measures for detecting subtle deficits of early stage or prodromal AD. The importance of this last point cannot be overstated as the value of any test will be in its ability to accurately differentiate diagnostic features and identify markers of further decline.

Neurocognitive variability across tests has been measured before in aging and lifespan research [7], but with inconsistent nomenclature, including 'dispersion' [14, 15], 'within-person variability' [16], 'within-person across neuropsychological test variability' [17], and 'intra-individual differences' [7]. Here we define WIV as inconsistent relative strengths and weaknesses in test performances within or across domains [7, 18]. Thus, WIV is estimated within a single person, but across several tasks in several domains and therefore provides an index of evenness, or consistency, of neurocognitive performance. Specifically, a low WIV value indicates a relatively consistent within-individual performance profile, whereas a high WIV value indicates uneven performance profiles [15]. Measuring WIV therefore could be another, more sensitive way, to document the emergence of problematic individual performance differences. This type of variability differs from 'intra-individual variability' (e.g., across-trial IIV) within a given test [19, 20], which typically focuses on consistency in performance speed. WIV has been operationalized two ways: 1) variability associated with measuring one individual at one time point across multiple neurocognitive tasks; or 2) measuring one individual on a single task across multiple occasions [7]. It is important to note that when WIV is small, mean performance is a robust metric; however, if WIV is high, the utility of mean performance diminishes. Thus, exclusive reliance upon mean performance without considering WIV may lead to inaccurate conclusions [21, 22]. Consequently, WIV has emerged as a useful construct for assessing cognitive performance in many disorders. WIV is higher than normal in individuals with cognitive decline, Parkinson's disease, frontotemporal dementia, ADHD (for reviews see [7, 21, 22]), and in dementia [10]. Here, we MCI and AD from the Alzheimer's Disease Neuroimaging Initiative (ADNI) sample and compared WIV in those individuals who transition from one diagnostic category to another within one year. Using a measure of across-task variability, we hypothesized 1) that WIV would improve upon diagnostic classification of mean neurocognitive performance and 2) WIV would be a sensitive index (e.g., lower at baseline, more change over time) in individuals who transition from MCI to AD.

Uncorrected Author Proof

MATERIALS AND METHODS

Study population: The Alzheimer's disease neuroimaging initiative

Data used in this study were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (http://adni.loni.usc.edu) and approval for this project was granted [23]. The ADNI was launched in 2003 as a public-private partnership, led by Principal Investigator Michael W. Weiner, MD. The primary goal of ADNI has been to test whether serial magnetic resonance imaging (MRI), positron emission tomography (PET), other biological markers, and clinical and neuropsychological assessments can be combined to measure the progression of MCI and early AD. For up-to-date information, see http://www.adni-info.org.

The neuropsychological data were collected in 229 healthy (HC), 397 MCI, and 193 AD individuals. There were significantly more males diagnosed with MCI at baseline $\chi^2(2) = 12.37, p = 0.002$. Overall, HC were older than MCI patients ($p = 0.023$), but not the AD patients (Table 1). HC attained higher levels of education than the AD patients ($p < 0.001$), but not the MCI patients (Table 1).

Diagnostic assessments included history and physical and neurologic examinations conducted by experienced clinicians. On the basis of these data, a consensus diagnosis was established using standardized clinical criteria for AD, MCI [24], or other neurological or psychiatric conditions presenting with cognitive impairment (see [25] for description). Screening assessments included the Mini-Mental State Examination (MMSE; [26]) and Clinical Dementia Rating [2]. Informed consent for the use of all data was obtained from all persons, in accord with university institutional review board–approved protocols. As expected, MMSE differed by group with AD<MCI<HC ($F(1, 809)=454.73$, $p < 0.001$).

The ADNI neuropsychological battery includes tests that assess multiple cognitive domains including memory, executive functioning, attention and language [24]. Detailed descriptions of the tasks including administration and scoring instructions can be found here: http://www.adni-info.org/Scientists/ADNIStudyProcedures.html.

For the current analysis, memory tests included immediate and delayed recall of the Rey Auditory Verbal Learning Test (RAVLT) [27] and immediate and delayed recall from the Wechsler Memory Scale-Revised [28] Logical Memory subtests. Executive function and attention tasks included the Digit Span Test (forward and backward) [28], the Trail Making Tests (Part A and B) [29], and the Digit Symbol Substitution Test [30]. Test of language function included semantic word-list generation (animal and vegetable fluency) and visual confrontation naming (Boston Naming Test (BNT); [31]). Additional neuropsychological measures included the Alzheimer's Disease Assessment Scale–Cognitive subscale (ADAS-Cog; [32] and the Clock Drawing Test [33]). Total scores for each task were used as outcome measures. For measures with immediate and delayed memory scores, each score was considered separately. For the Trail Making Test the difference in time between the completion of A and B was used as the outcome measure. For the current study, neuropsychological data from the baseline visit and 12-month visit were included.

