Journal of the International Neuropsychological Society (2024), **30**, 689–696
doi:10.1017/S1355617724000213

Research Article

Identification of amnestic mild cognitive impairment among Black and White community-dwelling older adults using NIH Toolbox Cognition tablet battery

Taylor Rigby[1,2,3], Allyson M. Gregoire[1,4], Johnathan Reader[1,4], Yonatan Kahsay[1,3], Jordan Fisher[1,3], Anson Kairys[3], Arijit K. Bhaumik[1,3,4], Annalise Rahman-Filipiak[1,3], Amanda Cook Maher[1,3], Benjamin M. Hampstead[1,3,4], Judith L. Heidebrink[1,4], Voyko Kavcic[5] and Bruno Giordani[1,3]

[1]Michigan Alzheimer's Disease Research Center, Ann Arbor, MI, USA, [2]Department of Veterans Affairs Medical Center, Geriatric Research Education and Clinical Center, Ann Arbor, MI, USA, [3]Department of Psychiatry, University of Michigan, Ann Arbor, MI, USA, [4]Department of Neurology, University of Michigan, Ann Arbor, MI, USA and [5]Institute of Gerontology, Wayne State University, Detroit, MI, USA

Abstract

Objectives: Identify which NIH Toolbox Cognition Battery (NIHTB-CB) subtest(s) best differentiate healthy controls (HC) from those with amnestic mild cognitive impairment (aMCI) and compare the discriminant accuracy between a model using a priori "Norm Adjusted" scores versus "Unadjusted" standard scores with age, sex, race/ethnicity, and education controlled for within the model. Racial differences were also examined. **Methods:** Participants were Black/African American (B/AA) and White consensus-confirmed (HC = 96; aMCI = 62) adults 60–85 years old that completed the NIHTB-CB for tablet. Discriminant function analysis (DFA) was used in the Total Sample and separately for B/AA ($n = 80$) and White participants ($n = 78$). **Results:** Picture Sequence Memory (an episodic memory task) was the highest loading coefficient across all DFA models. When stratified by race, differences were noted in the pattern of the highest loading coefficients within the DFAs. However, the overall discriminant accuracy of the DFA models in identifying HCs and those with aMCI did not differ significantly by race (B/AA, White) or model/score type (Norm Adjusted versus Unadjusted). **Conclusions:** Racial differences were noted despite the use of normalized scores or demographic covariates—highlighting the importance of including underrepresented groups in research. While the models were fairly accurate at identifying consensus-confirmed HCs, the models proved less accurate at identifying White participants with an aMCI diagnosis. In clinical settings, further work is needed to optimize computerized batteries and the use of NIHTB-CB norm adjusted scores is recommended. In research settings, demographically corrected scores or within model correction is suggested.

Keywords: Computerized neuropsychological assessment; computerized cognitive assessment; cognition; psychometrics; norms; cross-cultural; discriminant function analysis

(Received 25 September 2023; final revision 22 March 2024; accepted 1 May 2024)

Currently, it is estimated that 6.5 million persons over the age of 65 years are living with Alzheimer's disease (AD) and related dementias in the USA, and the number is projected to be 12.7 million by the year 2050 (Gaugler et al., 2022). Studies have shown that proactive management of AD and related dementias can improve the quality of life of affected individuals and their caregivers (Grossberg et al., 2010; Vickrey et al., 2006; Voisin & Vellas, 2009). As more treatments are under study or become available for AD, it is increasingly important to identify people at risk for AD and related dementias as early as possible, in part through accurately identifying individuals with mild cognitive impairment (MCI), as well as those with normal cognition.

A diagnosis of MCI refers to cognitive decline that is not normal for a person's age but generally does not affect that person's ability to carry out most activities of daily living (Gauthier et al., 2006). MCI is classified as one of two types based on a person's symptoms: amnestic (memory issues predominate) or non-amnestic (other cognitive issues predominate; Alzheimer's Association, 2022; Petersen et al., 2018; Ward et al., 2013). Though a portion of those diagnosed with MCI may remain stable or revert to preclinical cognition (Petersen et al., 2018), the risk for AD is significantly higher in amnestic MCI versus non-amnestic MCI (Alzheimer's Association, 2022; Kaduszkiewicz et al., 2014; Ward et al., 2013). It is estimated 10%–15% of individuals with MCI go on to develop a form of dementia each year (Alzheimer's Association, 2022) and about 1/3 of people with MCI develop dementia due to AD within five years (Alzheimer's Association, 2022; Ward et al., 2013). Accurate identification and diagnosis of those with MCI is critical for helping individuals, their families, and physicians prepare for future treatment and care.

Corresponding author: T. Rigby; Email: taylor.rigby@va.gov

Cite this article: Rigby T., Gregoire A.M., Reader J., Kahsay Y., Fisher J., Kairys A., Bhaumik A.K., Rahman-Filipiak A., Maher A.C., Hampstead B.M., Heidebrink J.L., Kavcic V., & Giordani B. (2024) Identification of amnestic mild cognitive impairment among Black and White community-dwelling older adults using NIH Toolbox Cognition tablet battery. *Journal of the International Neuropsychological Society*, **30**: 689–696, https://doi.org/10.1017/S1355617724000213

As the older adult population with dementia grows, disparities have emerged in the prevalence of all-cause dementia among different races. Older non-Hispanic Black/African American (B/AA) and Hispanic Americans are disproportionately more likely than older Whites to have AD or other dementias (Dilworth-Anderson et al., 2008; Power et al., 2021; Rosselli et al., 2022; Steenland et al., 2016; Yaffe et al., 2013). There is also evidence that a missed or delayed diagnosis of AD and other dementia types is more common among B/AA and Hispanic older adults than among White older adults (Clark et al., 2005; Fitten et al., 2001; Gianattasio et al., 2019; Lin et al., 2021), which then contributes to a delay of care that may impact disease trajectory and outcomes. Further, despite the increased risk posed to B/AA older adults for developing all-cause dementia, B/AA adults are largely underrepresented in research seeking to understand these diseases (Rosselli et al., 2022).

The NIH Toolbox Cognition Battery (NIHTB-CB) is one module within the larger computerized NIH Toolbox for the Assessment of Neurological and Behavioral Function that was developed as an assessment tool to provide clinical researchers a common metric for cross-study comparisons (Weintraub et al., 2013). The NIHTB-CB was designed to be a brief (30-min), computerized, widely accessible, and easily administered cognitive screener for ages 3–85 that is available in both English and Spanish (Gershon et al., 2013). The battery consists of seven tests measuring five cognitive domains (i.e., executive functioning, episodic memory, processing speed, working memory, language; Weintraub et al., 2013), which are separated broadly into "fluid" or dynamic thinking skills (executive function, episodic memory, processing speed, working memory) and "crystallized" or skills that remain relatively stable in adulthood (language; Heaton et al., 2014). The original NIHTB-CB was unadjusted for demographic factors (Heaton et al., 2014; Weintraub et al., 2013), with subsequent normative samples providing corrections for age, sex, race/ethnicity, and education (Casaletto et al., 2015).

