Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Construct Validity of the NIH Toolbox Cognition Battery in Individuals With Stroke

Noelle E. Carlozzi,
Department of Physical Medicine and Rehabilitation, University of Michigan

David S. Tulsky,
Center for Health Assessment Research and Translation, Departments of Physical Therapy and Psychological and Brain Sciences, University of Delaware

Timothy J. Wolf,
Occupational Therapy and Department of Neurology, Washington University

Siera Goodnight,
Department of Physical Medicine and Rehabilitation, University of Michigan

Robert K. Heaton, and
Department of Psychiatry, University of California, San Diego

Kaitlin B. Casaletto
Department of Neurology, University of California, San Francisco

Alex W. K. Wong and **Carolyn M. Baum**
Occupational Therapy and Department of Neurology, Washington University

Richard C. Gershon and
Department of Medical Social Sciences and Department of Preventative Medicine, Northwestern University

Allen W. Heinemann
Center for Rehabilitation Outcomes Research, Shirley Ryan AbilityLab, Chicago, Illinois, and Department of Physical Medicine and Rehabilitation, Northwestern University Feinberg School of Medicine.

Abstract

Objective: The National Institutes of Health (NIH) Toolbox (NIHTB) for the Assessment of Behavior and Neurological Function Cognition Battery (NIHTB-CB) provides a brief assessment (approximately 30 min) of key components of cognition. This article examines construct validity to support the clinical utility of the NIHTB-CB in individuals with stroke.

Research Method: A total of 131 individuals with stroke ($n = 71$ mild stroke; $n = 60$ moderate/ severe stroke) completed the NIHTB-CB. Univariate analyses were conducted to examine the

Correspondence concerning this article should be addressed to Noelle E. Carlozzi, PhD, Department of Physical Medicine and Rehabilitation, University of Michigan, North Campus Research Complex, Building 14, 2800 Plymouth Road, Ann Arbor, MI 48109-2800. carlozzi@med.umich.edu.

Author Manuscript

cognitive profiles of the two different stroke groups (mild vs. moderate/severe stroke) on NIHTB-CB measures and composite scores. Pearson correlations were conducted between NIHTB-CB and established measures to examine convergent and discriminant validity. Effect sizes and clinical impairment rates for the different NIHTB-CB measures and composite scores were also examined.

Results: Participants experiencing moderate to severe stroke had poorer performance than did individuals with mild stroke on several of the NIHTB cognition measures. Evidence of convergent validity was provided by moderate to strong correlations between the NIHTB measures and the corresponding standard neuropsychological test (Pearson rs ranged from 0.31 to 0.88; median = .60). Evidence of discriminant validity was provided by smaller correlations between different cognitive domains than correlations of measures within the same domain. Effect sizes for composite and subtest scores regarding stroke severity were generally moderate-to-large. In addition, 42% of the sample were exhibiting mild cognitive impairment (i.e., 2 low scores on fluid tests).

Conclusions: Findings provide support for the construct validity of the NIHTB-CB in individuals with stroke.

Introduction

Stroke is one of the most common disabling medical conditions in adults. Almost 800,000 people each year in the United States experience a stroke, and an estimated 6.8 million Americans are living with chronic symptoms associated with stroke (Go et al., 2013). Stroke is characterized by chronic physical, psychological, cognitive, and functional impairments that affect activity, participation, and quality of life (Desrosiers et al., 2008; Edwards, Hahn, Baum, & Dromerick, 2006; White, Magin, & Pollack, 2009). According to the American Heart Association and the American Stroke Association (AHA/ASA) Clinical Practice Guidelines for Adult Stroke Rehabilitation Care, stroke rehabilitation that is initiated early after stroke can support stroke recovery and minimize functional disability (Bates et al., 2005). This process starts with a thorough evaluation by the health care team to understand the patient's strengths and limitations.

