



Research Article

Validation of the NIH Toolbox-Cognition Battery against legacy neurocognitive measures in adults with cognitive impairments: An exploratory analysis

Emily A. Kringle¹ , Enrico M. Novelli², Meryl A. Butters³ and Elizabeth R. Skidmore⁴

¹University of Illinois at Chicago, Department of Medicine, Chicago, USA, ²University of Pittsburgh, Department of Medicine, Pittsburgh, USA, ³University of Pittsburgh, Department of Psychiatry, Pittsburgh, USA and ⁴University of Pittsburgh, Department of Occupational Therapy, Pittsburgh, USA

Abstract

Objective: The purpose of this exploratory study was to describe associations between NIH Toolbox-Cognition Battery subtests and legacy measures of neurocognitive function in two samples with neurological conditions (stroke and sickle cell disease (SCD)). **Method:** This exploratory secondary analysis uses data from two studies that assessed cognition at one time point using the NIH Toolbox-Cognition Battery, the Repeatable Battery for the Assessment of Neuropsychological Status (RBANS), and subtests from the Delis-Kaplan Executive Functions System (DKEFS). People with stroke ($n = 26$) and SCD ($n = 64$) were included. Associations between the NIH Toolbox-Cognition Battery subtests and corresponding legacy measures were examined using linear correlations, Bland–Altman analysis, and Lin's Concordance Correlation Coefficient. **Results:** Linear correlations and Lin's Concordance Correlation Coefficient were poor to strong in both samples on NIH Toolbox-CB subtests: Flanker Inhibitory Control and Attention ($r = .35$ to $.48$, Lin CCC = $.27$ to $.37$), Pattern Comparison Processing Speed ($r = .40$ to $.65$, Lin CCC = $.37$ to $.62$), Picture Sequence Memory ($r = .19$ to $.55$, Lin CCC = $.18$ to $.48$), Dimensional Change Card Sort ($r = .39$ to $.77$, Lin CCC = $.38$ to $.63$), Fluid Cognition Composite ($r = .88$ to $.90$, Lin CCC = $.60$ to $.79$), and Total Cognition Composite ($r = .64$ to $.83$, Lin CCC = $.60$ to $.78$). Bland–Altman analyses demonstrated wide limits of agreement across all subtests (-3.17 to 3.78). **Conclusions:** The NIH Toolbox-Cognition Battery subtests may behave similarly to legacy measures as an overall assessment of cognition across samples at risk for neurological impairment. Findings should be replicated across additional clinical samples.

Keywords: cognition; memory; perception; language; comprehension; executive function; anemia; sickle cell; stroke

(Received 23 August 2021; final revision 12 May 2022; accepted 18 May 2022; First Published online 5 September 2022)

Introduction

Cognition is commonly measured using time-intensive paper and pencil batteries such as the Repeatable Battery for the Assessment of Neuropsychological Status (RBANS) or the Delis-Kaplan Executive Functions System (DKEFS). However, these legacy measures can be difficult to apply in epidemiological studies. Epidemiological studies that include adults with cognitive impairment based on the behavioral phenotype, rather than underlying etiology, require assessments that can be administered efficiently and interpreted across clinical conditions. The National Institutes of Health Toolbox Cognition Battery (NIH Toolbox-CB) aims to balance brevity with precision for accurate measurement of cognitive function (Weintraub et al., 2013).

The NIH Toolbox-CB is a performance-based measure that assesses attention, memory, language, and executive functions. Items are administered using an electronic tablet and completed within 30 min. NIH Toolbox-CB items demonstrated convergent validity with legacy neuropsychological measures of attention, memory, language, and executive functions across the lifespan in healthy, racially and ethnically diverse samples (Carlozzi et al., 2015;

Dikmen et al., 2014; Gershon et al., 2014; Mungas et al., 2014; Tulskey et al., 2014; Zelazo et al., 2014). However, validation in healthy samples assumes variance that cannot be assumed to translate to specific clinical populations (Delis et al., 2003). Precision of the NIH Toolbox-CB among cognitively impaired samples with diverse etiologies remains unclear. Prior studies that employed the NIH Toolbox-CB among samples with neurological conditions described the ability to distinguish among samples based on expected impairments (e.g., distinguish among stroke, traumatic brain injury, spinal cord injury) (Carlozzi et al., 2017a). Within homogenous populations, the NIH Toolbox-CB classified severity of impairments categorically (Carlozzi et al., 2017b; Hackett et al., 2018). However, the association between continuous scores on NIH Toolbox-CB and scores on legacy measures of cognition among samples at elevated risk for cognitive impairment remain unclear. The present analysis is a step toward this.

Adults with chronic stroke and those with sickle cell disease (SCD) are examples of two populations whose behavioral phenotype may include cognitive impairments. Stroke affects 80 million adults globally, and among these, 32–43% experience persistent impairments in executive functions, memory, attention, processing

Corresponding author: Emily A. Kringle, email: kringle@uic.edu

Cite this article: Kringle E.A., Novelli E.M., Butters M.A., & Skidmore E.R. (2023) Validation of the NIH Toolbox-Cognition Battery against legacy neurocognitive measures in adults with cognitive impairments: An exploratory analysis. *Journal of the International Neuropsychological Society*, 29: 472–479, <https://doi.org/10.1017/S1355617722000406>

speed, or language (Sexton et al., 2019). These impairments are attributed to lesions in the brain caused by acute adult-onset vascular ischemia or hemorrhage (Zhao et al., 2018). SCD is an inherited blood disorder that causes diffuse organ damage, including extensive central nervous system pathology. People with SCD sustain silent cardiovascular events beginning in childhood. These events, including transient ischemic attacks, ischemic strokes, seizures, and high intracranial pressure, are not consistently detected clinically, and result in subtle cognitive impairments (Edwards et al., 2007). Further, silent cardiovascular events and chronic hypoxia negatively influence cognitive development resulting in impairments in attention, memory, processing speed, and executive functions that persist in adulthood (Jorgensen et al., 2017; Kirkham & Datta, 2006). Among adults with SCD, 36% score 1 standard deviation and 5% score 2 standard deviations below the population mean in the Processing Speed Index of the Wechsler Adult Intelligence Scale III (Vichinsky et al., 2010). Furthermore, approximately 50% of adults with SCD have silent cerebral infarcts which are associated with cognitive impairments (Kassim et al., 2016). Abnormalities in the basal ganglia, thalamus, and white matter integrity have been associated with neurocognitive impairments among adults with SCD (Mackin et al., 2014; Vichinsky et al., 2010). While the underlying etiology of neurocognitive impairment varies between stroke and SCD, both groups experience poor outcomes in daily living associated with cognitive impairments (Mole & Demeyere, 2018; Sanger et al., 2016).