Although the NIHTB-CB may have potential use as a clinical cognitive screener to help identify individuals appropriate for referral for more comprehensive neuropsychological assessment, the utility of the newer computerized NIHTB-CB for tablet has not been well established for clinical characterization. The measures within the NIHTB-CB have demonstrated acceptable reliability and construct validity as compared to traditional paper–pencil methods (Casaletto et al., 2015; Hackett et al., 2018; Heaton et al., 2014; Scott et al., 2019; Weintraub et al., 2013), but the test battery lacks clear support as a standalone replacement for traditional neuropsychological assessment methods (Garcia et al., 2023; Hackett et al., 2018; Scott et al., 2019). Further, despite substantial efforts devoted to the development of representative normative samples for the NIHTB-CB, there remains a scarcity of published studies delineating the performance of underrepresented populations on this cognitive assessment tool. This knowledge gap may inadvertently disregard the potential for performance disparities among distinct racial and ethnic groups.

Aims

In our study, we aimed to determine how well NIHTB-CB tablet subtest scores differentiate those characterized by consensus diagnosis as either healthy controls (HCs) or those with amnestic mild cognitive impairment (aMCI), using National Alzheimer's Coordinating Center (NACC) criteria. We further sought to compare the discriminant ability between a model using "Norm Adjusted" T-scores provided by NIHTB-CB that have been a priori adjusted for age, sex, race/ethnicity, and education (Casaletto et al., 2015) to a second model using NIHTB-CB "Unadjusted" standard scores with the same demographic variables controlled for within the model. No prior predictions were made, as this aim was largely exploratory. We also aimed to examine possible differences between the models when stratified by race (B/AA and White). We hypothesized that there would be no significant differences between the two subsamples, as scores would be either norm adjusted or adjusted in the model for age, sex, race/ethnicity, and education prior to entering the model. We also sought to identify which subtest(s) within the NIHTB-CB for tablet accounted for the largest proportion of difference between HCs and aMCIs. On this point the literature is mixed. Traditionally, Fluid measures have been shown to be particularly sensitive to changes in cognitive status (Heaton et al., 2014; Weintraub et al., 2013). However, studies of older adults in an all B/AA sample (Kairys et al., 2022), a majority B/AA sample (Garcia et al., 2023), and a majority White sample (Hackett et al., 2018) found that crystalized measures on the NIHTB-CB were also important in differentiating those with MCI from HCs.

Methods

Participants

Participants were recruited through the Michigan Alzheimer's Disease Research Center and allied projects. This research was completed in accordance with the Helsinki Declaration. This study was reviewed and approved by the human subjects Institutional Review Board at the University of Michigan Medical School in Ann Arbor, MI, USA. All participants signed consent as per the human subjects University of Michigan Medical School Institutional Review Board in Ann Arbor, MI, USA prior to participation in the study. If the competency of a participant was questionable, a trained study team member administered a decision-making assessment tool to gauge their understanding of the research study and their rights as a participant. For those participants deemed not able to give informed consent, the participant's assent as well as the written informed consent of their legal representative (durable power of attorney, guardian, or next-of-kin as applicable by local laws and regulations) was obtained. All participants then completed the NACC – Unified Data Set (UDS) Version 3 evaluation which included a multidomain medical, neurological, social, and neuropsychological evaluation; participants were then diagnosed at the Michigan Alzheimer's Disease Research Center using NACC consensus conference criteria (Weintraub et al., 2009; Weintraub et al., 2018).

Participants were community-dwelling older adults and were included in analyses if they were between 65 and 85 years of age, classified by consensus diagnosis as either having no clinically significant cognitive impairment/healthy control (HC; $n = 96$) or probable MCI with amnestic features (aMCI; $n = 62$), identified as either B/AA ($n = 80$) or White ($n = 78$), and were not missing any scores. Our sample originally included those classified as non-amnestic MCI ($n = 23$); however, our subsample of non-amnestic MCI was so small that it could not be reasonably included in the analyses and was, therefore, excluded from all analyses. NIHTB-CB assessments using the tablet version for iPad were conducted in English up to 10 days before UDS visits and up to 18 days after UDS assessments with 96.8% of assessments taking place on the same day. NIHTB-CB results were not available to the consensus panel.

https://doi.org/10.1017/S1355617724000213

Table 1. Sample characteristics

	Total Sample	HC	aMCI		B/AA	White	
	N = 158	*n* = 96	*n* = 62	*p-Value*	*n* = 80	*n* = 78	*p-Value*
Education (M/SD)	15.9 (2.6)	16.1 (2.7)	15.6 (2.4)	.25	15.7 (2.6)	16.1 (2.6)	.31
Age (M/SD)	70.2 (6.7)	69.0 (6.4)	72.1 (6.8)	<.01*	69.9 (5.9)	70.4 (7.5)	.60
Female (*n*%)	113 (71.5%)	75 (78.1%)	38 (61.3%)	.03*	64 (80.0%)	49 (62.8%)	.03*
B/AA (*n*%)	80 (50.6%)	43 (44.8%)	37 (59.7%)	.10			
White (*n*%)	78 (49.4%)	53 (55.2%)	25 (40.3%)				
NIH Toolbox Cognition Battery (M/SD)							
Crystalized Abilities							
Picture Vocabulary	112.8 (9.1)	114.4 (9.1)	110.2 (8.5)	<.01*	107.5 (7.3)	118.2 (7.4)	<.01*
Oral Reading Recognition	107.4 (6.9)	108.1 (7.1)	106.3 (6.5)	.09	103.9 (6.8)	111.0 (5.1)	<.01*
Fluid Abilities							
Flanker Inhibitory Control and Attention	90.0 (8.5)	91.8 (7.3)	87.3 (9.5)	<.01*	87.5 (7.9)	92.6 (8.4)	<.01*
Dimensional Change Card Sort	98.6 (8.8)	100.5 (8.2)	95.7 (8.9)	<.01*	97.1 (7.9)	100.1 (9.4)	.03*
List Sorting Working Memory	93.8 (11.9)	97.4 (10.4)	88.2 (12.0)	<.01*	89.5 (11.4)	98.2 (10.8)	<.01*
Pattern Comparison Processing Speed	84.3 (14.5)	86.9 (14.6)	80.2 (13.6)	<.01*	80.3 (12.3)	88.3 (15.5)	<.01*
Picture Sequence Memory	93.5 (12.1)	98.7 (11.3)	85.5 (8.2)	<.01*	90.9 (11.8)	96.3 (11.8)	<.01*

Note: To examine for group differences, independent measures *t*-test was used on continuous variables and chi-square statistic on categorical. Scores used in the NIH Toolbox portion of this table were unadjusted standard scores ($M = 100$, SD = 15) available through NIH Toolbox Cognition for tablet. B/AA = Black/African American; HC = Healthy Controls; aMCI = Mild Cognitive Impairment with amnestic features; *M*/SD = Mean/Standard Deviation.
*Significance $p < .05$.