Approximately two out of three patients with stroke have some level of cognitive dysfunction (Salter, Teasell, Bitensky, Foley, & Bhogal, 2008). The extent of cognitive dysfunction is one of the primary concerns of poststroke rehabilitation. Even though existing guidelines recommend a thorough evaluation of cognitive status to guide treatment planning, clinical constraints often preclude time-intensive assessments in acute care settings. As a consequence, assessment of cognition following stroke typically consists of screening tools to detect gross changes in cognitive status; for example, Montreal Cognitive Assessment (MOCA; Bates et al., 2005; Godefroy et al., 2011). While screening tools may be moderately sensitive to global impairment, they are not comprehensive and, by design, target limited domains of cognitive function (Godefroy et al., 2011). Few instruments provide a

brief but relatively comprehensive evaluation of multiple domains of cognitive function. The National Institutes of Health (NIH) Toolbox for the Assessment of Behavioral and Neurological Function Cognition Battery (NIHTB-CB) was developed for this purpose. The NIHTB-CB is a brief assessment of key components of cognition that can be used across the life span (Gershon et al., 2010). The NIHTB-CB was designed to evaluate processing speed (PS), executive function, episodic memory, working memory, and language in approximately 30 min. The NIHTB-CB meets clinical demands of brevity but has the advantage of providing assessments across five cognitive domains. Thus, the NIHTB-CB could be used to screen for cognitive strengths and weaknesses and to provide the basis for a targeted, more in-depth assessment of cognition when appropriate, as well as highlight aspects of cognition that might be used to maximize outcomes. For example, an individual who has intact language but performance deficits for episodic memory might be encouraged to use compensatory strategies, such as creating lists (which requires language and comprehension) to aid in memory recall.

Previous research provides preliminary support for validity of the NIHTB in individuals with stroke (Carlozzi et al., 2017). With regard to the Cognition Battery, this work indicated that many individuals with stroke had significant impairments on fluid cognition measures (i.e., measures that reflect biological processes that change over the course of the life span and are typically sensitive to brain injury), supporting construct validity. This study also reported that impairment rates for crystallized cognition (i.e., measures that rely on language and comprehension that are more resilient to brain injury) were slightly elevated, which is consistent with findings of individuals with stroke (Caplan et al., 1990; Wall, Isaacs, Copland, & Cumming, 2015), again supporting construct validity. Furthermore, cognitive performance profiles across individuals with stroke, traumatic brain injury, and spinal cord injury supported known group validity.

While previous work provides preliminary support for the validity of the NIHTB-CB in individuals with stroke, data are lacking regarding convergent and discriminant validity. In addition, evidence of differential cognitive performance associated with stroke severity has not been reported. Establishing known group validity for cognitive performance relative to stroke severity is critical to establishing the clinical utility of this measure. Thus, this article focuses on evidence of convergent and discriminant validity and known-groups validity of the NIHTB-CB in community-dwelling individuals with stroke.

Method

Participants

We recruited 131 participants with medically confirmed diagnoses of stroke. The participants were recruited as part of a larger multi-site study (Tulsky & Heinemann, 2017). Stroke severity was classified according to the National Institutes of Health Stroke Scale (NIHSS; Goldstein, Bertels, & Davis, 1989; Kwah & Diong, 2014); 54% were classified as having a mild stroke (scores of 1–5; $n = 71$) and 46% had a moderate/severe stroke (scores of 6–24; $n = 54$ moderate and $n = 6$ severe). Individuals with moderate/severe stroke were combined because of the small number of participants with severe strokes, and because this distribution is typical of individuals served in rehabilitation settings. Participants were at

least 18 years old. Participants were excluded if they demonstrated evidence of aphasia on the Frenchay Aphasia Screening Test (Enderby & Crow, 1996; Enderby, Wood, Wade, & Hewer, 1987; Salter, Jutai, Foley, Hellings, & Teasell, 2006), could not read or understand English at a fifth-grade level (as determined by the Wide Range Of Achievement–4th Edition Reading Test; Wilkinson & Robertson, 2006), or had vision that was poorer than 20/100 (as determined by the Lighthouse Near Visual Acuity Test, 2nd edition; Bailey & Lovie, 1976; Ferris & Bailey, 1996). Participants were administered the NIHTB-CB battery and neuropsychological tests as part of a larger battery. Participants ranged in age from 22–83 years ($M = 57.5$; $SD = 12.6$) and were evenly distributed across gender (51% male). The median time since stroke was 29.0 months (range = 12.5–87.3; $M = 31.5$; $SD = 11.8$). Data were collected in accordance with local institutional review board requirements; analyses highlighted in this article focus on hypotheses establishing validity evidence for individuals with stroke. Some analyses also include scores from the NIHTB Motor Battery (Reuben et al., 2013); the 9-Hole Pegboard Dominant Hand Dexterity Test was used as a measure of dominant hand motor function or stroke-related dysfunction (Reuben et al., 2013), because this could affect performance on NIHTB-CB tests with timed responses (Pattern Comparison, Flanker, Dimensional Change Card Sort). Furthermore, 43% of mild and 67% of those with moderate/severe strokes had low dominant hand pegboard times ($t < 40$; $\chi^2 = 7.41$, $p < .01$).