People with stroke and SCD represent clinical populations that are at risk for cognitive impairment. Neurocognitive assessment can be particularly challenging in these populations due to the presence of fatigue and pain (Chakravorty & Williams, 2015; Duncan et al., 2012). Simultaneously, neurocognitive assessment is important in these populations who are at high risk for cognitive impairments. The NIH Toolbox-CB may be particularly advantageous because of the brevity of the test, combined with the opportunity to obtain a full cognitive profile. It is currently unclear if continuous scores on the NIH Toolbox-CB behave similarly across distinct samples with cognitive impairment. Thus, the aim of this secondary analysis was to explore associations between continuous scores on the NIH Toolbox-CB and legacy measures of cognition in two clinical populations with neurocognitive impairments of varied underlying etiology. We expected to observe significant associations between corresponding measures to the extent that the different subtests tap the same underlying cognitive process. That is, that subtests which measure similar underlying processes would have strong associations.

Methods

This is an exploratory secondary analysis of data from two studies. These studies contained two samples: people with stroke (cerebrovascular accident, CVA) and people with SCD. Data from each sample was analyzed separately.

Participants

Participants provided written informed consent. Study procedures were approved by The University of Pittsburgh Institutional Review Board and conducted in compliance with the Helsinki Declaration.

Participants with stroke

Community-dwelling people with chronic stroke (CVA) were participants in an intervention study that promoted engagement in

daily activities to reduce sedentary behavior (ClinicalTrials.gov NCT03305731). Data were collected from December 2017 to December 2018. People who were: (1) greater than 6 months post-stroke; (2) ambulatory in the community; (3) reported greater than 6 hr of sedentary time daily (related to the parent study); and (4) resided within 50 miles of our research institution, were included. People who: (1) were currently participating in rehabilitation therapies (occupational, physical, or speech therapy); or who had: (2) current major depressive disorder, psychiatric disorder, or substance abuse disorder (Patient Health Questionnaire-9, PRIME-MD/Mini Neuropsychiatric Interview, (Kroenke et al., 2001; Spitzer et al., 1994); (3) cancer, in current treatment; or (4) diagnosis of neurodegenerative disorder, were excluded.

Participants with SCD

Participants with SCD were recruited from the University of Pittsburgh Medical Center (UPMC) Adult Sickle Cell Program outpatient clinic. Data were collected from October 2016 to December 2018 as part of a longitudinal study of neuroradiological biomarkers of cognitive function in SCD (ClinicalTrials.gov NCT02946905). All patients with HbSS, HbSC and HbS/ β -thalassemia – the three most prevalent genotypes of SCD – older than 18, and able to provide informed consent were informed about the study by staff members during their routine clinic visit and offered entry into the study if they were in steady-state SCD (Ballas, 2012). Eligibility criteria also included: (1) English-speaking; and (2) currently receiving routine follow-up care at the UPMC Adult Sickle Cell Program. Exclusion criteria included: (1) pregnancy as determined by a positive urine human chorionic gonadotropin test at the time of informed consent and (2) acute medical problem including acute vaso-occlusive crisis.

Measures

Assessments were conducted by independent raters trained to criterion and supervised by a senior neuropsychologist (MB). Assessments were administered during one or two testing sessions. If two testing sessions were required, they were scheduled within the same week. The CVA sample completed assessments in their homes in a quiet testing area free of distractions. The SCD sample completed assessments in the research clinic. All participants from both samples received monetary compensation for completion of assessments (unrelated to effort; all participants in the same parent study received the same amount).

NIH toolbox-CB

The NIH Toolbox-CB is a performance-based assessment of cognitive functions conducted using a mobile application (NIH Toolbox v.1.21, Glinberg & Associates Inc, Madison, WI) on an iPad Air 2 (Apple, Cupertino, CA, USA) device which has a 9.7" display. Algorithms within the mobile application compute age-corrected standard scores based on population mean (100) and standard deviation (15), which were used in the present analysis. The NIH Toolbox-CB was validated across the lifespan (ages 3–85) among a healthy population and demonstrated good test-retest reliability in adults aged 20–85 (ICC > 0.72) on all subtests (Weintraub et al., 2013). This assessment battery contains seven subtests and was designed to take no more than 30 min. The subtests include: (1) Flanker Inhibitory Control and Attention; (2) Pattern Comparison Processing Speed; (3) List Sorting Working Memory; (4) Picture Sequence Memory; (5) Oral Reading Recognition; (6) Picture Vocabulary; and (7) Dimensional

Change Card Sort. Administration procedures defined by the test developers were followed (including practice trials) and alternate test forms were used.

Flanker inhibitory control and attention

A row of stimuli (arrows) are presented at the center of the tablet screen. Participants indicate the direction of the center arrow by tapping the corresponding response button on the screen. Sometimes the center arrow matches the others, and sometimes it does not. Twenty test items are completed. The score for this subtest is based on speed.

Pattern comparison processing speed

Two stimuli are presented on the screen that may be the same or different. Participants are instructed to use their dominant hand to tap *yes* if the stimuli are the same and *no* if the stimuli are not the same. Participants are instructed to go as quickly as they can. The test continues until either 130 test items were completed or 85 s have passed. The score for this subtest is based on speed.

List sorting working memory

During condition 1, a series of stimuli that belong to the same category (either food or animals) are presented sequentially on the tablet screen. Participants are instructed to verbally indicate the stimuli that they observed on the screen from smallest to largest. During condition 2, a series of stimuli from two categories (food and animals) are presented. Participants are instructed to verbally indicate the stimuli that they observed on the screen beginning with food (from smallest to largest) and then animals (from smallest to largest). The assessor scores each response as correct (1) or incorrect (0).

Picture sequence memory

Images and verbal statements of events that might occur at a single event (such as Going to the Park) are presented sequentially and assigned the corresponding position around the edge of the tablet screen. Images are then scrambled in the center of the tablet screen and participants must replicate the sequence by dragging images to their appropriate position around the edge of the screen. Two test trials are completed, first with 15 items and second with 18 items.

Oral reading recognition

This is a computer adapted test in which participants are presented with words on the tablet that they must read out loud. The assessor scores correct (1) and incorrect (0) responses based on the NIH-CB pronunciation guide.

Picture vocabulary

This is a variable length computer adapted test in which a word is read aloud and four images are presented on the screen. Participants must select the image that corresponds with the word.

Dimensional change card sort

A stimulus is presented at the center of the screen, and two response options are provided on the screen. The participant is instructed to tap the response option which matches one of two dimensions (shape or color). Practice trials for each dimension are completed, and then thirty test items that include both dimensions are completed. The score for this subtest is based on speed.