Assessment measures

National Institutes of Health Toolbox Cognition Battery (NIHTB-CB): The NIHTB-CB is a computerized cognitive screener that takes approximately 30 minutes to administer and is validated to use from ages 3 to 85 (Gershon et al., 2013). Individual subtest performances (Weintraub et al., 2013) as well as composite summary scores for Crystallized, Fluid, and Total Cognition are provided (Heaton et al., 2014). The Crystallized Cognition Composite includes the subtests: Oral Reading Recognition (reading and pronunciation) and Picture Vocabulary (receptive vocabulary). Measures comprising the Fluid Composite include the subtests: Dimensional Change Card Sort (executive function/set-shifting), Flanker Inhibitory Control and Attention (executive function/attention), List Sorting Working Memory (working memory), Pattern Comparison Processing Speed (complex processing speed), and Picture Sequence Memory (episodic memory). Specific test details, procedures, and extensive psychometric evaluation are available elsewhere (Heaton et al., 2014; Weintraub et al., 2013).

Statistical analyses

This study is unique because it provides analyses of two types of NIHTB-CB reported scores. The first are "Norm Adjusted" T-scores ($M = 50$, SD = 10) that have been a priori adjusted for age, sex, race/ethnicity, and education (Casaletto et al., 2015). The second are "Unadjusted" standard scores ($M = 100$, SD = 15) that were standardized using the overall NIHTB-CB norming sample (without regard to demographic variables) and then, for research purposes in this paper, have been adjusted for age, sex, race/ethnicity, and education of our participants through statistical modeling.

Prior to analysis, all measures were screened visually for univariate and multivariate outliers. Univariate outliers were assessed with boxplots, and several were identified. However, only one multivariate outlier was identified among the Norm Adjusted scores based on a chi-square statistic of the Mahalanobis distance with a *p*-value < .001; this outlier was removed from the analytic sample with 158 participants remaining in the final sample. The assumption of multivariate normality was assessed with a multivariate qq plot and was reasonably met. Demographic data and NIHTB-CB subtest performance were examined for group differences using independent measures *t*-test on continuous variables and chi-square on categorical variables (see Table 1). Deeper exploration of the frequency, range, means, and standard deviations of the subtest Picture Sequence Memory was done post hoc to investigate possible difference by diagnosis when stratified by race (see Appendix A).

To determine how well NIHTB-CB tablet subtest scores differentiate those characterized by consensus diagnosis (the "Gold Standard") as either HCs or those with aMCI, a series of discriminant function analyses (DFAs) with leave-one-out cross-validation were performed. All individual NIHTB-CB subtests scores were used in the DFA analyses (summary scores were not included due to overlapping correlations). DFA allows for evaluation of the unique contribution of each variable in rank order. A positive or negative coefficient loading, respectively, increases or decreases the final total score used to discriminate groups. It is, therefore, recommended that the absolute value of the coefficients (see Tables 2 and 3) be used to interpret which tests had more influence over the model outcome (see Table 4). In each model, leave-one-out cross-validation was used to assess model accuracy; meaning that each observation was left out of the model in turn and group membership can be predicted from the loadings of each variable across functions (see Cross-Validation percentage in each model; Table 4).

Separate DFA analyses were run for the Total Sample ($N = 158$) and then separately for B/AA participants ($n = 80$) and White participants ($n = 78$). The same analytic samples were used in the Norm Adjusted score models and the Unadjusted score models to allow for direct comparison of the models by sample (Total Sample ($N = 158$), B/AA ($n = 80$), and White ($n = 78$)). In the DFA models using the Norm Adjusted scores (a priori adjusted for age, sex, race/ethnicity, education), all variables were entered into the models in a single step with no additional control variables included in the models (see Tables 2–4). In the DFA models using Unadjusted scores, the control variables (age, sex, race/ethnicity, and education) were added in a single step concurrent with the

https://doi.org/10.1017/S1355617724000213

Table 2. Discriminant function analyses differentiating those with amnestic mild cognitive impairment from healthy controls in total sample and by race using the norm adjusted NIH Toolbox Cognition for tablet T-scores

	Discriminant function standardized score		
Variable	Total Sample	B/AA	White
Crystalized Abilities			
Picture Vocabulary	−.03	.25	−.34*
Oral Reading Recognition	.09	−.10	.39*
Fluid Abilities			
Flanker Inhibitory Control and Attention	.16	.11	.18
Dimensional Change Card Sort	−.11	−.04	−.22
List Sorting Working Memory	−.45*	−.58*	−.19
Pattern Comparison Processing Speed	−.17	−.28	−.01
Picture Sequence Memory	−.76*	−.72*	−.85*

Note: Coefficients are standardized structure matrix scores that identified the discriminant function. The absolute value of coefficient loadings represents the unique contribution of each measure in rank order. All variables contributed to the model. Scores used were the norm a priori adjusted (age, sex, race/ethnicity, and education) T-scores ($M = 50$, SD = 10) available through NIH Toolbox Cognition for tablet. B/AA = Black/African American.
*Coefficients with an absolute value of at least .30 were interpreted.