NIHTB-CB Core Measures

The NIHTB Picture Vocabulary Test (Gershon et al., 2013, 2014) measures receptive vocabulary and requires participants to identify which of four photographs matches the meaning of an orally presented word. Scores are based on the results of a computer adaptive test, and represent the participant's word knowledge.

The NIHTB Oral Reading Recognition Test (Gershon et al., 2013, 2014) measures literacy and quality of education; it requires participants to read and pronounce letters and words. Scores are based on the results of a computer adaptive test, and represent the participant's level of letter/word knowledge.

The NIHTB Picture Sequence Memory Test (Bauer et al., 2013; Dikmen et al., 2014) requires sequencing a series of pictures that are presented on a computer screen; it measures episodic memory. Sequence length varies from 6–18 pictures depending on the participant's age. Scores reflect the number of adjacent pairs that are identified correctly over two trials; the maximum score is one item less than the sequence length that was presented (i.e., for a sequence length of 18, the maximum score is 17).

The NIHTB Pattern Comparison Processing Speed Test (Carlozzi et al., 2014; Carlozzi, Tulsky, Kail, & Beaumont, 2013) requires participants to identify whether two visual patterns are the same or not; it measures processing speed. Patterns were identical or vary on one of three dimensions: color (all ages), adding/taking something away (all ages), or one versus many. Scores are the number of correct items out of 130 completed in 90 s.

The NIHTB List Sorting Working Memory Test (Tulsky et al., 2013, 2014) involves size order sequencing of familiar stimuli and measures working memory. Stimuli are presented

visually and orally; participants are required to place them in size order. There is both a one- and two-list version of this task; in the one-list version, participants sequence items from a single category (i.e., food OR animals), while in the two-list version, participants sequence items in two different categories (i.e., reporting animals in size order, followed by food in size order). Scores are the combined total items correct on the one- and two-list versions (maximum = 28), with higher scores indicating better performance.

The NIHTB Flanker Inhibitory Control and Attention Test (Zelazo et al., 2013, 2014) assesses inhibitory control, a component of executive functioning. This test requires participants to focus on a stimulus (i.e., arrow), that is flanked by two arrows; the participant must focus on the middle stimuli while ignoring the flanking stimuli and indicate in which direction the arrow is facing (consistent or inconsistent with the flanking arrows). Scores are based on a combination of accuracy and response time. Accuracy is defined as 0.125 times the number of correct responses. For participants with 80% accuracy, final scores were equal to accuracy scores. For participants with >80% accuracy, a reaction time (RT) score was calculated based on the participant's median response time on correct, inconsistent flanking trials from the mixed block administration. To calculate scores, RTs <100 ms or greater than 3 SDs from the participant's average RT were considered outliers and removed from further analysis. Median RTs for each participant were calculated. Next, because of a positive skew in RTs, a log (Base 10) transformation was used to normalize scores. These log values were rescaled from a log(500) – log(3000) range to a 0–5 range and reversed such that smaller log values were at the upper end of the 0–5 range whereas larger log values were at the lower end. These rescaled RT scores were added to the accuracy scores for participants with >80% accuracy.

The NIHTB Dimensional Change Card Sort (DCCS) Test (Zelazo et al., 2013, 2014) assesses cognitive flexibility (i.e., task switching/set shifting), a component of executive functioning. The first portion of this test requires matching along one dimension(i.e., color), whereas the second requires matching along the other dimension (i.e., shape), and the third involves matching between both dimensions. Scores are based on a combination of accuracy and response time using the process outlined under the NIHTB Flanker Inhibitory Control and Attention Test.