Neuropsychological within-individual-variability (WIV)

Performance values were transformed to their standard equivalents based on the means and standard deviation (SD) of the healthy sample. An index of WIV across tasks was calculated for each participant in SD units (see [17]). This index of variability has been used in other studies [9, 17, 34], and reflects variation *within a single person* across several neurocognitive tasks and is therefore an index of an individual's evenness or dispersion of performance across neuropsychological domains. WIV was calculated at baseline and at the 12-month follow-up. In addition, a global index of neurocognitive performance (GNP) was calculated by averaging the standardized scores (z-scores) across all neuropsychological tasks. Three individuals were excluded from analysis for having completed fewer than half of the tasks. All other participants completed a minimum of 9 of the 12 neuropsychological tasks.

$$\text{Within-individual variability} = \sqrt{\sum_{k=1}^{K}\left(\frac{(Z_{ik}-A_i)^2}{K-1}\right)}$$

where Z_{ik} is the kth test score for the ith individual and:

$$A_i = \sum_{k=1}^{K}\frac{Z_{ik}}{K}$$

Uncorrected Author Proof

Table 1
Demographic characteristics ~~and~~ GNP and WIV performance for AD, MCI, and healthy participants

Mean (SD)	Healthy ($n = 229$)	MCI ($n = 397$)	AD ($n = 193$)
Age	75.93 (5.03)	74.80 (7.44)	75.36 (7.47)
Education	16.04 (2.87)	15.67 (3.04)	14.07 (3.13)
Sex (% Female)	48%	36%	47%
MMSE baseline	29.11 (1.00)*#	27.03 (1.79)#	23.34 (2.06)
MMSE 12 months	29.13 (1.17)*#	26.40 (2.87)#	21.13 (4.46)
CDR baseline	0 (0)*#	0.50 (0.03)#	0.74 (0.25)
CDR 12 months	0.04 (0.14)*#	0.53 (0.18)#	0.99 (0.50)
Number of Neuropsychological Tasks completed baseline	11.74 (0.49)	11.97 (0.20)	11.90 (0.38)
Number of individuals at follow-up with sufficient data	208	327	144
GNP baseline	0.00 (0.50)*#	–1.35 (0.80)#	–2.42 (0.92)
GNP 12 months	0.00 (0.53)*#	–1.42 (1.01)#	–2.79 (1.4)
WIV baseline	0.87 (0.23)*#	1.45 (0.48)#	1.89 (0.62)
WIV 12 months	0.81 (0.24)*#	1.53 (0.57)#	2.08 (0.65)

*$p < 0.001$ as compared to MCI; #$p < 0.001$ as compared to AD. AD, Alzheimer's disease; CDR, Clinical Dementia Rating; GNP, global neuropsychological performance; MCI, mild cognitive impairment; MMSE; Mini-Mental State Examination; WIV, within-individual variability.

Statistical analysis

WIV values were entered as dependent measures in ANCOVAs with MMSE, age, and sex included as factors and education ~~was included~~ as a covariate. *Post-hoc* contrasts were used to examine interactions; Satterthwaite corrections were used when equal variances could not be assumed; corrected degrees of freedom are reported where appropriate. Pearson correlations were performed between demographic and performance variables. Cohen's *d* values are presented for group-specific WIV comparisons. The diagnosis accuracy for each measure (or combination of measures) was calculated as the area under the receiver operating characteristic (ROC) curve (AUC). The AUC measure represents the mean sensitivity value for all possible values of specificity and larger AUC values indicate more accurate classification of participants. The logistic ROC analysis used a 10-fold cross-validation approach to estimate AUC and optimal cut-off score. AUCs were compared using DeLong's non-parametric method [35]. A cut-off score for each measure that best differentiated diagnostic groups was determined using the Youden Index [36], which maximizes the tradeoffs between sensitivity and specificity. The classification accuracy (probability of correct classification of subject with or without impairment at a given cut-off score) was calculated based upon these cut-off scores (Table 2). Diagnostic accuracy of the GNP and WIV (or combination of measures) were compared via Chi-Square analysis. Classification accuracy of each measure was compared using the Wilcoxon Signed Rank test. All statistical analyses were performed in R [37].

RESULTS

Within-individual variability and global neurocognitive performance in AD and MCI

WIV differed by diagnostic group, $F(2, 810) = 258.42$, $p < 0.0001$, but not by age ($p = 0.68$) or sex ($p = 0.44$) and there were no significant interactions (Fig. 1). As hypothesized, AD patients had higher WIV than MCI [$t(305.31) = 8.76$, $p < 0.001$; Cohen's $d = 0.80$] or HC [$t(238.46) = 21.58$, $p < 0.001$; Cohen's $d = 2.18$] and MCI had higher WIV than HC [$t(610.16) = 20.25$, $p < 0.001$; Cohen's $d = 1.54$]. These effects remained significant in an age and gender matched subsample (see Supplementary Material). Exploratory group-specific analysis found that in HC, age $F(1, 220) = 5.01$, $p = 0.02$, was significantly related to WIV, while there were no effects of MMSE, sex, nor any interactions. In the MCI group, a lower MMSE score $F(1, 391) = 24.18$, $p < 0.001$ and lower education, $F(1, 391) = 8.62$, $p = 0.004$, were associated with higher WIV; age, sex, and the interactions were not significant. In AD, lower MMSE scores were associated with higher WIV, $F(1, 391) = 10.05$, $p < 0.002$; age, sex, and the interactions were not significant. Associations between MMSE and WIV are displayed in Fig. 2. As expected, GNP differed by diagnostic group, $F(2, 801) = 519.42$, $p < 0.0001$, by age $F(1, 801) = 5.20$,

Uncorrected Author Proof

Table 2
Diagnostic parameters for GNP, WIV and the combination of the two metrics across AD, MCI, and HC