Table 3. Discriminant function analyses differentiating those with amnestic mild cognitive impairment from healthy controls in the total sample and by race using the unadjusted NIH toolbox cognition for tablet standard scores

	Discriminant function standardized score		
Variable	Total Sample	B/AA	White
Crystalized Abilities			
Picture Vocabulary	−.23	−.05	−.39*
Oral Reading Recognition	.26	.31*	.24
Fluid Abilities			
Flanker Inhibitory Control and Attention	−.02	.01	−.07
Dimensional Change Card Sort	−.10	.15	−.27
List Sorting Working Memory	−.27	−.26	−.23
Pattern Comparison Processing Speed	−.03	−.09	.07
Picture Sequence Memory	−.74*	−.74*	−.72*
Covariates			
Age	.06	.15	−.02
Sex	−.31*	−.44*	−.19
Education	−.11	−.33*	.23
Race	.02	−−	−−

Note: Coefficients are standardized structure matrix scores that identified the discriminant function. The absolute value of coefficient loadings represents the unique contribution of each measure in rank order. All variables contributed to the model. Scores used were unadjusted standard scores ($M = 100$, SD = 15) available through NIH Toolbox Cognition for tablet, and then correcting for differences within the model using age, sex, race/ethnicity, and education as covariates. B/AA = Black/African American.
*Coefficients with an absolute value of at least .30 were interpreted.

subtest variables (see Tables 3 and 4). A two-sample proportional test was used to determine whether there was a statistically significant ($p < .05$) difference between the proportion of diagnoses correctly differentiated across model type (Norm Adjusted versus Unadjusted) in the Total Sample and when stratified by race (Table 4). A two-sample proportional test was also used to determine whether there was a statistically significant difference ($p < .05$) between the proportion of diagnoses correctly differentiated by race (B/AA versus White) in the Norm Adjusted model and Unadjusted model, respectively (see results section).

Results

Sample characteristics can be found in Table 1. HCs did not significantly differ from those with aMCI in education. Age

Table 4. Discriminant function analysis results by model type

	Norm Adjusted	Unadjusted	*p-Value*
Total Sample			.70
% Correctly Identified	77.2%	79.1%	
% Correctly Identified After Cross-Validation	72.8%	75.3%	
% of HC	84.4%	81.3%	
% of aMCI	54.8%	66.1%	
All B/AA Sample			.86
% Correctly Identified	71.3%	78.8%	
% Correctly Identified After Cross Validation	67.5%	70.0%	
% HC	74.4%	69.8%	
% of aMCI	59.5%	70.3%	
All White Sample			1.00
% Correctly Identified	78.2%	82.1%	
% Correctly Identified After Cross-Validation	71.8%	73.1%	
% HC	83.0%	86.8%	
% of aMCI	48.0%	44.0%	

Note: Norm Adjusted = model using a priori norm adjusted (age, sex, race/ethnicity, and education) T-scores ($M = 50$, SD = 10); Unadjusted = model using unadjusted standard scores ($M = 100$, SD = 15) and, then, correcting for differences within the model using age, sex, race/ethnicity, and education as covariates; *p-value* = a two-sample proportional test was used to determine whether there was a statistically significant difference between the proportion of diagnoses correctly predicted by each model (based on a $p < .05$) between the Norm Adjusted and Unadjusted models; B/AA = Black/African American; HC = Healthy Controls; aMCI = Mild Cognitive Impairment with amnestic features.

differed significantly between HCs and those with aMCI in the Total Sample, with HCs being approximately 3 years younger than those with aMCI on average. There were significantly more female HCs in the total sample than there were females with aMCI. HCs consistently scored higher on NIHTB-CB subtests than those with aMCI when using non-demographically corrected scores. Apart from the crystalized measure Oral Reading Recognition, all differences between HCs and those with aMCI were significant. In a deeper post hoc exploration of the subtest Picture Sequence Memory by diagnosis when stratified by race we found that the range of scores was more restricted in the White sample than in the B/AA when using both Norm Adjusted and Unadjusted scores (see Appendix A). There was also greater score range overlap between the performance of HCs and those with aMCI in the White sample than in the B/AA sample when using both Norm Adjusted and Unadjusted scores (see Appendix A).

The B/AA sample did not significantly differ from the White sample in education or age, and there was no significant difference between the samples on these variables when further broken down into HCs and those with aMCI. However, there were significantly more female participants in the B/AA sample than the White sample. White participants scored significantly higher on all NIHTB-CB subtests than B/AA participants when using non-demographically corrected scores; though, it should be noted that the percentage of those with aMCI was greater in the B/AA sample (59.7%) than in the White sample (40.3%).

Table 2 shows the individual contribution of tests within the NIHTB-CB using the Norm Adjusted scores (a priori adjusted for age, sex, race/ethnicity, and education) in the Total Sample and stratified by race. Picture Sequence Memory was the subtest that accounted for the largest proportion of the between-group difference in the Total Sample and B/AA sample, followed by List Sorting Working Memory. In the White sample, Picture Sequence Memory was again the coefficient with the highest loading followed by Oral Reading Recognition and Picture Vocabulary.

https://doi.org/10.1017/S1355617724000213

DFAs that controlled for age, sex, race/ethnicity, and education within the models using the Unadjusted NIHTB-CB standard scores can be seen in Table 3. This table shows the individual contribution of NIHTB-CB subtests in the Total Sample and stratified by race. In the Total Sample analysis, Picture Sequence Memory was the subtest that accounted for the largest proportion of the between-group difference followed by sex. In the B/AA sample, Picture Sequence Memory was again the coefficient with the highest loading followed by sex, education, and Oral Reading Recognition, respectively. In the White sample, Picture Sequence Memory was the coefficient with the highest loading followed by Picture Vocabulary.

Table 4 shows a side-by-side comparison of DFA results by model type (Norm Adjusted versus Unadjusted) in the Total Sample and stratified by race (B/AA; White). The Norm Adjusted and Unadjusted DFAs did not significantly differ in their ability to discriminate consensus-confirmed HCs from those with a consensus diagnosis of aMCI across samples (Total Sample, B/AA, White). Further, neither the Norm Adjusted ($p = .68$) nor Unadjusted models ($p = .80$) significantly differed in their overall discriminant ability when comparing model type by race. Though the overall accuracy of the models did not significantly differ by race, the White sample models were notably worse at identifying those with aMCI than the B/AA sample models. This finding indicates that in the White sample the ability to discriminate between the two diagnoses (HC versus aMCI) is relatively weak in both the Fully Adjusted and Unadjusted models.

Discussion

We found that the DFA models using Norm Adjusted scores produced a similar pattern of results to the models using Unadjusted scores that were corrected within the analyses for age, sex, race/ethnicity, and education. Both White sample models were notably worse at identifying those with aMCI than the B/AA sample models. To explore this difference, we reexamined frequency plots, ranges, means, and the standard deviations of the subtest that accounted for the largest proportion of the between-group difference by sample type in each model—Picture Sequence Memory. We found that the White sample had a restricted range of scores when compared to the B/AA sample and there was greater score range overlap between HCs and those with aMCI in the White sample than in the B/AA sample. Thus, the coefficient loadings in the White models are describing a small between-group difference that is being reflected in the relatively low discriminant accuracy of the models. Similar to our findings in the White sample, a study found that the NIHTB-CB yielded a moderate level of specificity but demonstrated chance level sensitivity to subtle cognitive impairment in a racially diverse sample (Buckley et al., 2017).