NIHTB-CB Composite Scores

NIHTB Fluid Cognitive Composite score (Heaton et al., 2014) reflects a variety of abilities that are involved in adapting to novel cognitive tasks and are especially sensitive to normal aging and acquired brain dysfunction. This Composite Score combines performances on the DCCS Test, Flanker Test of Executive Function-Inhibitory Control and Attention, Picture Sequence Memory Test of Episodic Memory, List Sorting Working Memory Test, and Pattern Comparison Processing Speed Test.

The NIHTB Crystallized Cognitive Composite score (Heaton et al., 2014) reflects accumulated verbal knowledge, skills, and education. It includes the Picture Vocabulary and Oral Reading scores.

The NIHHTB Overall Cognitive Composite score (Heaton et al., 2014) is an index of overall cognition and represents both fluid and crystallized abilities. This score is the average of the Fluid and Crystallized Composites.

Measure of Motor Functioning

The NIHTB Motor Battery (Reuben et al., 2013) 9-Hole Pegboard Dexterity Test, Dominant Hand (Wang et al., 2011) measures dominant hand motor functioning. Scores reflect time to completion (in seconds) with higher scores indicating worse functioning. Scores were Winsorized for the eight participants who were unable to complete this test because of motor impairments.(i.e., these participants were given a T score that was 1 point lower than the lowest obtained score for the rest of the sample).

NIHTB Normative Standards

For all of the NIH Toolbox Cognition Battery (NIHTB-CB) measures and composite scores described previously, as well as the 9-Hole Pegboard Dexterity Test, Dominant Hand from the NIHTB Motor Function battery, demographically corrected normative scores were utilized. Normative standards were developed in a cohort of neurological healthy adults ($N =$ 972) to determine deviations from expected levels of performances. Details regarding these norms are presented in detail in Casaletto et al. (2015). In brief, multiple fractional polynomial models were used to regress the normalized NIHTB-CB scores of each test separately for each race/ethnicity (i.e., Caucasian, African American, Hispanic White) on demographic characteristics (i.e., age, education, gender). The residuals from these models were corrected to enhance the homogeneity of the variances across demographics (age, sex, education, race/ethnicity). The corrected residuals were standardized and rescaled to form individual T scores. The resulting fully corrected T score ($M = 50$, $SD = 10$) for each test therefore represents an individual's neurocognitive performance compared with age-, education-, sex- and race/ethnicity-matched peers; these scores were used in the primary analyses and analyses examining effect sizes for the NIHTB-CB measures. Age corrected T scores ($M = 50$, $SD = 10$) for each test representing an individual's neurocognitive performance compared with age-matched peers were used in analyses that involved examining convergent and discriminant validity.

Validation Measures

Table 1 provides the NIHTB measure matched with a corresponding validation measure.

Auditory Verbal Learning Test (Rey) (National Institutes of Health & Northwestern University, 2017) is a list of 15 unrelated words presented three times in the same order; this measure is an abbreviated version of the Rey Auditory Verbal Learning Test (RAVLT; Strauss, Sherman, & Spreen, 2006) that was designed specifically for the NIHTB-CB (National Institutes of Health & Northwestern University, 2017). Participants recall as many of the words as possible. Age-corrected scores were used in convergent and discriminant validity analyses and reflect the number of correct answers out of 45. It serves as a criterion measure for the NIHTB Picture Sequence Memory Test.

Brief-Visuospatial Memory Test—Revised (BVMT-R; Benedict, 1997; Benedict, Schretlen, Groninger, Dobraski, & Shpritz, 1996) requires participants to reproduce six geometric figures presented in a 2×3 array after viewing the images for 10 s. Participants draw as many of the figures as they can remember after each of three learning trials. Age corrected scores were used in convergent and discriminant validity analyses and reflect the number correct. This test also serves as a criterion measure for the NIHTB Picture Sequence Memory Test.

The PPVT-IV (Dunn & Dunn, 2007) measures receptive vocabulary skills. Participants identify which of four pictures reflects a word spoken by the examiner. Age corrected scores were used in convergent and discriminant validity analyses and reflect the number of correct answers out of a possible 228. It serves as a criterion measure for the NIHTB Picture Vocabulary Test.

Wide Range Achievement Test 4th edition (WRAT-4) Reading Subtest (Wilkinson & Robertson, 2006) is a norm-referenced test in which participants name letters and read aloud words out of context. The words are listed in order of decreasing familiarity and increasing phonological complexity. Age corrected scores were used in convergent and discriminant validity analyses and reflect the number of words that were pronounced correctly. It serves as a criterion measure for the NIHTB Oral Reading Recognition Test.