		HC versus AD	HC versus MCI	MCI versus AD
GNP	AUC (±95% CI)	**0.99 (0.99–1.0)**	0.94 (0.92–0.96)	**0.81 (0.78–0.85)**
	Sensitivity/Specificity	**0.98/0.96**	0.85/0.88	**0.71/0.76**
	Youden Index	**0.94**	0.73	**0.47**
	Cutoff[#]	**–0.86**	–0.52	**–1.85**
	Classification Accuracy	**94%**	85%	**74%**
WIV	AUC (±95% CI)	0.96 (0.95–0.98)	0.89 (0.86–0.91)	0.72 (0.68–0.76)
	Sensitivity/Specificity	0.88/0.93	0.79/0.83	0.64/0.73
	Youden Index	0.81	0.62	0.37
	Cutoff	1.26	1.09	1.64
	Classification Accuracy	90%	80%	70%
GNP+WIV	AUC (±95% CI)	0.99 (0.99–1.0)	**0.95 (0.94–0.97)**[*]	0.81 (0.78–0.85)
	Sensitivity/Specificity	0.98/0.97	**0.85/0.93**	0.71/0.76
	Youden Index	0.95	**0.78**	0.47
	Classification Accuracy	98%	**94%**	83%

Bold text indicated the best model for predicting group. [*]Overall AUC is significantly improved with the addition of a measure of variability (WIV). [#]GNP Cutoff scores are reported as a Z-score. AD, Alzheimer's disease; AUC, area under the curve; CI, confidence interval; GNP, global neuropsychological performance; HC, healthy participants; MCI, mild cognitive impairment; WIV, within-individual variability.

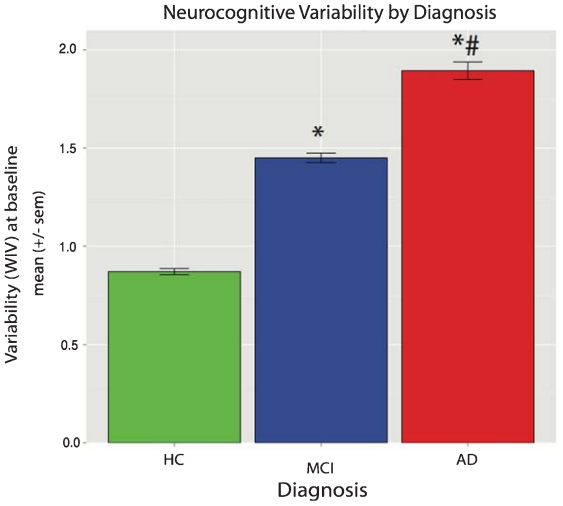

Fig. 1. Neurocognitive variability is high in AD and MCI. Mean across-task WIV in HC, MCI, and AD. WIV was associated with diagnosis. [*]$p < 0.001$ as compared to HC. [#]$p < 0.001$ as compared to MCI.

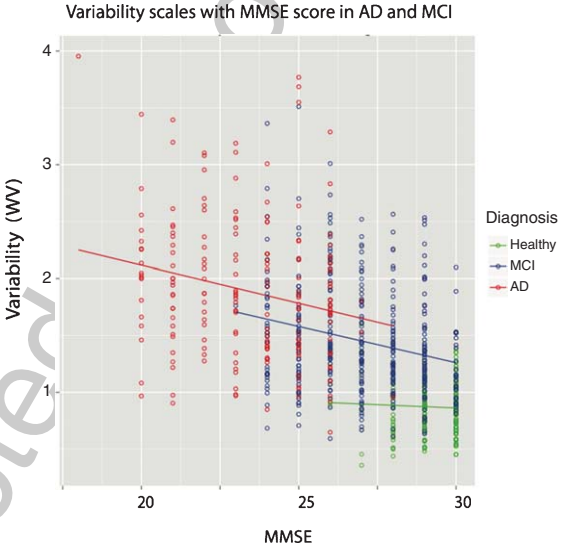

Fig. 2. Neurocognitive variability is associated with MMSE performance. Correlation between MMSE and WIV was significant in AD and MCI, but not HC.

$p < 0.03$, but not sex ($p = 0.07$), and there were no significant interactions. As seen in Table 1, HC outperformed MCI, who in turn outperformed AD. Plots of performance by task are shown in the Supplementary Material.

The association between WIV and GNP was measured using Pearson correlation. Higher variability was associated with poorer GNP across diagnostic group (Pearson r = –0.78, $p < 0.001$). This relationship was more prominent in the two patient groups: AD (r = –0.67, $p < 0.001$) and MCI (r = –0.62, $p < 0.001$), but was also significant in the HC (r = –0.23, $p < 0.001$). In addition, variability at baseline was associated with variability at 12-month follow-up (r = 0.73, $p < 0.001$). Again, this association was strongest in the two patient groups: AD (r = 0.58, $p < 0.001$), MCI (r = 0.54, $p < 0.001$); however, HC (r = 0.37, p < 0.001) also showed a significant positive correlation.