Table 1. NIH Toolbox-CB and corresponding legacy measures

NIH toolbox	Legacy measure
Subtests	
Flanker inhibitory control and attention	DKEFS color-word interference condition 3: inhibition DKEFS color-word interference condition 4: inhibition/switching
Pattern comparison processing speed	RBANS Coding
Picture sequence memory	RBANS figure recall RBANS story recall RBANS list recall
Oral reading recognition	Wide range achievement test 4-reading
Dimensional change card sort	DKEFS color-word interference condition 4: inhibition/switching DKEFS test of trail making condition 4: switching DKEFS verbal fluency condition 3: category switching
List sorting working memory	No corresponding subtest
Picture vocabulary	No corresponding subtest
Composite/index scores	
Fluid cognition	RBANS total index
Crystallized cognition	No corresponding composite score
Total cognition	RBANS total index

Legacy neurocognitive measures

The Repeatable Battery for the Assessment of Neuropsychological Status (RBANS), Delis-Kaplan Executive Functions System (DKEFS), and Wide Range Achievement Test-4 (WRAT4) are well validated and widely applied measures of cognitive functions (e.g., Delis et al., 2001; Karr et al., 2019; McFarland, 2020; Randolph et al., 1998; Wilkinson & Robertson, 2006). The RBANS consists of 12 subtests that assess cognitive functioning across 5 domains: attention, language, immediate memory, delayed memory, and visuospatial/constructional. Subtests are described in detail by Randolph et al. (1998). Age corrected scaled scores were computed for each subtest. In addition, scores from each subtest were used to derive the total index score (Randolph et al., 1998). Age-corrected scaled scores from DKEFS Color-Word Interference (Condition 3: Inhibition, Condition 4: Inhibition/Switching), Test of Trail Making (Condition 4: Switching), and Verbal Fluency (Condition 3: Category Switching) were used to assess executive functions. The WRAT4-Reading Subtest raw score was used to compute an age-corrected standard score (Wilkinson & Robertson, 2006). Based on the designs of the parent studies, the WRAT4 was only administered to the SCD sample.

Statistical analyses

Four types of analyses were used to describe association between scores on the NIH Toolbox-CB and corresponding legacy measures: (1) linear correlations; (2) Bland-Altman analysis; (3) Lin's Concordance Correlation Coefficient; and (4) dichotomous agreement using percent agreement and Cohen's Kappa. Prior to conducting these analyses, corresponding subtests were identified by a senior neuropsychologist (MB) and guided by prior validity studies (Table 1, Carozzi et al., 2015; Dikmen et al., 2014; Gershon et al., 2014; Mungas et al., 2014; Tulskey et al., 2014; Zelazo et al., 2014). Analyses were conducted using SPSS Statistics for Windows, version 27.0 (SPSS Inc, Chicago, Ill., USA) and Microsoft Excel. Across all subtests, age-corrected scores were converted to Z-scores using the population mean

and standard deviation for each measure ($M[SD]$ for NIH Toolbox-CB and WRAT4-Reading = 100[15]; DKEFS = 10[3]; RBANS used age-specific population mean and standard deviation). Because the NIH Toolbox-CB Total Cognition Composite Score and RBANS Total Index Scores are both scaled using population $M(SD) = 100(15)$, Z -scores were not computed for total scores. Data were analyzed by sample. Distributions were examined for each subtest and normality was assessed using Shapiro-Wilk ($\alpha = .05$). Linear correlations between continuous scores on NIH Toolbox subtests and corresponding legacy subtests were assessed using Pearson's r (or Spearman's ρ , if non-normally distributed). Correlations were interpreted as weak (.1 to .3), moderate (.4 to .6), strong (.7 to .9), and perfect (1.0) (Akoglu, 2018). Then association between continuous scores on the NIH Toolbox-CB and corresponding legacy subtests were assessed using Bland-Altman analyses (Giavarina, 2015). First, new variables were created for each comparison that described: (1) difference between Z -scores; and (2) mean Z -scores. A one-sample t -test was conducted to determine if the mean difference between Z -scores differed from 0 ($\alpha = .05$). We report the magnitude and direction of difference from 0. Plots of the mean Z -scores versus difference between Z -scores and simple linear regression models were examined to determine if there were differences in magnitude and direction of agreement across sample means. Limits of agreement were computed. Next Lin's CCC (r) was computed for each comparison to describe the degree of precision and accuracy of NIH Toolbox-CB subtests relative to corresponding subtests on legacy measures (Barnhart et al., 2007; Lawrence & Lin, 1989). The CCC was interpreted using the same classifications as for the linear correlations described above, following Altman's recommendation to interpret the CCC similar to Pearson's r (Altman, 1991). Point estimates of the CCC within 95% confidence intervals were plotted to enable visual comparison among samples. Lastly, composite scores on the NIH Toolbox and RBANS were dichotomized to consider scores $\leq 1.5 SD$ as impaired, and all others not impaired. This is a widely used cut point for impairment on neuropsychological measures that represents scores at or below approximately the 10th percentile (Ciafone et al., 2020). Percent agreement and Cohen's Kappa were computed. Agreement based on Cohen's Kappa was interpreted as slight (.0 to .2), fair (.2 to .4), moderate (.4 to .6), substantial (.6 to .8), and almost perfect (greater than .8) (Watson & Petrie, 2010).

Results

Participants

Participant characteristics are displayed in Table 2. Analyses were completed by group.

Stroke

The stroke group ($n = 26$) averaged 68.85 (11.22) years of age and 46.2% were male. The majority of the group was white (76.9%), and had an average of 14.76 (2.93) years of education. Participants had sustained ischemic (84.6%) or hemorrhagic (15.4%) strokes affecting the right hemisphere (42.3%), on average 27.23 (13.76) months prior to enrollment in the research study.

SCD

The SCD group ($n = 64$) averaged 36.20 (11.92) years of age and 51.6% were male. The majority of this group was black (96.2%) and had an average of 13.52 (1.98) years of education.

Table 2. Participant characteristics

Mean (SD) or % (n)	Stroke ($n = 26$)	Sickle cell disease ($n = 64$)
Age, years	68.85 (11.22)	36.20 (11.92)
Gender, male	46.2 (12)	51.6 (33)
Race		
White	76.9 (20)	3.1 (2)
Black	19.2 (5)	96.9 (62)
American Indian	3.8 (1)	0.0 (0)
Education		
Years	14.76 (2.93)*	13.52 (1.98)
Stroke-specific		
Hemisphere		-
Right	42.3 (11)	
Left	46.2 (12)	
Bilateral	11.5 (3)	
Type		-
Ischemic	84.6 (22)	
Hemorrhagic	15.4 (4)	
Chronicity, months	27.23 (13.76)	-
Sickle cell disease-specific		
Genotype		
Hb-SS	-	46.9 (30)
Hb S/beta thalassemia	-	17.2 (11)
Hb-C	-	35.9 (23)

*Note. $n = 25$; exact years of education unknown for $n = 1$ with less than high school education.