Picture Sequence Memory (a test of episodic memory) was the subtest that accounted for the largest proportion of between-group difference in each model. This finding is unsurprising as episodic memory is often one of the earliest cognitive domains to decline in population-based studies of preclinical Alzheimer's dementia, and measures of episodic memory are one of the best predictors of progression from MCI to Alzheimer's dementia in most longitudinal studies (Bastin & Salmon, 2014). Other studies of the NIHTB-CB have found Picture Sequence Memory to be an important contributor to differentiating HCs from those with MCI (Garcia et al., 2023; Kairys et al., 2022) and differentiating those with aMCI from those with non-amnestic MCI (Garcia et al., 2023). However, a study conducted in a memory clinic setting found that many cognitively impaired participants had difficulty completing the task Picture Sequence Memory and it was ultimately dropped from the analyses and replaced by another memory task (Rey Auditory Verbal Learning Test; Hackett et al., 2018). Thus, while memory tasks are important in differentiating those with a memory deficit from those without, Picture Sequence Memory may be too difficult to use as a standalone screener in a clinical setting.

When stratifying by race using NIHTB-CB standard scores, Oral Reading Recognition (a single word reading task) was among the largest unique contributors accounting for between-group difference in the B/AA sample, while Picture Vocabulary (a receptive language ability task) was in the White sample. Similarly, a recent study with a majority B/AA older adult sample found Oral Reading Recognition to be one of the strongest unique contributors in identifying HCs from those with MCI (both amnestic and non-amnestic MCI) when using the NIHTB-CB standard scores for tablet (Garcia et al., 2023). Another study using NIHTB-CB with an all B/AA older adult sample found that Oral Reading Recognition and Picture Vocabulary best differentiated HCs from those with aMCI when using NIHTB-CB Norm Adjusted scores for laptop (Kairys et al., 2022). In our sample, we found, when using the NIHTB-CB Norm Adjusted scores for tablet, that crystalized abilities were among the strongest contributors to the model identifying HCs from those with aMCI in the White sample but not in the B/AA sample. These findings are interesting, as both receptive language ability tasks and single word reading tasks are thought to generally remain stable across the lifespan or become "crystallized" and are, thus, thought to provide a reasonably accurate estimate of premorbid cognitive functioning in older adults with and without neurodegenerative disease (Grober et al., 1991; Snitz et al., 2000). However, individuals with MCI have been shown to perform worse than HCs on a receptive language task (Peabody Picture Vocabulary Test; Jokel et al., 2019) and those with AD have been shown to perform worse than those with MCI on that same task (Snitz et al., 2000). Thus, decline in language ability may be more important than is generally acknowledged for identifying those declining cognitively. However, neither Oral Reading Recognition nor Picture Vocabulary were strong discriminating variables in the Unadjusted or Norm Adjusted models when the total (race combined) sample was used; this was due to the opposing coefficient loadings (i.e., negative versus positive) between the B/AA and White samples. Because this effect washed away once the race stratified samples were unified, it is important to consider how combining racial samples can sometimes obscure important differences—lending further support to the call to conduct more studies that focus on understudied minority populations and cross-study comparisons of factor invariance.

When looking at the models using unadjusted NIHTB-CB scores, sex was a strong discriminating variable in the Total Sample and in the B/AA models, but not in the White models. Sex effects have been consistently shown across studies with differences typically favoring males in visuospatial tasks specifically and favoring females more broadly but particularly in episodic memory tasks (Heaton, 2004; Lippa et al., 2010). However, there were significantly more females in the HC group than in the aMCI group and there were more female HCs in the B/AA sample than in the White. Thus, it is difficult to parse apart whether the relatively high coefficient loading is related to sex effects, or the uneven distribution of sex among variables (i.e., HC versus aMCI).

https://doi.org/10.1017/S1355617724000213

Similarly, the B/AA group differences are likely driving the sex effects seen in the Total Sample model using Unadjusted scores.

Despite there being no significant difference in average level of education between participants when stratified by race, we found that White participants scored significantly higher on average than B/AA participants on all NIHTB-CB subtests. This finding is consistent with a previous finding using NIHTB-CB (Casaletto et al., 2015) and a finding using traditional neuropsychological measures (Manly et al., 1998). Further, education was among the largest unique contributors in identifying HCs from those with MCI in the B/AA sample in the model using Unadjusted NIHTB-CB scores. Conversely, education was not a strong discriminating variable in the Total Sample model or the White sample model using Unadjusted NIHTB-CB scores. Together, these results reiterate the meaningful and complicated impact of demographic factors on assessment outcomes. Historically, neuropsychologists have sought to address cultural disparities by making demographically adjusted norms. However, such norms are not able to account for all sociocultural or individual factors (Byrd & Rivera-Mindt, 2022; Manly, 2005; Rosselli et al., 2022). Nevertheless, when such norms are used appropriately, they offer greater diagnostic accuracy (Manly, 2005; Werry et al., 2019) and have been recently shown to reduce the association between education and cognitive performance in a racially diverse sample of older adults (Mungas et al., 2021). The NIHTB-CB norms specifically have been shown to successfully reduce the impact of demographic variables on performance (Casaletto et al., 2015); thus, use of the norm corrected scores provided by NIHTB-CB or within model demographic correction is suggested.

Limitations and future directions

One limitation of this study is that we were not able to include older adults over the age of 85 due to a lack of norms in the field for those above the age of 85. Other studies such as the Advancing Reliable Measurement in Alzheimer's Disease and Cognitive Aging Study are seeking to address this problem by extending the NIHTB norms to those over the age of 85 (Weintraub et al., 2022). HCs in our sample were younger by 3 years on average than those diagnosed with aMCI. This finding is unsurprising as increasing age in adults is consistently and strongly associated with poorer performances (Heaton, 2004) as well as a higher risk of cognitive impairment associated with MCI and all-cause dementia (Alzheimer's Association, 2022; Gaugler et al., 2022). Those that participated in this study with a consensus diagnosis of MCI all had amnestic features. Future studies may attempt to recruit larger numbers of those with a diagnosis of non-amnestic MCI to examine potential differences. There was an uneven distribution of sex among HCs and those with aMCI in our study sample, as well as differences in distribution of sex by race. Future studies may aim to recruit an equal number of males and females, though this continues to be an issue in many ongoing longitudinal studies of cognitive change. Differences were found between models when stratified by race. For example, although both models using the B/AA sample were more accurate at identifying those with aMCI than the models using the White sample, it should be noted that the B/AA participants were better balanced in terms of the number of aMCI versus HCs, and the White sample demonstrated a restricted range with small between-group difference on the strongest discriminating variable (Picture Sequence Memory). It is likely that these differences also impacted the results of the combined sample model. Though, the NIHTB-CB norms have been shown to successfully reduce the impact of demographic variables on performance (Casaletto et al., 2015), it is unclear to what degree the differences between racial groups in our study are due to differences in sample distribution and demographic factors. Future DFA studies should aim to recruit equal numbers of racial groups by diagnosis. Further, because the between-group difference of HCs versus those with aMCI appears small, future studies should aim to recruit larger sample sizes.