ROC analysis of GNP and WIV in AD, MCI, and HC

The ROC analysis was used to evaluate the diagnostic accuracy of each measure (GNP and WIV) to

Uncorrected Proof

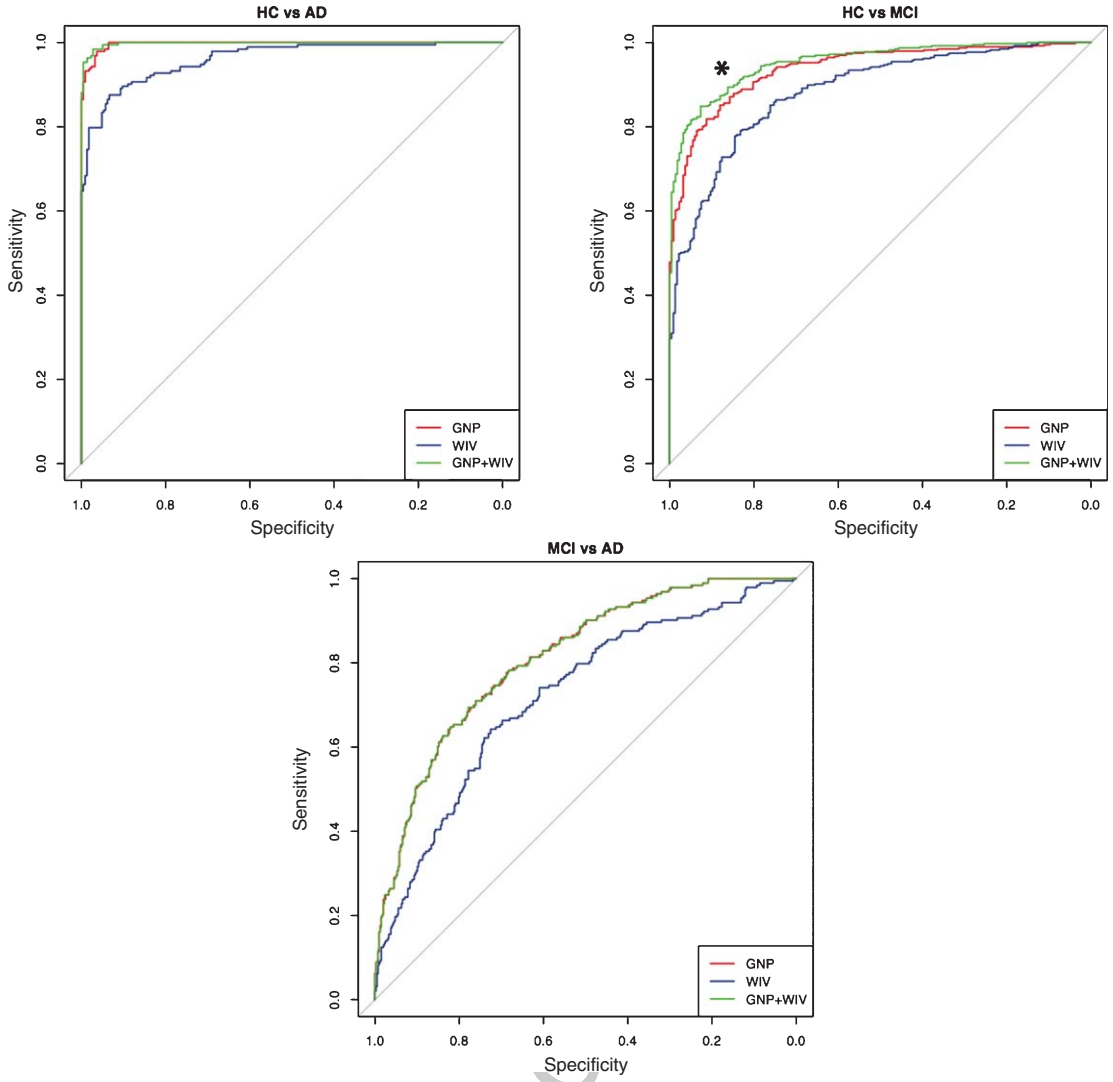

Fig. 3. Variability aids in diagnostic classification in MCI. ROC curve of the GNP, WIV, and GNP+WIV in HC versus AD, HC versus MCI, and AD versus MCI. Including WIV in diagnostic classification improved differentiation of HC and MCI (*$p < 0.01$).

discriminate AD and MCI from each other and from healthy cognitive subjects. Graphic representations of the ROC curves are provided in Fig. 3, and Table 2 shows clinically relevant cut-offs for the GNP, WIV, and combination of the two measures. The diagnostic accuracies of GNP and WIV were excellent for HC versus AD, with AUCs > 0.96. Diagnostic accuracies were lower, but still very good in the GNP (0.94) and WIV (0.89) for HC versus MCI. Diagnostic accuracies of both measures were the lowest when differentiating AD from MCI, yet the AUCs were still moderate to good: GNP (0.81) and WIV (0.72). A comparison of AUC between GNP and WIV is presented in the Supplementary Material.

Combining WIV and GNP improves diagnostic accuracy in MCI

Combining WIV with GNP significantly improved diagnostic accuracy as compared to using either measure alone for discriminating MCI from HC (Table 2/Fig. 4, green lines). Specifically, considering WIV in addition to GNP improved diagnostic accuracy when differentiating HC from MCI [$Z = 3.32$, $p < 0.0001$]. This was due to an increase in specificity (improvement in identifying HC). GNP alone was best for differentiating HC from AD and MCI from AD; WIV did not significantly increase diagnostic discrimination (AUC) in these two comparisons.