Participants in this group carried the Hb-SS (46.9%), Hb S/beta thalassemia (17.2%), and Hb-C (35.9%) genotype.

Primary outcomes

Sample mean scores (Table 3) on the NIH Toolbox-CB fell within 1 standard deviation above or below the population mean (85 to 115) on all subtests and composite scores in the CVA sample except for Flanker Inhibitory Control and Attention, $M(SD) = 84.8(11.0)$ and Pattern Comparison Processing Speed, $M(SD) = 78.9(26.6)$, and on all subtests and composite scores in the SCD sample except for Flanker Inhibitory Control and Attention, $M(SD) = 78.9(14.7)$ and Fluid Cognition Composite Score, $M(SD) = 82.5(17.6)$. On legacy measures, sample mean scores fell within 1 standard deviation above or below the population mean (7 to 13) on all subtests in the CVA sample except for RBANS Figure Copy, $M(SD) = 6.8(5.8)$, and on all subtests in the SCD sample except for RBANS Coding, $M(SD) = 5.9(4.1)$ and DKEFS Test of Trail Making Condition 4, $M(SD) = 5.7(3.9)$.

Association between continuous scores on NIH Toolbox-CB and legacy measures

Results of the correlational analyses, Bland-Altman analyses, and Lin's Concordance Correlation Coefficients are presented in Table 4. Bland-Altman plots for all comparisons are available in the supplementary materials. Based on linear correlations, strong agreement was observed between NIH Toolbox-CB Fluid Cognition Composite Scores and RBANS Total Index Scores (CVA, $r = .90$, $p < .05$; SCD, $r = .88$, $p < .05$), and NIH Toolbox-CB Total Cognition Composite Scores and RBANS Total Index Scores (CVA, $r = .83$, $p < .05$). Lin's CCC demonstrated strong concordance between NIH Toolbox-CB Fluid and Total Cognition scores versus RBANS Total Index Scores in the CVA sample ($r = .78$ to $.79$, $p < .05$) and only moderate concordance in the SCD sample ($r = .60$, $p < .05$). The limits of agreement were wide in both samples when comparing NIH Toolbox-CB Fluid Cognition and NIH Toolbox-CB Total Cognition scores with

Table 3. Distribution of age-corrected scores on cognitive measures

	Stroke CVA, N = 26		Sickle cell disease SCD, N = 64	
	n	M(SD)	n	M(SD)
NIH Toolbox-CB				
Flanker inhibitory control and attention	26	84.8 (11.0)	57	78.9 (14.7)
Pattern comparison processing speed	26	78.9 (26.6)	57	87.2 (25.1)
List sorting working memory	24	97.1 (12.8)	58	89.9 (13.1)
Picture sequence memory	26	94.1 (15.1)	57	96.6 (15.2)
Oral reading recognition	25	107.9 (15.0)	58	95.2 (15.6)
Picture vocabulary	26	104.1 (13.8)	58	94.3 (12.6)
Dimensional change card sort	25	95.8 (14.7)	56	85.2 (20.2)
Crystallized cognition composite score	25	107.2 (13.9)	54	94.6 (13.1)
Fluid cognition composite score	24	86.5 (18.2)	52	82.5 (17.6)
Total cognition composite score	24	96.7 (16.0)	52	86.7 (14.9)
RBANS				
Picture naming	26	10.5 (4.5)	64	8.7 (5.4)
Semantic fluency	26	7.8 (3.3)	64	7.8 (3.9)
Line orientation	26	9.2 (4.8)	64	7.6 (4.1)
Figure copy	26	6.8 (5.8)	64	7.0 (4.9)
Digit span	26	10.0 (3.5)	64	8.5 (3.2)
Coding	26	7.0 (4.5)	64	5.9 (4.1)
List learning	26	9.2 (4.5)	64	8.1 (3.5)
Story memory	26	8.8 (4.5)	64	7.6 (3.8)
List recall	26	9.2 (4.3)	64	7.1 (4.1)
List recognition	25	8.5 (5.1)	64	7.1 (9.5)
Figure recall	26	9.1 (3.6)	64	7.7 (3.7)
Story recall	26	8.6 (4.5)	64	8.2 (3.5)
Total index score	26	91.7 (19.9)	62	81.2 (14.7)
DKEFS				
Color-word interference condition 3	24	8.6 (3.9)	63	8.2 (3.5)
Color-word interference condition 4	22	9.6 (3.2)	63	7.7 (3.2)
Test of trail making condition 4	24	8.1 (4.3)	64	5.7 (3.9)
Verbal fluency condition 3	25	8.4 (4.9)	64	7.7 (3.4)
WRAT4				
Reading	-	-	64	91.7 (14.2)

NIH Toolbox-CB subtest and composite, population mean(SD) = 100(15).
 RBANS subtest, population mean(SD) = 10(3).
 RBANS total index, population mean(SD) = 100(15).
 DKEFS subtest, population mean(SD) = 10(3).
 WRAT4 subtest, population mean(SD) = 100(15).

Note. Table S1 in the supplementary file contains NIH Toolbox-CB T-scores for this same sample, which are fully corrected for age, education, sex, and race/ethnicity.

the RBANS Total Index Score (Table 4). Among subtests, strong linear correlations were observed between NIH Toolbox-CB ORR and WRAT4 (SCD, $r = .81, p < .05$) and NIH Toolbox-CB DCCS and DKEFS Trails 4 (CVA, $r = .77, p < .05$). Lin's CCC demonstrated strong concordance between NIH Toolbox-CB ORR and WRAT 4 (SCD, $r = .82, p < .05$), and only moderate concordance between NIH Toolbox-CB DCCS and DKEFS Trails 4 (CVA, $r = .63, p < .05$). The limits of agreement were wide for both subtests (Table 4). Concordance between NIH Toolbox-CB subtests and corresponding legacy measures is depicted in Figure 1.

Agreement between impairment classification on NIH Toolbox-CB and legacy measures

Impairment classifications determined by NIH Toolbox-CB Composite Scores (Fluid Cognition and Total Cognition) were examined against the RBANS Total Index Score (Table 5).