Conclusions

In our study we compared two types of methods using scores available through NIHTB-CB: a priori adjusted (age, sex, race/ethnicity, and education) T-scores and unadjusted standard scores that were corrected for age, sex, race/ethnicity, and education within the analyses. Our findings indicate that either method can be used to produce similar results when identifying HCs and those with aMCI. However, despite the use of normalized scores or demographic covariates, differences between models when stratified by race were noted—emphasizing the need to continue efforts to include underrepresented groups in research seeking to understand AD and other dementia types. We found that the White sample models were less successful at identifying those with aMCI across model type than the B/AA sample. The use of the norm corrected scores is strongly recommended for use in a clinical setting, and norm corrected scores or within model demographic correction are suggested when using NIHTB-CB in research settings.

Our findings do not provide clear support for use of the NIHTB-CB as a standalone screener in a clinical setting; however, NIHTB-CB has relative ease and efficiency of administration when compared to traditional neuropsychological methods and has been shown to perform as well as traditional neuropsychiatric test batteries at classifying aMCI in a clinical setting (Hackett et al., 2018). Though, further work is needed to optimize computerized batteries for use in clinical settings. For example, there were subtests in the NIHTB-CB that provided comparatively little information in differentiating HCs from those with aMCI, and some tasks may be too difficult for cognitively impaired individuals to complete reliably (Hackett et al., 2018). Further, pairing the complete or partial NIHTB-CB (a measure with specificity) with a more sensitive measure like the Montreal Cognitive Assessment may be helpful (Larner, 2019; Nasreddine et al., 2005), but more research is needed. In research settings, it will be important to consider the aims and constraints of the study when determining whether to use NIHTB-CB (e.g., ease and efficiency of administration, minimizing false negatives/sensitivity versus minimizing false positives/specificity).

Supplementary material. The supplementary material for this article can be found at https://doi.org/10.1017/S1355617724000213.

Acknowledgments. We would like to thank Dr Ana Daugherty, Assistant Professor at the Wayne State Institute of Gerontology and Department of Psychology, and Jian Kang, Professor of Biostatistics at the University of Michigan School of Public Health, for conferring with us regarding the statistical modeling included in this paper. We would also like to thank Dr Subhamoy Pal, Statistician at the Michigan Alzheimer's Disease Center, for his helpful input.

Funding statement. This work was supported by the US National Institute on Aging (P30 AG053760, P30 AG072931) and the US National Institute of Health (V.K., R01 AG054484, R21 AG046637; B.H., RO1 AG058724; U2C AG057441).

Competing interests. None.

References

Alzheimer's Association (2022). *More than normal aging: Understanding mild cognitive impairment*. Alzheimer's Disease Facts and Figures. https://www.alz.org/media/Documents/alzheimers-facts-and-figures-special-report-2022.pdf.

Bastin, C., & Salmon, E. (2014). Early neuropsychological detection of Alzheimer's disease. *European Journal of Clinical Nutrition*, *68*(11), 1192–1199. https://doi.org/10.1038/ejcn.2014.176

Buckley, R. F., Sparks, K. P., Papp, K. V., Dekhtyar, M., Martin, C., Burnham, S., Sperling, R. A., & Rentz, D. M. (2017). Computerized cognitive testing for use in clinical trials: A comparison of the NIH toolbox and cogstate C3 batteries. *The Journal of Prevention of Alzheimer's Disease*, *4*(1), 3–9. https://doi.org/10.14283/jpad.2017.1

Byrd, D. A., & Rivera-Mindt, M. G. (2022). Neuropsychology's race problem does not begin or end with demographically adjusted norms. *Nature Reviews Neurology*, *18*(3), 125–126. https://doi.org/10.1038/s41582-021-00607-4

Casaletto, K. B., Umlauf, A., Beaumont, J., Gershon, R., Slotkin, J., Akshoomoff, N., & Heaton, R. K. (2015). Demographically corrected normative standards for the English version of the NIH toolbox cognition battery. *Journal of the International Neuropsychological Society*, *21*(5), 378–391. https://doi.org/10.1017/S1355617715000351

Clark, P. C., Kutner, N. G., Goldstein, F. C., Peterson-Hazen, S., Garner, V., Zhang, R., & Bowles, T. (2005). Impediments to timely diagnosis of Alzheimer's disease in African Americans. *Journal of the American Geriatrics Society*, *53*(11), 2012–2017. https://doi.org/10.1111/j.1532-5415.2005.53569.x

Dilworth-Anderson, P., Hendrie, H. C., Manly, J. J., Khachaturian, A. S., & Fazio, S. (2008). Diagnosis and assessment of Alzheimer's disease in diverse populations. *Alzheimer's & Dementia*, *4*(4), 305–309. https://doi.org/10.1016/j.jalz.2008.03.001

Fitten, L. J., Ortiz, F., & Pontón, M. (2001). Frequency of Alzheimer's disease and other dementias in a community outreach sample of hispanics. *Journal of the American Geriatrics Society*, *49*(10), 1301–1308. https://doi.org/10.1046/j.1532-5415.2001.49257.x

Garcia, S., Askew, R. L., Kavcic, V., Shair, S., Bhaumik, A. K., Rose, E., Campbell, S., May, N., Hampstead, B. M., Dodge, H. H., Heidebrink, J. L., Paulson, H. L., & Giordani, B. (2023). Mild cognitive impairment subtype performance in comparison to healthy older controls on the NIH toolbox and cogstate. *Alzheimer Disease & Associated Disorders*, *37*(4), 328–334. https://doi.org/10.1097/WAD.0000000000000587

Gaugler, J., James, B., Johnson, T., Reimer, J., Solis, M., Weuve, J., & Hohman, T. J. (2022). Alzheimer's disease facts and figures. *Alzheimers & Dementia*, *18*(4), 700–789. https://doi.org/10.1002/alz.12638