Uncorrected

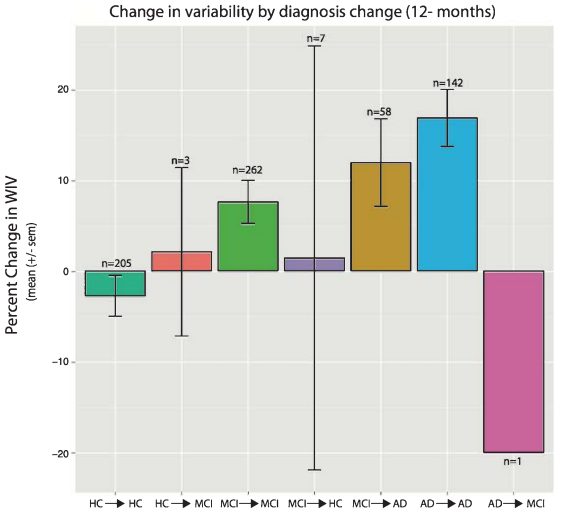

Fig. 4. Change in variability over 12 months is highest in MCI individuals that transition to AD. Percent change of WIV for all individuals with follow-up neurocognitive assessments at 12 months. Data is displayed for groups of individuals that both showed a diagnostic transition and those that remained stable.

However, using the combination of optimal and clinically relevant cut-off scores (a score below either cut-off score) of the GNP and/or WIV to classify individuals resulted in an increase in classification accuracy of 4% at the optimal cut-off for HC versus AD ($V = 153$, $p < 0.01$), 9% for HC and MCI ($V = 1711$, $p < 0.001$) and 11% for MCI and AD ($V = 1378$, $p < 0.001$). Cut-off scores for each measure are provided in Table 2.

WIV is higher in individuals that transition from MCI to AD

The majority of individuals (83%) had neuropsychological data available at 12-month follow-up (Table 1). Approximately 10% of individuals had a diagnostic change within 12 months, the vast majority transitioning from MCI to AD (85% or 58/68 subjects). Since the majority of diagnostic change was seen within the MCI group, follow-up statistical analyses were conducted only within the MCI→MCI and MCI→AD group. The change in variability was calculated as WIV at 12 months minus WIV at baseline. The change in WIV is shown for all groups in Fig. 4. Age did not differ between the MCI→MCI and MCI→AD groups ($p = 0.21$). The MCI→AD group $F(1, 314) = 6.51$, $p = 0.01$) had more change in WIV over a 12 month period (Fig. 4). On average, there was a 12% increase in variability in the MCI→AD group, while the MCI→MCI group had approximately a 6% increase over the same time period. In addition, the MCI→AD group had higher baseline WIV [$t(77.88) = 3.20 = 0.002$] and higher WIV at 12 months [$t(97.05) = 3.16 = 0.001$]. Importantly, HC→HC showed *lower* WIV at 12 months (~3% reduction), while AD→AD patients showed, on average, a 17% increase in WIV. There were a small number of individuals ($n = 7$) with a change from MCI→HC, however this group was too small to perform meaningful statistical analysis.

DISCUSSION

We report the utility of measuring neurocognitive WIV in individuals at risk for dementia. We report lower GNP and higher WIV in MCI and AD as compared to healthy older adults. Higher variability was associated with higher MMSE and poorer overall neurocognitive performance; however, this was more prominent in MCI and AD patients than in healthy individuals. Most importantly, we show that diagnostic change from MCI to AD corresponds with greater baseline WIV and a larger 12-month change in WIV. In addition, we show that WIV can add to the differentiation between diagnostic groups, particularly between healthy individuals and MCI.

The preponderance of the neurocognitive research in dementia emphasizes mean group differences in neurocognitive performance, which typically ignores within-individual variability. As we show, within-individual variability appears to be a useful tool to monitor individual differences in neurocognition, and aids in diagnostic differentiation of MCI from healthy individuals. WIV has not been thoroughly studied in AD and MCI, but is more common in aging and lifespan research. Our findings indicate that WIV is associated with general cognitive performance (i.e., MMSE), but not age; these findings parallel other large studies of aging [14, 15]. Yet, the specific aims of these two studies were not to directly compare healthy individuals to those with MCI or AD, although Lindenberger & Baltes provide exploratory findings in a small cohort of individuals with dementia. Another study, which measured the deviation of measures of cognition from crystalized intelligence in healthy aging, found age to be associated with higher variability [38], particularly when compared to healthy young adults. In addition, Rabbitt (1993) concludes that when neurocognitive function begins to decline in old age these abilities do not "all go together when they go" [38] (p. 385).

Uncorrected Author Proof

More recently, a study of limited neuropsychological data from the ADNI indicates that computing variability within items or across tests provides a useful summary measure of performance in AD, MCI, and healthy cognitive aging [39], and we show similar results when measuring global WIV in the ADNI sample. More recent evidence from another sample suggests that high WIV is ~~a~~ characteristic of patients with MCI [10], which we replicate. In addition, we show that WIV, in combination with GNP, may help to identify those individuals with a progressive pathological process earlier in the course of disease. Specifically, our measure of variability, which has been used before in aging research [7], examines intra-individual consistency of performance across neurocognitive domains. Our finding that AD and MCI groups were more varied in their performance than our HC group fits with prior reports in aging [21] and dementia [10]. For example, Reckess et al. [10] measured across-task WIV in 528 individuals, 395 with clinical symptoms. Their findings indicate that WIV increases with symptom severity in MCI, but shrinks in those with significant dementia. In conjunction with our findings, this indicates that WIV may be more sensitive in detecting subtle change in those people with ~~milder~~ cognitive impairments (such as MCI), but is less meaningful in frank dementia where mean neuropsychological performance is near floor and sufficient, for detecting dementia level impairment. We do not find lower WIV in advanced AD as compared to MCI; however, Reckess et al. [10] estimated variability using the standard deviation of performance for groups of individuals with specific MMSE scores, whereas we estimated variability over the entire performance spectrum.