Concordance Among Subtests of the NIH Toolbox-Cognition Battery and Legacy Measures of Cognition

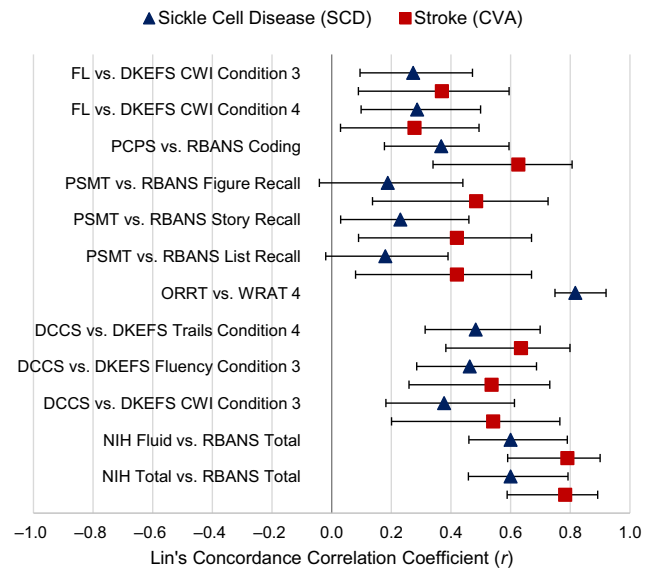


Figure 1. Concordance among subtests of the NIH Toolbox-Cognition Battery and legacy measures of cognition. Forest plot which illustrates the magnitude of Lin's Concordance Correlation Coefficient (r) in the sickle cell disease (blue triangles) and stroke (red squares) samples within 95% confidence intervals (error bars). Abbreviations: FL = Flanker Inhibitory Attention and Control; PCPS = Picture Comparison Processing Speed; PSMT = Picture Sequence Memory Test; ORRT = Oral Reading Recognition Test; DCCS = Dimensional Change Card Sort; NIH Total = NIH Toolbox Total Cognition Composite Score; DKEFS CWI = Delis-Kaplan Executive Functions Test Color-Word Interference; DKEFS Trails = Delis-Kaplan Executive Functions Test, Test of Trail Making; DKEFS Fluency = Delis-Kaplan Executive Functions Test Verbal Fluency; RBANS = Repeatable Battery for the Assessment of Neuropsychological Status; WRAT 4 = Wide Ranging Achievement Test, Version 4 Reading.

Across all comparisons, percent agreement ranged from 72 to 96%. Substantial agreement between the NIH Toolbox-CB Total Cognition Composite Score and the RBANS Total Index Score was detected in the CVA sample ($Kappa = .78, p < .05$). All other comparisons demonstrated only fair agreement ($Kappa = .34$ to $.36, p < .05$). Percent agreement between impairment classification on NIH Toolbox-CB subtests and corresponding legacy measures is available in the supplementary file (Table S2).

Discussion

Examining association between continuous scores on the NIH Toolbox-CB and legacy neuropsychological measures in samples at risk for cognitive impairments is a step toward validation of the NIH Toolbox-CB to characterize cognitive impairments across neurologically impaired populations. Scores on NIH Toolbox-CB had weak to strong associations with corresponding legacy measures in both samples. Wide limits of agreement ($>1 SD$ in either direction from the population mean) were observed across all subtests and composite scores. In addition, proportional bias detected on several subtests and the Fluid Cognition Composite Score in the SCD sample suggest that differences between individuals' scores on NIH Toolbox-CB and legacy measures may vary depending on the score. Group mean differences between Z-scores on NIH Toolbox-CB subtests and corresponding legacy measures were $<1 SD$ in either direction, suggesting that these subtests may provide

Table 4. Association between age-corrected scores on NIH Toolbox-CB and legacy measures

NIH Toolbox	Legacy measure	Sample	Linear Corr.	Bland-Altman analysis			Lin's CCC
			<i>r</i> **	Mean Δ (SE)	Limits of agreement (SE)	β (95% CI)	<i>r</i> (95% CI)
FL	CWI 3	CVA	.48*	-0.54 (0.23)*	-2.77, 1.70 (0.40)	-.70 (-1.17, -0.23)*	.37 (.09, .59)*
		SCD	.35*	-0.78 (0.16)*	-3.17, 1.60 (0.28)	-.28 (-0.65, 0.09)	.27 (.07, .45)*
	CWI 4	CVA	.44*	-0.78 (0.26)*	-3.16, 1.60 (0.45)	-.58 (-1.12, -0.04)*	.28 (.03, .49)*
		SCD	.35*	-0.65 (0.16)*	-2.97, 1.67 (0.27)	-.20 (-0.58, 0.18)	.29 (.07, .47)*
PCPS	Coding	CVA	.65*	-0.40 (0.27)	-3.12, 2.31 (0.47)	.20 (-0.18, 0.58)	.62 (.34, .81)*
PSM	Figure recall	CVA	.40*	0.49 (0.22)*	-2.80, 3.78 (0.39)	.31 (-0.04, 0.66)	.37 (.14, .56)*
		SCD	.49*	-0.10 (0.22)	-2.30, 2.10 (0.38)	-.23 (-0.72, 0.25)	.48 (.14, .72)*
PSM	Story recall	CVA	.19	0.22 (0.18)	-2.47, 2.92 (0.32)	-.20 (-0.65, 0.24)	.19 (-.06, .42)
		SCD	.55*	0.07 (0.27)	-2.63, 2.77 (0.47)	-.54 (-1.02, -0.05)*	.42 (.09, .67)*
	List recall	CVA	.30*	0.38 (0.17)*	-2.21, 2.96 (0.30)	-.16 (-0.58, 0.26)	.23 (-.00, .43)
		SCD	.45*	-0.12 (0.26)	-2.75, 2.51 (0.46)	-.49 (-0.98, 0.01)	.42 (.08, .67)*
ORR	WRAT4	CVA	.23	0.70 (0.19)*	-2.19, 3.60 (0.30)	-.43 (-0.85, -0.01)*	.18 (-.03, .38)
		SCD	-.81*	0.22 (0.07)*	-0.88, 1.32 (0.13)	.14 (-.02, 0.29)	.82 (.71, .88)*
DCCS	Trails 4	CVA	.77*	0.29 (0.21)	-1.70, 2.29 (0.36)	-.46 (-0.81, -0.11)*	.63 (.38, .80)*
		SCD	.51*	0.42 (0.17)*	-2.14, 2.98 (0.32)	.05 (-0.26, 0.36)	.48 (.27, .65)*
	Fluency 3	CVA	.62*	0.24 (0.26)	-2.29, 2.77 (0.45)	-.62 (-1.00, -0.23)*	.53 (.26, .73)*
		SCD	.48*	-0.22 (0.17)	-2.75, 2.31 (0.30)	.21 (-0.11, 0.53)	.46 (.24, .64)*
CWI 4	CVA	.55*	0.00 (0.20)	-1.82, 1.82 (0.34)	-.25 (-0.75, 0.24)	.54 (.20, .76)*	
	SCD	.39*	-0.20 (0.18)	-2.85, 2.46 (0.32)	.30 (-0.05, 0.66)	.38 (.14, .57)*	
Fluid cognition	Total Index	CVA	.90*	-8.46 (2.11)*	-28.70, 11.78 (3.65)	-.09 (-0.18, 0.36)	.79 (.59, .90)*
		SCD	.88*	0.60 (2.19)	-29.79, 30.99 (3.80)	.46 (0.15, 0.78)*	.60 (.41, .74)*
Total cognition	Total Index	CVA	.83*	1.71 (1.93)	-16.82, 20.23 (3.34)	-.05 (-0.31, 0.22)	.78 (.59, .89)*
		SCD	.64*	4.34 (1.68)*	-18.90, 27.58 (2.90)	.24 (-0.03, .51)	.60 (.41, .74)*