Wechsler Adult Intelligence Scale-Fourth Edition (WAIS-IV) Symbol Search (Wechsler, 2008) presents participants with two target symbols (designs) followed by a test series of symbols that include or do not include a target design. The participant searches the test series to identify the target symbol. The participant is allowed 120 s to complete as many as possible. Age-corrected scores were used in convergent and discriminant validity analyses and reflect number correct, minus number incorrect (maximum = 60). It serves as a criterion measure for the NIHTB Pattern Comparison Processing Speed Test.

WAIS-IV Coding (Wechsler, 2008) requires the participant to associate numbers and symbols using a key. This speeded measure is sensitive to deficits in processing speed but also requires motor coordination, short-term memory (STM), and visuoperceptual abilities (Tulsky, Saklofske, & Zhu, 2003). Age-corrected scaled scores were used in convergent and discriminant validity analyses and reflect the number correct. It also serves as a criterion measure for the NIHTB Pattern Comparison Processing Speed Test.

WAIS-IV Letter-Number Sequencing (Wechsler, 2008) presents participants with a mixed list of numbers and letters and asked to repeat the list by saying the numbers in ascending order and then letters in alphabetical order. This subtest has a strong working memory component (Crowe, 2000; Gold, Carpenter, Randolph, Goldberg, & Weinberger, 1997; Haut, Kuwabara, Leach, & Arias, 2000). Age-corrected scores were used in convergent and discriminant validity analyses and reflect the number of correct responses for letter-number strings ranging from 3 to 9 items (maximum 30 points). It serves as a criterion measure for the NIHTB List Sorting Working Memory Test.

Delis-Kaplan Executive Function System (D-KEFS) Color/Word Interference (Delis, Kaplan, & Kramer, 2001) measures ability to inhibit overlearned verbal responses.

Participants are timed while (a) naming of color patches, (b) reading color words printed in black ink, and (c) naming the color of the ink in which color words are printed in the "wrong" color of ink (i.e., red written in blue ink). The standard DKEFS Color/Word Interference Test also includes an additional switching task that was not administered as a part of this study. We examined age corrected scores on the interference trial in convergent and discriminant validity analyses. This test also serves as a criterion measure for NIH Toolbox executive function subtests (Dimensional Change Card Sort, Flanker Inhibitory Control and Attention Test).

NIHTB and Validation Measure Administration and Scoring Certification Process

The NIHTB assessments and validation measures were administered by examiners with at least a bachelor's degree. They completed a certification process for test administration prior to their working with study participants. Two PhD-level licensed Clinical Psychologists who were also involved in the development of the NIHTB-CB and received formal training to administer these measures (NEC and DST) provided training to review the standardized methods of administering and scoring the NIHTB measures and neuropsychological tests. The examiners practiced the assessments and completed a minimum of three practice cases before completing the formal certification process which required an additional practice administration that was observed by one of the trainers. Trainers provided examiners detailed feedback regarding test administration and scoring of standard neuropsychological tests for which computerized scoring was not available. Examiners who strictly followed standardized protocol procedures were certified and tested participants; examiners who failed the certification process completed additional practice sessions until achieving certification. Examiners repeated the recertification process annually to assure adherence to standardized procedures and minimize drift from the protocol.

NIHTB and Validation Measure Scoring Process

After test administration certification, a PhD-level Clinical Psychologist (NEC) reviewed 10 de-identified test protocols from each examiner and provided feedback about deviations from protocol. Two examiners independently scored validation measures; the PhD-level Clinical Psychologist (NEC) reviewed and reconciled discrepant scores.

Data Analysis

Univariate analyses examined the main effects and interactions between stroke laterality and stroke severity; these analyses did not yield significant findings (all $p > .05$); thus, we did not include laterality as a covariate in our analyses. A univariate analysis examined group (mild vs. moderate/severe stroke) and the seven NIHTB-CB Core Measures and NIHTB Composite Scores as the dependent variables to compare the groups' cognitive profiles using fully corrected NIHTB scores. For NIHTB-CB tests that rely on motor responses (i.e., Pattern Comparison, Flanker and DCCS), analyses were ran both with and without motor function as a covariate (as noted above, however, participants with moderate/severe strokes tended to have worse motor impairment, $p < .01$). In addition, univariate analyses also examined stroke severity and the 9 standard neuropsychological tests as the dependent variables. For standard neuropsychological tests that rely on speeded motor responses (i.e.,

WAIS-IV Coding and WAIS-IV Symbol Search), analyses were run both with and without motor function as a covariate.