Most importantly, WIV appears to provide a sensitive measure to detect small, but potentially meaningful change over short periods of time in MCI. We find MCI individuals that transition to clinical dementia show higher WIV at baseline and a larger increase in WIV over a 12-month period. Thus, higher WIV may reflect domain specific deterioration of cognitive performance. This inconsistency in performance suggests that relative declines in one area of cognition as compared to another are an important signal of overall deterioration of the neural system. This suggests that WIV provides researchers and clinicians a tool that is sensitive to subtle changes in the disease course that traditional approaches fail to detect. Likewise, our findings suggest that WIV may help reduce some of the heterogeneity in the definition of MCI; however, these measures need to be validated in larger samples, prospectively. Similar to changes in average performance, inconsistencies in variability are likely due to underlying changes in the neural architecture; however, this remains to be confirmed. Other studies [22, 40] suggest consistency, or lack thereof, may be a sensitive index of performance over time that is related, in part, to neurotransmitter function [19, 41] or white matter integrity [9]. While WIV provides an index of deterioration it does not identify the specific domain affected. Nonetheless, it is noteworthy that we were able to detect these subtle changes in WIV, at least within the relatively short time parameters of this study. Future studies should consider measuring the relationship between response slowing/variability at the task level and brain function (i.e., diffusion MRI) in dementia.

While our findings are intriguing, there are several limitations to consider. First, we use global scores for performance and variability and these are calculated across domain. Given the modest number of tests in each neurobehavioral domain, the global scores may have been biased by the psychometric characteristics of the test instruments used. Although age-related differences in GNP and WIV appear systematic, there are likely multiple factors that contribute to variability. Differences in WIV therefore may depend to a large extent on the specific constellation of abilities being measured; if so, its generalizability needs further scrutiny. Directly measuring WIV in neurocognitive performance provides a general view of neurocognitive ability. However, Cole et al. [40] argue that using a composite index of neurocognitive domains, such as WIV, provides a better index of consistency in neurocognitive ability. Furthermore, WIV can be advantageous in elucidating common underlying mechanisms of information processing that result in increased variability [42]. This approach may be more sensitive to detecting change over time by taking advantage of the variability within an individual to aid in determining individuals at-risk for dementia. Future longitudinal follow-up studies could expand upon the variability findings by elucidating the specific neurocognitive domains responsible for higher variability in MCI and AD and by examining variability across time *within a specific test*. The contribution of WIV to diagnostic classification was significant, but small. We acknowledge that there is only a small increase in the AUC of the ROC curve for HC versus MCI, however we believe the more important finding is the increase in classification accuracy at the proposed cut-offs since this translates to mean performance scores that could be used clinically. Given

Uncorrected Author Proof

the heterogeneity in the MCI diagnosis, it is unlikely that a global measure of variability will suffice for detecting very subtle performance changes. Measuring variability on a trial-wise basis may further improve the capability of variability in classification. Future studies that implement computerized neurocognitive testing and record reaction time on a trial-by-trial basis should consider evaluating variability in MCI and AD. Moreover, our results support the potential utility of WIV as a dimension or metric related to cognitive decline. Initial evaluation of the construct validity of this new dimension suggests that WIV correlates negatively with average performance. The negative association between higher WIV and lower GNP data adds validity to our measure as the findings align with the substantial literature showing decreases in neurocognitive ability with increasing age and dementia [21]. A single measure of variability over a battery of tasks does not replace a thorough neuropsychological evaluation, as variability is common even in healthy adults [43]. However, measurement of within-person variability as a complementary measure to traditional mean performance metrics provides a generalizable index of neurocognitive performance that is informative and potentially useful in the study of dementia.

Despite being a major research focus in recent years [5], identifying individuals at risk for developing dementia using traditional neurocognitive assessments remains challenging. This is likely due, in part, to defining and diagnosing MCI (and more recently in individuals with subjective memory complaints) in terms of acquired impairment in neurocognitive domains specifically affected in AD [44]. While this approach ensures clinical continuity between MCI and AD, traditional tests tend to be insensitive to early, subtle deficits. Moreover, this approach assumes that early deficits in at-risk individuals are the result of higher order (cortico-cortical) neural disruptions; however, recent work suggests that lower level neural deficits may be more prominent in MCI [45]. We believe that early identification and monitoring of disease progression can be improved upon in MCI by increasing the specificity of neurocognitive testing by measuring variability in neurocognitive performance. The present study of within-individual variability provides evidence that capturing variability in neurocognitive performance is a useful index in MCI and dementia, and may be a beneficial screening tool. Specifically, increases in performance variability may index vulnerability and potential transition to dementia. Moreover, the use of a patient-centered metric, such as WIV, is ideal for monitoring subtle change in performance over extended periods of time, may reflect the unfolding of neurocognitive dysfunction and may be associated with brain deterioration that would go undetected if only mean performance is considered.