Note. Mean Δ = [NIH Toolbox Z-score] - [Legacy measure Z-score] for all subtests; [NIH Toolbox scaled score] - [Legacy measure scaled score] for total composite and total index scores. Statistical significance indicates whether Mean Δ differs from 0.

FL = Flanker Inhibitory Control and Attention; PCPS = Pattern Comparison Processing Speed; PSM = Picture Sequence Memory; DCCS = Dimensional Change Card Sort; Total Cognition = Total Cognition Composite; Fluid Cognition = Fluid Cognition Composite; CWI 3 = DKEFS Color-Word Interference Condition 3: Inhibition; CWI 4 = DKEFS Color-Word Interference Condition 4: Switching; Coding = RBANS Coding; Figure Recall = RBANS Figure Recall; Story Recall = RBANS Story Recall; List Recall = RBANS List Recall; Trails 4 = DKEFS Test of Trail Making Condition 4: Switching; Fluency 3 = DKEFS Verbal Fluency Condition 3: Category Switching; Total Index = RBANS Total Index; WRAT4 = Wide Range Achievement Test Version 4-Reading.

**p* < .05.

**Pearson's *r* for all except DCCS versus Trails 4 (CVA and SCD samples), FL versus CWI 3 (CVA sample), PSMT versus Story Recall (CVA and SCD samples), ORR versus WRAT4 (SCD sample). Spearman's *rho* was used in these cases due to non-normal distribution.

Table 5. Agreement in impairment classifications between RBANS total index and NIH Toolbox composite measures

	Sample	% Agreement	Kappa
Fluid cognition	CVA	79%	.36*
	SCD	72%	.36*
Total cognition	CVA	96%	.78*
	SCD	74%	.34*

Note. Impaired = 1.5 SD below population mean. Kappa values interpreted as fair = .21 to .40, moderate = .41 to .60, substantial = .61 to .80.

**p* < .05.

adequate estimates of group mean cognitive function among people with CVA and SCD. Further, despite limited overlap of cognitive functions assessed by the NIH Toolbox-CB and the RBANS, the NIH Toolbox-CB Total Cognition Composite Score had moderate to strong associations, and the NIH Toolbox-CB Fluid Cognition Composite Score had strong associations, with the RBANS Total Index Score. This suggests that the NIH Toolbox-CB Total Cognition Composite Score may be a valid measure of overall cognitive function among people with CVA and SCD.

Associations observed between NIH Toolbox-CB subtests and corresponding legacy measures were similar to those reported in prior validation studies of healthy populations on Pattern Comparison Processing Speed (*r* = .40 to .65 vs. *r* = .36 to .54, Carlozzi et al., 2015) and Oral Reading Recognition (*r* = .81 vs. *r* = .86) (Gershon et al., 2014). The present findings support the use of NIH Toolbox-CB ORR among people with SCD. On measures of executive functions, the magnitude of association was

smaller in the present study relative to healthy validation samples when comparing the Flanker Inhibitory Control and Attention subtest with Color Word Interference (Condition 3, Inhibition, *r* = .35 to .48 vs. *r* = .52) and the Dimensional Change Card Sort with Color Word Interference in the SCD sample (Condition 4, Inhibition/Switching *r* = .39 vs. *r* = .55) (Zelazo et al., 2014). A similar association between the Dimensional Change Card and Color Word Interference was detected in the CVA sample versus prior healthy validation studies (Condition 4, Inhibition/Switching *r* = .55 vs. *r* = .55) (Zelazo et al., 2014). Prior validation studies did not examine the Flanker Inhibitory Control and Attention subset relative to the Color Word Interference: Condition 4 (Inhibition/Switching) subtest, nor did they examine the Dimensional Change Card Sort relative to switching subtests on the Test of Trail Making and Verbal Fluency. The present findings suggest that switching may be related to the NIH Toolbox-CB Flanker Inhibitory Control and Attention, but that these subtests may tap different underlying constructs. Further, only one legacy measure of switching (Test of Trail Making) had a strong association with the Dimensional Change Card Sort in the CVA sample (*r* = .77). Moderate associations between Dimensional Change Card Sort and additional legacy measures of switching (*r* = .48 to .62) suggest that these subtests measure related but different cognitive functions among people with CVA and SCD. The Picture Sequence Memory Test was not associated with Figure Recall or List Recall in the SCD sample, and was only moderately associated with Figure Recall and List Recall in the CVA sample, and Story Recall in both samples. Differences in the stimuli themselves (e.g. story, abstract image, words) and the length of time elapsed

between presentation of the stimuli and recall on these subtests may contribute to these discrepancies.

An overall cognition score may be computed on both the NIH Toolbox-CB and the RBANS. Only 2 of the 7 NIH Toolbox-CB subtests overlap with cognitive functions assessed by the RBANS (Pattern Comparison Processing Speed, Picture Sequence Memory). Despite these differences, the NIH Toolbox-CB Fluid and Total Cognition Composite Scores and the RBANS Total Index Scores had moderate to strong associations. Further, these associations were stronger than associations on individual subtests and corresponding legacy measures (except for Oral Reading Recognition). This suggests that the NIH Toolbox Fluid and Total Cognition Composite Scores may represent the *g*-factor proposed by Spearman (1904) and be useful as a general measure of cognition among people with CVA and SCD.