Pearson correlations were computed to evaluate relationships between the NIHTB measures and standard neuropsychological tests to evaluate convergent validity using age-corrected scores. Pearson correlations were computed to evaluate discriminant validity among cognitive domains of function among the NIHTB measures. Evidence of discriminant validity consisted of lower correlations (0.1 differences) with measures of a *different* cognitive construct. Across measures, correlations less than 0.3 were considered weak, 0.3–0.6 adequate, and 0.6 or greater were good to very good evidence of convergent validity; evidence of discriminant validity consisted of lower correlations with selected measures of a *different* cognitive construct (Campbell & Fiske, 1959). For NIHTB-CB tests that rely on motor responses (i.e., Pattern Comparison, Flanker, and DCCS), as well as standard neuropsychological tests that rely on speeded motor responses (i.e., WAIS-IV Coding, and WAIS-IV Symbol Search) analyses were conducted both with and without motor function as a covariate.

Effect sizes are reported as Cohen's *d*, with cutoffs of .20, .50, and .80 indicating small, medium, and large effects, respectively (Cohen, 1992). Clinical impairment rates were calculated based on Holdnack and colleagues' (2017) approach to identify base rates for clinically significant impairment. Thus, for individuals with Crystallized *t* scores 58 the cutoff for clinical impairment for fluid tests was a score <44, for individuals with Crystallized *t* scores between 50 and 57 the cutoff for clinical impairment for fluid tests was a score <41, for individuals with Crystallized *t* scores between 43 and 49 the cutoff for clinical impairment for fluid tests was a score <38, for individuals with Crystallized *t* scores <43 the cutoff for clinical impairment for fluid tests was a score <35. We calculated the percent of participants who met Diagnostic and Statistical Manual of Mental Disorders (5th ed., *DSM–5*; American Psychiatric Association, 2013) criteria for at least a mild cognitive disorder defined as the presence of at least two low scores on an NIHTB-CB fluid test.

Results

Missing Data

Table 2 displays missing data, stratified by stroke severity and NIHTB-CB subtest. There were no group differences between those with mild versus those with moderate/severe stroke for rates of missing data.

Demographic Characteristics

The demographic characteristics of the mild and moderate/severe stroke groups are presented in Table 2. Groups did not differ on age, $t(129) = .10$, $p = .92$, time since injury, $t(129) = -0.18$, $p = .86$, gender, $\chi^2(1, N = 131) = .25$, $p = .37$, or education, $t(127) = 828$, $p = .41$. There were no significant group differences on race, $\chi^2(1, N = 131) = 3.484$, $p = .062$ and ethnicity, $\chi^2(1, N = 129) = .16$, $p = .69$. The average NIHTB composite, subtest, and supplemental fully corrected scaled scores for the different stroke groups are presented in Table 3.

Univariate analyses examined stroke severity, the two NIHTB composite scores, and seven NIHTB subtest scores as the dependent variables. In the first set of analyses that did not control for motor function, individuals with moderate/severe stroke performed worse than those with mild stroke on both NIHTB composites and revealed significant group differences on all NIHTB subtests except Picture Vocabulary and List Sorting (Table 4); in all cases with significant group differences, individuals with moderate/severe stroke performed worse than those with mild stroke. There were no significant differences in the second set of analyses when motor function was included as a covariate for tests that rely on speeded motor responses (i.e., Pattern Comparison, Flanker, and DCCS; Table 4).

Univariate analyses also examined stroke severity and performance on the nine standard neuropsychological tests. In the first set of analyses that did not control for motor function, individuals with moderate/severe stroke performed worse than those with mild stroke on all measures (Table 5); in all cases, individuals with moderate/severe stroke performed worse than those with mild stroke. There were no significant differences in the second set of analyses when motor function was included as a covariate for tests that rely on speeded motor responses (i.e., WAIS-IV Coding and WAIS-IV Symbol Search; Table 5).