ACKNOWLEDGMENTS

This work was supported by NIA AG10124, NIMH K01 MH102609 (Dr. Roalf), and the Institute of Aging and Alzheimer's Disease Core Center Pilot Funding Program (Dr. Roalf).

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

Authors' disclosures available online (http://j-alz.com/manuscript-disclosures/16-0259r1).

Uncorrected Author Proof

SUPPLEMENTARY MATERIAL.

The supplementary material is available in the electronic version of this article: http://dx.doi.org/10.3233/JAD-160259

REFERENCES

[1] Rossetti HC, Cullum CM, Hynan LS, Lacritz L (2010) The CERAD neuropsychological battery total score and the progression of Alzheimer's disease. *Alzheimer Dis Assoc Disord* **24**, 138.

[2] Morris JC, Edland S, Clark C, Galasko D, Koss E, Mohs R, Van Belle G, Fillenbaum G, Heyman A (1993) The Consortium to Establish a Registry for Alzheimer's Disease (CERAD) Part IV. Rates of cognitive change in the longitudinal assessment of probable Alzheimer's disease. *Neurology* **43**, 2457.

[3] Song X, Mitnitski A, Zhang N, Chen W, Rockwood K (2013) Dynamics of brain structure and cognitive function in the Alzheimer's disease neuroimaging initiative. *J Neurol Neurosurg Psychiatry* **84**, 71-78.

[4] Petersen RC (2004) Mild cognitive impairment as a diagnostic entity. *J Intern Med* **256**, 183-194.

[5] Winblad B, Palmer K, Kivipelto M, Jelic V, Fratiglioni L, Wahlund LO, Nordberg A, Bäckman L, Albert M, Almkvist O (2004) Mild cognitive impairment–beyond controversies, towards a consensus: Report of the International Working Group on Mild Cognitive Impairment. *J Intern Med* **256**, 240-246.

[6] Lonie JA, Tierney KM, Ebmeier KP (2009) Screening for mild cognitive impairment: A systematic review. *Int J Geriatr Psychiatry* **24**, 902-915.

[7] MacDonald SW, Li SC, Backman L (2009) Neural underpinnings of within-person variability in cognitive functioning. *Psychol Aging* **24**, 792-808.

[8] Roalf DR, Gur RC, Almasy L, Richard J, Gallagher RS, Prasad K, Wood J, Pogue-Geile MF, Nimgaonkar VL, Gur RE (2013) Neurocognitive performance stability in a muliplex multigenerational study of schizophrenia. *Schizophr Bull* **39**, 529-537.

[9] Roalf DR, Ruparel K, Verma R, Elliott MA, Gur RE, Gur RC (2013) White matter organization and neurocognitive performance variability in schizophrenia. *Schizophr Res* **143**, 172-178.

[10] Reckess GZ, Varvaris M, Gordon B, Schretlen DJ (2014) Within-person distributions of neuropsychological test scores as a function of dementia severity. *Neuropsycholog* **28**, 254.

[11] Roalf DR, Moberg PJ, Xie SX, Wolk DA, Moelter ST, Arnold SE (2013) Comparative accuracies of two common screening instruments for classification of Alzheimer's disease, mild cognitive impairment, and healthy aging. *Alzheimers Dement (Amst)* **9**, 529-537.

[12] Paajanen T, Hänninen T, Tunnard C, Mecocci P, Sobow T, Tsolaki M, Vellas B, Lovestone S, Soininen H (2010) CERAD neuropsychological battery total score in multinational mild cognitive impairment and control populations: The AddNeuroMed study. *J Alzheimers Dis* **22**, 1089-1097.

[13] Petersen RC, Smith GE, Waring SC, Ivnik RJ, Tangalos EG, Kokmen E (1999) Mild cognitive impairment: Clinical characterization and outcome. *Arch Neuro* **56**, 303.

[14] Christensen H, Mackinnon A, Korten A, Jorm A, Henderson A, Jacomb P (1999) Dispersion in cognitive ability as a function of age: A longitudinal study of an elderly community sample. *Neuropsychol Dev Cogn B Aging Neuropsychol Cogn* **6**, 214-228.

[15] Lindenberger U, Baltes PB (1997) Intellectual functioning in old and very old age: Cross-sectional results from the Berlin Aging Study. *Psychol Aging* **12**, 410-432.

[16] Salthouse TA (2007) Implications of within-person variability in cognitive and neuropsychological functioning for the interpretation of change. *Neuropsychology* **21**, 401.

[17] Holtzer R, Verghese J, Wang C, Hall CB, Lipton RB (2008) Within-person across-neuropsychological test variability and incident dementia. *J Am Med Assoc* **300**, 823-830.

[18] Snitz BE, Macdonald AW, 3rd, Carter CS (2006) Cognitive deficits in unaffected first-degree relatives of schizophrenia patients: A meta-analytic review of putative endophenotypes. *Schizophr Bull* **32**, 179-194.

[19] Stuss DT, Murphy KJ, Binns MA, Alexander MP (2003) Staying on the job: The frontal lobes control individual performance variability. *Brain* **126**, 2363-2380.