Gauthier, S., Reisberg, B., Zaudig, M., Petersen, R. C., Ritchie, K., Broich, K., Belleville, S., Brodaty, H., Bennett, D., Chertkow, H., Cummings, J. L., de Leon, M., Feldman, H., Ganguli, M., Hampel, H., Scheltens, P., Tierney, M. C., Whitehouse, P., & Winblad, B. (2006). Mild cognitive impairment. *The lancet*, *367*(9518), 1262–1270. https://doi.org/10.1016/S0140-6736(06)68542-5

Gershon, R. C., Wagster, M. V., Hendrie, H. C., Fox, N. A., Cook, K. F., & Nowinski, C. J. (2013). NIH toolbox for assessment of neurological and behavioral function. *Neurology*, *80*(11_supplement_3), S2–S6. https://doi.org/10.1212/WNL.0b013e3182872e5f

Gianattasio, K. Z., Prather, C., Glymour, M. M., Ciarleglio, A., & Power, M. C. (2019). Racial disparities and temporal trends in dementia misdiagnosis risk in the United States. *Alzheimer's & Dementia: Translational Research & Clinical Interventions*, *5*(1), 891–898. https://doi.org/10.1016/j.trci.2019.11.008

Grober, E., Sliwinsk, M., & Korey, S. R. (1991). Development and validation of a model for estimating premorbid verbal intelligence in the elderly. *Journal of Clinical and Experimental Neuropsychology*, *13*(6), 933–949. https://doi.org/10.1080/01688639108405109

Grossberg, G. T., Christensen, D. D., Griffith, P. A., Kerwin, D. R., Hunt, G., & Hall, E. J. (2010). The art of sharing the diagnosis and management of Alzheimer's disease with patients and caregivers: Recommendations of an expert consensus panel. *The Primary Care Companion for CNS Disorders*, *12*(1), 26659. https://doi.org/10.4088/PCC.09cs00833oli

Hackett, K., Krikorian, R., Giovannetti, T., Melendez-Cabrero, J., Rahman, A., Caesar, E. E., Chen, J. L., Hristov, H., Seifan, A., Mosconi, L., & Isaacson, R. S. (2018). Utility of the NIH toolbox for assessment of prodromal Alzheimer's disease and dementia. *Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring*, *10*(1), 764–772. https://doi.org/10.1016/j.dadm.2018.10.002

Heaton, R. K. (2004). *Revised comprehensive norms for an expanded halstead-reitan battery: Demographically adjusted neuropsychological norms for African American and Caucasian adults, professional manual*. Psychological Assessment Resources.

Heaton, R. K., Akshoomoff, N., Tulsky, D., Mungas, D., Weintraub, S., Dikmen, S., Beaumont, J., Casaletto, K. B., Conway, K., Slotkin, J., & Gershon, R. (2014). Reliability and validity of composite scores from the NIH toolbox cognition battery in adults. *Journal of the International Neuropsychological Society*, *20*(6), 588–598. https://doi.org/10.1017/S1355617714000241

Jokel, R., Seixas Lima, B., Fernandez, A., & Murphy, K. J. (2019). Language in amnestic mild cognitive impairment and dementia of Alzheimer's type: Quantitatively or qualitatively different? *Dementia and Geriatric Cognitive Disorders Extra*, *9*(1), 136–151. https://doi.org/10.1159/000496824

Kaduszkiewicz, H., Eisele, M., Wiese, B., Prokein, J., Luppa, M., Luck, T., Jessen, F., Bickel, H., Mosch, E., Pentzek, M., Fuchs, A., Eifflaender-Gorfer, S., Weyerer, S., Konig, H.-H., Brettschneider, C., van den Bussche, H., Maier, W., Scherer, M., Riedel-Heller, S. G., & The Study on Aging, Cognition, and Dementia in Primary Care Patients (AgeCoDe) Study Group (2014). Prognosis of mild cognitive impairment in general practice: Results of the German ageCoDe study. *The Annals of Family Medicine*, *12*(2), 158–165. https://doi.org/10.1370/afm.1596

Kairys, A., Daugherty, A., Kavcic, V., Shair, S., Persad, C., Heidebrink, J., Bhaumik, A., & Giordani, B. (2022). Laptop-administered NIH toolbox and cogstate brief battery in community-dwelling black adults: Unexpected pattern of cognitive performance between MCI and healthy controls. *Journal of the International Neuropsychological Society*, *28*(3), 239–248. https://doi.org/10.1017/S135561772100028X

Larner, A. J. (2019). Evaluating cognitive screening instruments with the, likelihood to be diagnosed or misdiagnosed, measure. *International Journal of Clinical Practice*, *73*(2), e13265. https://doi.org/10.1111/ijcp.13265

Lin, P-J., Daly, A. T., Olchanski, N., Cohen, J. T., Neumann, P. J., Faul, J. D., Fillit, H. M., & Freund, K. M. (2021). Dementia diagnosis disparities by race and ethnicity. *Medical care*, *59*(8), 679–686. https://doi.org/10.1097/MLR.0000000000001577

Lippa, R. A., Collaer, M. L., & Peters, M. (2010). Sex differences in mental rotation and line angle judgments are positively associated with gender equality and economic development across 53 nations. *Archives of sexual behavior*, *39*(4), 990–997. https://doi.org/10.1007/s10508-008-9460-8

Manly, J. J. (2005). Advantages and disadvantages of separate norms for African Americans. *The Clinical Neuropsychologist*, *19*(2), 270–275. https://doi.org/10.1080/13854040590945346

Manly, J. J., Miller, S. W., Heaton, R. K., Byrd, D., Reilly, J. U. D. Y., Velasquez, R. J., Saccuzzo, D. P., Grant, I., & HIV Neurobehavioral Research Center (HNRC) Group (1998). The effect of African-American acculturation on neuropsychological test performance in normal and HIV-positive individuals. *Journal of the International Neuropsychological Society*, *4*(3), 291–. https://doi.org/302.10.1017/S1355617798002914

Mungas, D., Shaw, C., Hayes-Larson, E., DeCarli, C., Farias, S. T., Olichney, J., & Saucedo, H. H., Gilsanz, P., Glymour, M. M., Whitmer, R. A., & Mayeda, E. R. (2021). Cognitive impairment in racially/ethnically diverse older adults: Accounting for sources of diagnostic bias. *Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring*, *13*(1), e12265. https://doi.org/10.1002/dad2.12265