Convergent and Discriminant Validity

Table 6 reports evidence of convergent validity. We found moderate to large correlations between the NIHTB measures and their corresponding neuropsychological test regardless of whether motor function was included as a covariate. Moderate relationships were found between Picture Sequence Memory, Pattern Comparison, Flanker, and DCCS, and their neuropsychological comparison measure(s) (r's ranged from .40 to .67), and strong relationships were found between the two crystallized NIHTB-CB tests (Picture Vocabulary and Oral Reading Recognition) and their comparator (r's ranged from .87 to .88); this pattern of findings did not change when motor function was included as a covariate for tests relying on motor responses. For List Sorting, there were moderate correlations with WAIS-IV Letter Number Sequencing (r = .66). In general, correlations between measures of different cognitive domains were smaller than correlations of measures of the same domain, providing evidence of discriminant validity (Table 7). Specifically, for Picture Sequencing, correlations with one of its primary comparator measures (i.e., BVMT) was higher (r = .66) than correlations with tests from any other domain (rs ranged from .21 to .54; median = .38). For List Sorting, correlations with its primary comparator (i.e., WAIS-IV Letter Number; r = .64) was higher than correlations with tests on any other domain (rs ranged from .28 to .61). In addition, similar findings were observed for Pattern Comparison, where correlations with other processing speed measures (i.e., WAIS-IV Coding and WAIS-IV Symbol Search) were higher (rs ranged from .59-.67; median = .63) than correlations with tests on any other domain (rs ranged from .25 to .47; median = .36). A similar and more robust pattern of correlations was seen for the two crystallized NIHTB-CB measures (Oral Reading Recognition and Picture Vocabulary) where the relationships with comparator measures were stronger with other measures of language (for Oral Reading recognition rs ranged from .77 to .88 and for Picture Vocabulary, rs ranged from .74 to .87; median = .81) than they were with tests in any other domain (rs ranged from .14 to .66; median = .40). Finally, the pattern of findings was less robust for the two executive function measures (DCCS and

Flanker), where moderate relationships were found with measures in most domains (rs ranged from .46 to .54 with the standard neuropsychological executive function measures, rs ranged from .33 to .57 with measures of Episodic memory, rs ranged from .37 to .55 with measures of Language, rs ranged from .60 to .65 with other Processing Speed Measures, and rs ranged from .41 to .49 with the working memory measure). Controlling for motor function for tests that relied on Motor Responses (i.e., analyses that included Pattern Comparison, Flanker, DCCS, WAIS-IV Coding, and WAIS-IV Symbol Search), yielded similar findings.

Other Demographic Comparisons

We computed effect sizes to describe the influence of stroke severity on NIHTB performance. Table 8 shows that the effect sizes for the measures of fluid cognition were generally small to moderate for the mild stroke group and moderate to large for the moderate/severe stroke groups (Cohen, 1977). For the mild group, the largest effects were seen for Pattern Comparison, List Sorting, Flanker, and DCCS; the largest effects for the moderate/severe group were for Flanker, DCCS and Pattern Comparison. Table 8 also indicates rates of impairment ranged from 31.0% to 37.4% for the entire sample for individual NIHTB-CB test, and 42% had clinical impairment on the fluid composite; 42.0% of the sample would meet *DSM–5* criteria for at least a mild cognitive disorder.

Discussion

We evaluated the construct validity of the NIHTB-CB in individuals with stroke by comparing individuals with mild versus moderate/severe stroke. Individuals with moderate/ severe stroke performed worse than those with mild stroke on both NIHTB Composite measures (i.e., Fluid Cognition and Crystallized Cognition), as well as all on most of tests of fluid reasoning (i.e., Picture Sequence Memory, Flanker, DCCS, Pattern Comparison) and oral reading (i.e., Oral reading Recognition). For NIHTB-CB tests that rely on speeded motor function (Pattern Comparison, DCCS and Flanker), group differences were nonsignificant, suggesting that group differences are partially confounded with difficulties with motor function. Users should exercise caution when using the NIHTB-CB tests that rely on motor responses in individuals with stroke when drawing conclusions about cognitive impairment when clinically significant problems in motor function exist.