[20] Klein C, Wendling K, Huettner P, Ruder H, Peper M (2006) Intra-subject variability in attention-deficit hyperactivity disorder. *Biol Psychiatry* **60**, 1088-1097.

[21] MacDonald SWS, Nyberg L, Backman L (2006) Intra-individual variability in behavior: Links to brain structure, neurotransmission and neuronal activity. *Trends Neurosci* **29**, 474-480.

[22] Hultsch DF, Strauss E, Hunter MA, MacDonald SWS (2008) Intraindividual variability, cognition, and aging. In *The handbook of aging and cognition*, Craik FIM, Salthouse TA, eds. Psychology Press, New York, pp. 491-556.

[23] Mueller SG, Weiner MW, Thal LJ, Petersen RC, Jack CR, Jagust W, Trojanowski JQ, Toga AW, Beckett L (2005) Ways toward an early diagnosis in Alzheimer's disease: The Alzheimer's Disease Neuroimaging Initiative (ADNI). *Alzheimers Dement (Amst)* **1**, 55-66.

[24] Landau SM, Harvey D, Madison CM, Reiman EM, Foster NL, Aisen PS, Petersen RC, Shaw LM, Trojanowski JQ, Jack CR (2010) Comparing predictors of conversion and decline in mild cognitive impairment. *Neurology* **75**, 230-238.

[25] Johnson JK, Gross AL, Pa J, McLaren DG, Park LQ, Manly JJ, Alzheimer's Disease Neuroimaging, Initiative (2012) Longitudinal change in neuropsychological performance using latent growth models: A study of mild cognitive impairment. *Brain Imaging Behav* **6**, 540-550.

[26] Folstein MF, Folstein SE, McHugh PR (1975) Mini-mental state: A practical method for grading the cognitive state of patients for the clinician. *J Psychiatr Res* **12**, 189-198.

[27] Rey A (1964) *The clinical examination in psychology*, Presses Universitaires de France, Paris.

[28] Wechsler D (1987) WMS-R: *Wechsler memory scale-revised, Psychological* Corporation.

[29] Reitan RM (1958) Validity of the Trail Making Test as an indicator of organic brain damage. *Percept Mot Skills* **8**, 271-276.

[30] Wechsler D (1981) *WAIS-R manual: Wechsler adult intelligence scale-revised,* Psychological Corporation.

[31] Kaplan EF, Goodglass H, Weintraub S (1983) *The Boston naming test,* 2nd. Lea & Febiger, Philadelphia.

[32] Rosen WG, Mohs RC, Davis KL (1984) A new rating scale for Alzheimer's disease. *Am J Psychiatry* **141**, 1356-1364.

Uncorrected Author Proof

[33] Goodglass H, Kaplan E (1983) Clock Drawing Test. *Assessment of Aphasia and Related Disorders. 2nd ed.* Lea and Febiger, Philadelphia, PA.

[34] Roalf DR, Gur RE, Ruparel K, Calkins ME, Satterthwaite TD, Bilker W, Hakonarson H, Harris LJ, Gur RC (2014) Within-individual variability in neurocognitive performance: Age and sex-related differences in children and youths from ages 8 to 21. *Neuropsychology* **28**, 506-518.

[35] DeLong ER, DeLong DM, Clarke-Pearson DL (1988) Comparing the areas under two or more correlated receiver operating characteristic curves: A nonparametric approach. *Biometrics* 837-845.

[36] Youden WJ (1950) Index for rating diagnostic tests. *Cancer* **3**, 32-35.

[37] R-Core-Team (2012) R Foundation for Statistical Computing, Vienna, Austria.

[38] Rabbitt P (1993) Does it all go together when it goes? The Nineteenth Bartlett Memorial Lecture. *Q J Exp Psychol* **46**, 385-434.

[39] Tractenberg RE, Pietrzak RH (2011) Intra-individual variability in Alzheimer's disease and cognitive aging: Definitions, context, and effect sizes. *PLoS One* **6**, e16973.

[40] Cole VT, Weinberger DR, Dickinson D (2011) Intra-individual variability across neuropsychological tasks in schizophrenia: A comparison of patients, their siblings, and healthy controls. *Schizophr Res* **129**, 91-93.

[41] Winterer G, Weinberger DR (2004) Genes, dopamine and cortical signal-to-noise ratio in schizophrenia. *Trends Neurosci* **27**, 683-690.

[42] Li S, Aggen SH, Nesselroade JR, Baltes PB (2001) Short-term fluctuations in elderly people's sensorimotor functioning predict text and spatial memory performance: The Macarthur Successful Aging Studies. *Gerontology* **47**, 100-116.

[43] Binder LM, Iverson GL, Brooks BL (2009) To err is human: "Abnormal" neuropsychological scores and variability are common in healthy adults. *Arch Clin Neuropsychol* **24**, 31-46.

[44] Tales A, Wilcock GK, Phillips JE, Bayer A (2014) Is there more to subjective cognitive impairment than meets the eye? Raising awareness. *J Alzheimers Dis* **41**, 665-666.

[45] Tales A, Leonards U, Bompas A, Snowden RJ, Philips M, Porter G, Haworth J, Wilcock G, Bayer A (2012) Intra-individual reaction time variability in amnestic mild cognitive impairment: A precursor to dementia? *J Alzheimers Dis* **32**, 457-466.

Uncorrected Author Proof