Nasreddine, Z. S., Phillips, N. A., Bédirian, Vérie, Charbonneau, S., Whitehead, V., Collin, I., Cummings, J. L., & Chertkow, H. (2005). The montreal cognitive assessment, MoCA: A brief screening tool for mild cognitive impairment. *Journal of the American Geriatrics Society*, *53*(4), 695–699. https://doi.org/10.1111/j.1532-5415.2005.53221.x

Petersen, R. C., Lopez, O., Armstrong, M. J., Getchius, T. S. D., Ganguli, M., Gloss, D., Gronseth, G. S., Marson, D., Pringsheim, T., Day, G. S., Sager, M.,

Stevens, J., & Rae-Grant, A. (2018). Practice guideline update summary: Mild cognitive impairment: Report of the guideline development, dissemination, and implementation subcommittee of the American academy of neurology. *Neurology*, *90*(3), 126–135. https://doi.org/10.1212/WNL.0000000000004826

Power, M. C., Bennett, E. E., Turner, R. W., Dowling, N. M., Ciarleglio, A., Glymour, M. M., & Gianattasio, K. Z. (2021). Trends in relative incidence and prevalence of dementia across non-hispanic black and white individuals in the United States, 2000-2016. *JAMA neurology*, *78*(3), 275–284. https://doi.org/10.1001/jamaneurol.2020.4471

Rosselli, M., Uribe, I. V., Ahne, E., & Shihadeh, L. (2022). Culture, ethnicity, and level of education in Alzheimer's disease. *Neurotherapeutics*, *19*(1), 26–54. https://doi.org/10.1007/s13311-022-01193-z

Scott, E. P., Sorrell, A., & Benitez, A. (2019). Psychometric properties of the NIH toolbox cognition battery in healthy older adults: Reliability, validity, and agreement with standard neuropsychological tests. *Journal of the International Neuropsychological Society*, *25*(8), 857–867. https://doi.org/10.1017/S1355617719000614

Snitz, B. E., Bieliauskas, L. A., Crossland, A., Basso, M. R., & Roper, B. (2000). PPVT-R as an estimate of premorbid intelligence in older adults. *The Clinical Neuropsychologist*, *14*(2), 181–186.

Steenland, K., Goldstein, F. C., Levey, A., & Wharton, W. (2016). A meta-analysis of Alzheimer's disease incidence and prevalence comparing African-Americans and Caucasians. *Journal of Alzheimer's Disease*, *50*(1), 71–76. https://doi.org/10.3233/JAD-150778

Vickrey, B. G., Mittman, B. S., Connor, K. I., Pearson, M. L., Della Penna, R. D., Ganiats, T. G., R. W. DeMonte, Chodosh, J., Cui, X., Vassar, S., Duan, N., & Lee, M. (2006). The effect of a disease management intervention on quality and outcomes of dementia care: A randomized, controlled trial. *Annals of internal medicine*, *145*(10), 713–726. https://doi.org/10.7326/0003-4819-145-10-200611210-00004

Voisin, T., & Vellas, B. (2009). Diagnosis and treatment of patients with severe Alzheimer's disease. *Drugs & aging*, *26*(2), 135–144. https://doi.org/10.2165/0002512-200926020-00005

Ward, A., Tardiff, S., Dye, C., & Arrighi, H. M. (2013). Rate of conversion from prodromal Alzheimer's disease to Alzheimer's dementia: A systematic review of the literature. *Dementia and Geriatric Cognitive Disorders Extra*, *3*(1), 320–332. https://doi.org/10.1159/000354370

Weintraub, S., Besser, L., Dodge, H. H., Teylan, M., Ferris, S., Goldstein, F. C., Giordani, B., Kramer, J., Loewenstein, D., Marson, D., Mungas, D., Salmon, D., Welsh-Bohmer, K., Zhou, X.-H., Shirk, S. D., Atri, A., Kukull, W. A., Phelps, C., & Morris, J. C. (2018). Version 3 of the Alzheimer disease centers' neuropsychological test battery in the uniform data set (UDS). *Alzheimer Disease & Associated Disorders*, *32*(1), 10–17. https://doi.org/10.1097/WAD.0000000000000223

Weintraub, S., Dikmen, S. S., Heaton, R. K., Tulsky, D. S., Zelazo, P. D., Bauer, P. J., Carlozzi, N. E., Slotkin, J., Blitz, D., Wallner-Allen, K., Fox, N. A., Beaumont, J. L., Mungas, D., Nowinski, C. J., Richler, J., Deocampo, J. A., Anderson, J. E., Manly, J. J., Borosh, B., Havlik, R., Conway, K., Edwards, E., Freund, L., King, J. W., Moy, C., Witt, E., & Gershon, R. C. (2013). Cognition assessment using the NIH toolbox. *Neurology*, *80*(11 Supplement 3), S54–S64. https://doi.org/10.1212/WNL.0b013e3182872ded

Weintraub, S., Karpouzian-Rogers, T., Peipert, J. D., Nowinski, C., Slotkin, J., Wortman, K., Ho, E., Rogalski, E., Carlsson, C., Giordani, B., Goldstein, F., Lucas, J., Manly, J. J., Rentz, D., Salmon, D., Snitz, B., Dodge, H. H., Riley, M., Eldes, F., Ustsinovich, V., & Gershon, R. (2022). ARMADA: Assessing reliable measurement in Alzheimer's disease and cognitive aging project methods. *Alzheimer's & Dementia*, *18*(8), 1449–1460. https://doi.org/10.1002/alz.12497

Weintraub, S., Salmon, D., Mercaldo, N., Ferris, S., Graff-Radford, N. R., Chui, H., Cummings, J., DeCarli, C., Foster, N. L., Galasko, D., Peskind, E., Dietrich, W., Beekly, D. L., Kukull, W. A., & Morris, J. C. (2009). The Alzheimer's disease centers' uniform data set (UDS): The neuropsychological test battery. *Alzheimer Disease & Associated Disorders*, *23*(2), 91–101. https://doi.org/10.1097/WAD.0b013e318191c7dd

Werry, A. E., Daniel, M., & Bergström, B. (2019). Group differences in normal neuropsychological test performance for older non-hispanic white and black/African American adults. *Neuropsychology*, *33*(8), 1089–1100. https://doi.org/10.1037/neu0000579

Yaffe, K., Falvey, C., Harris, T. B., Newman, A., Satterfield, S., Koster, A., Ayonayon, H., Simonsick, E., & for the Health ABC Study (2013). Effect of socioeconomic disparities on incidence of dementia among biracial older adults: Prospective study. *BMJ*, *347*(dec19 5), f7051. https://doi.org/10.1136/bmj.f7051