Limitations

Results of the present analysis should be interpreted as exploratory because of limitations imposed by the purpose of the primary data collection. The CVA sample was smaller than the SCD sample. Although scores in analyses are age-adjusted for all measures, we did not adjust for race, ethnicity, or education. In addition, normative samples with unknown population-based equivalence were used to derive population norms for the DKEFS, RBANS, and WRAT4. We used these population-based $M(SD)$ to compute the standard scores used in the analyses, and remain mindful that this could contribute to bias. Scores on legacy measures of neuropsychological function are biased against Black people, largely reflective of disparities in social systems such as education quality and economic opportunity (Jean et al., 2019). There are varied published methods for adjusting NIH Toolbox-CB scores by race (Casaletto et al., 2015; Heaton et al., 2014) and extremely limited representation of Black people in neuropsychological research (Pugh et al., 2021). This is reflected in our predominantly white CVA sample. Future research which specifically recruits people from underrepresented racial and ethnic groups and applies optimal methods to produce fully adjusted scores accounting for race, ethnicity, education, and other social factors is important for building on these exploratory findings.

Future directions

Future studies that build on this work should aim to confirm findings among a broader range of clinical samples who have disorders associated with cognitive impairments. These studies will facilitate precise interpretation of NIH Toolbox-CB and support the use of this tool in broad epidemiological studies inclusive of people at risk for cognitive impairments.

Supplementary Material. To view supplementary material for this article, please visit <https://doi.org/10.1017/S1355617722000406>

Acknowledgements. We would like to acknowledge Michelle Zmuda and Isaac Delozier for recruiting participants and administering tests, and staff of the Occupational Therapy Cognitive Performance Laboratory for their role in participant recruitment and data collection. We would also like to acknowledge the blinded peer reviewers whose thoughtful feedback strengthened this paper.

Authors' contribution. Kringle EA: Conceptualization, Investigation, Formal analysis, Writing - Original Draft, Funding acquisition; Novelli E: Conceptualization, Investigation, Writing - Review & Editing, Funding acquisition; Butters MA: Conceptualization, Writing - Review & Editing, Supervision;

Skidmore ER: Conceptualization, Investigation, Resources, Writing - Review & Editing, Funding acquisition.

Funding statement. Research reported in this publication was supported by the National Heart Lung and Blood Institute (R01 HL127107-01A1, T32 HL134634, K23 HL159240), the National Center for Advancing Translational Sciences (UL1 TR001857) of the National Institutes of Health, the National Institute on Disability, Independent Living and Rehabilitation Research (90AR5023), the University of Pittsburgh School of Health and Rehabilitation Sciences PhD Student Award, and University of Pittsburgh Occupational Therapy Department Funds. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health, the National Institute on Disability, Independent Living and Rehabilitation Research, or the University of Pittsburgh.

Conflicts of interest. None.

References

- Akoglu, H. (2018). User's guide to correlation coefficients. *Turkish Journal of Emergency Medicine*, 18, 91–93.
- Altman, D. (1991). *Practical statistics for medical research*. Chapman and Hall.
- Ballas, S. K. (2012). More definitions in sickle cell disease: Steady state versus base line data. *American Journal of Hematology*, 87, 338–338.
- Barnhart, H. X., Haber, M. J., & Lin, L. I. (2007). An overview on assessing agreement with continuous measurements. *Journal of Biopharmaceutical Statistics*, 17, 529–569.
- Carlozzi, N., Beaumont, J. L., Tulskey, D. S., & Gershon, R. C. (2015). The NIH toolbox pattern comparison processing speed test: Normative data. *Archives of Clinical Neuropsychology*, 30, 359–368.
- Carlozzi, N., Goodnight, S., Casaletto, K., Goldsmith, A., Heaton, R., Wong, A., Baum, C. M., Gershon, R., & Tulskey, D. (2017a). Validation of the NIH toolbox in individuals with neurologic disorders. *Archives of Clinical Neuropsychology*, 32, 555–573.
- Carlozzi, N., Tulskey, D. S., Wolf, T. J., Goodnight, S., Heaton, R. K., Casaletto, K. B., Wong, A. W. K., Baum, C. M., Gershon, R. C., & Heinemann, A. W. (2017b). Construct validity of the NIH Toolbox Cognition Battery in individuals with stroke. *Rehabilitation Psychology*, 62, 443.
- 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, 378–391.
- Chakravorty, S., & Williams, T. N. (2015). Sickle cell disease: A neglected chronic disease of increasing global health importance. *Archives of Disease in Childhood*, 100, 48–53.
- Ciafone, J., Little, B., Thomas, A. J., & Gallagher, P. (2020). The neuropsychological profile of mild cognitive impairment in Lewy body dementias. *Journal of the International Neuropsychological Society*, 26, 210–225.
- Delis, D. C., Jacobson, M., Bondi, M. W., Hamilton, J. M., & Salmon, D. P. (2003). The myth of testing construct validity using factor analysis or correlations with normal or mixed clinical populations: Lessons from memory assessment; Testing construct validity; DC Delis et al. *Journal of the International Neuropsychological Society: JINS*, 9, 936.
- Delis, D. C., Kaplan, E., & Kramer, J. H. (2001). *Delis-Kaplan executive function system*. Psychological Corporation.
- Dikmen, S. S., Bauer, P. J., Weintraub, S., Mungas, D., Slotkin, J., Beaumont, J. L., Gershon, R., Temkin, R. K., & Heaton, R. K. (2014). Measuring episodic memory across the lifespan: NIH toolbox picture sequence memory test. *Journal of the International Neuropsychological Society*, 20, 611–619.
- Duncan, F., Wu, S., & Mead, G. E. (2012). Frequency and natural history of fatigue after stroke: A systematic review of longitudinal studies. *Journal of Psychosomatic Research*, 73, 18–27.
- Edwards, C. L., Raynor, R. D., Feliu, M., McDougald, C., Johnson, S., Schmechel, D., Wood, M., Bennett, G. G., Saurona, P., Bonner, M., Wellington, C., DeCastro, L. M., Whitworth, E., Abrams, M., Logue, P., Edwards, L., Martinez, S., & Whitfield, K. (2007). Neuropsychological assessment,

- neuroimaging, and neuropsychiatric evaluation in pediatric and adult patients with sickle cell disease (SCD). *Neuropsychiatric Disease and Treatment*, 3, 705.
- Gershon, R. C., Cook, K. F., Mungas, D., Manly, J. J., Slotkin, J., Beaumont, J. L., & Weintraub, S. (2014). Language measures of the NIH Toolbox Cognition Battery. *Journal of the International Neuropsychological Society*, 20, 642–651.
- Giavarina, D. (2015). Understanding bland Altman analysis. *Biochemia Medica: Biochemia Medica*, 25, 141–151.
- 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, 764–772.
- 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, 588–598.
- IBM Corp. (2020). *IBM SPSS Statistics for Windows, Version 127.0*. Armonk, NY: IBM Corp.
- Jean, K. R., Lindbergh, C. A., Mewborn, C. M., Robinson, T. L., Gogniat, M. A., & Miller, L. S. (2019). Education differentially buffers cognitive performance in black and white older adults. *The Journals of Gerontology: Series B*, 74, 1366–1375.
- Jorgensen, D. R., Metti, A., Butters, M. A., Mettenburg, J. M., Rosano, C., & Novelli, E. M. (2017). Disease severity and slower psychomotor speed in adults with sickle cell disease. *Blood Advances*, 1, 1790–1795.
- Karr, J. E., Hofer, S. M., Iverson, G. L., & Garcia-Barrera, M. A. (2019). Examining the latent structure of the Delis–Kaplan executive function system. *Archives of Clinical Neuropsychology*, 34, 381–394.
- Kassim, A. A., Pruthi, S., Day, M., Rodegheier, M., Gindville, M. C., Brodsky, M. A., DeBaun, M. R., & Jordan, L. C. (2016). Silent cerebral infarcts and cerebral aneurysms are prevalent in adults with sickle cell anemia. *Blood*, 127, 2038–2040.
- Kirkham, F. J., & Datta, A. K. (2006). Hypoxic adaptation during development: Relation to pattern of neurological presentation and cognitive disability. *Developmental Science*, 9, 411–427.
- Kroenke, K., Spitzer, R. L., & Williams, J. B. (2001). The PHQ-9: Validity of a brief depression severity measure. *Journal of General Internal Medicine*, 16, 606–613.
- Lawrence, I., & Lin, K. (1989). A concordance correlation coefficient to evaluate reproducibility. *Biometrics*, 45, 255–268.
- Mackin, R. S., Insel, P., Truran, D., Vichinsky, E. P., Neumayr, L. D., Armstrong, F. D., Gold, J. I., Kesler, K., Brewer, J., Weiner, M. W., and the Neuropsychological Dysfunction and Neuroimaging Adult Sickle Cell Anemia Study Group. (2014). Neuroimaging abnormalities in adults with sickle cell anemia: Associations with cognition. *Neurology*, 82, 835–841.
- McFarland, D. J. (2020). Factor-analytic evidence for the complexity of the Delis–Kaplan executive function system (D-KEFS). *Assessment*, 27, 1645–1656.
- Mole, J. A., & Demeyere, N. (2018). The relationship between early post-stroke cognition and longer term activities and participation: A systematic review. *Neuropsychological Rehabilitation*, 30, 346–370.
- Mungas, D., Heaton, R., Tulsky, D., Zelazo, P. D., Slotkin, J., Blitz, D., Lai, J.-S., & Gershon, R. (2014). Factor structure, convergent validity, and discriminant validity of the NIH Toolbox Cognitive Health Battery (NIHTB-CHB) in adults. *Journal of the International Neuropsychological Society*, 20, 579–587.
- Pugh, E., Robinson, A., De Vito, A. N., Bernstein, J. P., & Calamia, M. (2021). Representation of US Black Americans in neuropsychology research: How well do our reporting practices show that Black lives matter? *The Clinical Neuropsychologist*, 36, 214–226.
- Randolph, C., Tierney, M. C., Mohr, E., & Chase, T. N. (1998). The repeatable battery for the assessment of neuropsychological status (RBANS): Preliminary clinical validity. *Journal of Clinical and Experimental Neuropsychology*, 20, 310–319.
- Sanger, M., Jordan, L., Pruthi, S., Day, M., Covert, B., Merriweather, B., Rodeghier, M., DeBaun, M., & Kassim, A. (2016). Cognitive deficits are associated with unemployment in adults with sickle cell anemia. *Journal of Clinical and Experimental Neuropsychology*, 38, 661–671.
- Sexton, E., McLoughlin, A., Williams, D. J., Merriman, N. A., Donnelly, N., Rohde, D., Hickey, A., Wren, M.-A., & Bennett, K. (2019). Systematic review and meta-analysis of the prevalence of cognitive impairment no dementia in the first year post-stroke. *European Stroke Journal*, 4, 160–171.
- Spearman, C. (1904). "General intelligence," objectively determined and measured. *American Journal of Psychology*, 15, 201–292.
- Spitzer, R. L., Williams, J. B., Kroenke, K., Linzer, M., Verloin deGruy, F., Hahn, S. R., Brody, D., & Johnson, J. G. (1994). Utility of a new procedure for diagnosing mental disorders in primary care: The PRIME-MD 1000 study. *JAMA*, 272, 1749–1756.
- Tulsky, D. S., Carozzi, N., Chiaravalloti, N. D., Beaumont, J. L., Kisala, P. A., Mungas, D., Conway, K., & Gershon, R. (2014). NIH Toolbox Cognition Battery (NIHTB-CB): List sorting test to measure working memory. *Journal of the International Neuropsychological Society*, 20, 599–610.
- Vichinsky, E. P., Neumayr, L. D., Gold, J. I., Weiner, M. W., Rule, R. R., Truran, D., Kasten, J., Eggleston, B., Kesler, K., McMahon, L., Orringer, E. P., Harrington, T., Kalinyak, K., De Castro, L. M., Kutlar, A., Rutherford, C. J., Johnson, C., Bessman, J. D., Jordan, L. B., Armstrong, F. D., & the Neuropsychological Dysfunction and Neuroimaging Adult Sickle Cell Anemia Study Group. (2010). Neuropsychological dysfunction and neuroimaging abnormalities in neurologically intact adults with sickle cell anemia. *JAMA*, 303, 1823–1831.
- Watson, P. F. & Petrie, A. (2010). Method agreement analysis: A review of correct methodology. *Therogenology*, 73, 1167–1179.
- Weintraub, S., Dikmen, S. S., Heaton, R. K., Tulsky, D. S., Zelazo, P. D., Bauer, P. J., Carozzi, 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, S54–S64.
- Wilkinson, G., & Robertson, G. (2006). *Wide range achievement test (WRAT4)*. Psychological Assessment Resources.
- Zelazo, P. D., Anderson, J. E., Richler, J., Wallner-Allen, K., Beaumont, J. L., Conway, K. P., Gershon, R., & Weintraub, S. (2014). NIH Toolbox Cognition Battery (CB): Validation of executive function measures in adults. *Journal of the International Neuropsychological Society*, 20, 620–629.
- Zhao, L., Biesbroek, J. M., Shi, L., Liu, W., Kuijff, H. J., Chu, W. W., Abrigo, J. M., Lee, R. K. L., Leung, T. W. H., Lau, A. Y., Biessels, G. J., Mok, V., & Wong, A. (2018). Strategic infarct location for post-stroke cognitive impairment: A multivariate lesion-symptom mapping study. *Journal of Cerebral Blood Flow & Metabolism*, 38, 1299–1311.