Measures of processing speed (Pattern Comparison) and executive function (DCCS and Flanker) yielded the largest effect sizes (vs. the total normative group) for participants with severe strokes. Findings are consistent with earlier reports that describe deficits in processing speed (Barker-Collo & Feigin, 2006; de Bruijn, Synhaeve, van Rijsbergen, de Leeuw, Jansen, & de Kort, 2014; Su, Wuang, Lin, & Su, 2015), and executive function (Barker-Collo & Feigin, 2006; de Bruijn et al., 2014; Sörös, Harnadek, Blake, Hachinski, & Chan, 2015), and provide support for the validity of the NIBTB-CB in individuals with severe stroke. Although we excluded participants with dysphagia, group differences on the Crystallized Composite score suggest that that subtle levels of dysphagia were present.

Convergent and discriminant validity of the NIHTB-CB was supported by these findings. Each test demonstrated moderate to large correlations with its corresponding

neuropsychological test (Table 6) providing evidence of convergent validity. The strongest correlations were seen for measures of language; moderate correlations were demonstrated for all other domains. Discriminant validity was supported by smaller correlations among measures of dissimilar constructs (Table 7). For example, correlations among NIHTB-CB measures of language (for Oral Reading recognition rs ranged from .77 to .88 and for Picture Vocabulary, rs ranged from .74 to .87) and were higher than correlations with tests on any other domain (rs ranged from .21 to .54). A similar pattern was observed across other cognitive domains although less robust than what was seen for the language measures, with the exception executive function measures where moderate relationships were found, suggesting that these measures tap abilities related to multiple aspects of cognition, rather than only executive functioning. These tests are simple discrimination tasks characteristic of executive functioning tasks used in the experimental psychology literature, rather than neuropsychological assessments of other aspects of executive functioning, partially explaining the weaker evidence of discriminant validity. Controlling for motor function did not alter the pattern of findings.

Impairment rates observed here are higher than the general population for all NIHTB fluid measures. We expect 16% of the people in the general population to score 1 SD below the mean (Heaton, Miller, Taylor, & Grant, 2004). Our findings suggested overall impairment rates ranging from 31.0% to 37.4% for the NIHTB-CB fluid tests, with 42.0% of our sample meeting diagnostic criteria for at least a mild cognitive disorder. These findings are consistent with other studies examining cognition in stroke and support the construct validity of these measures (e.g., Barker-Collo & Feigin, 2006; Black, 2011; de Bruijn et al., 2014; Desmond, 2004; Hochstenbach, den Otter, & Mulder, 2003; Hochstenbach, Mulder, van Limbeek, Donders, & Schoonderwaldt, 1998; Salter et al., 2008; Schendel, Dronkers, & Turken, 2016).

Results provide evidence supporting the construct validity of the NIHTB-CB in individuals with stroke. Study strengths include a diverse sample of community dwelling individuals with stroke, medically documented stroke severity, as well as broad inclusion criteria; individuals were not excluded based on medication use, litigation status, previous psychiatric history, or previous history of learning disability, which is commonplace in the published literature. Findings should be generalizable to other heterogeneous stroke samples. Researchers should note the sensitivity of the NIHTB-CB processing speed measures consistent with others' work that indicates that processing speed measures are the most sensitive measures to cognitive insult (Carlozzi, Kirsch, Kisala, & Tulsky, 2015; DeLuca, Chelune, Tulsky, Lengenfelder, & Chiaravalloti, 2004; Donders, Tulsky, & Zhu, 2001; Gontkovsky & Beatty, 2006; Hawkins, 1998). However, it is important to note that the deficits that are observed for processing speed, as well as for the other NIHTB-CB measures that rely on speeded motor responses are due, at least in part, to motor functioning. This influence should be considered by clinicians when interpreting findings for individuals with stroke and significant motor impairments. However, because dominant hand motor speed was worse in participants with moderate/severe stroke, "corrections" for motor impairments biases against finding stroke severity differences on other tests.

It is important to acknowledge several study weaknesses. First, while our sample is heterogeneous, we did not evaluate participants' effort; it is possible that some participants exaggerated cognitive deficits. Future studies should include symptom validity testing for all participants to ensure that impaired performance is not related to poor effort. However, the pattern of NIHTB-CB results, with essentially normal performances on Crystallized cognition (especially oral reading) and much more apparent deficits on Fluid cognition measures, are consistent with expectations for a group with an acquired neurological disorder and certainly do not suggest a lack of effort. We did not evaluate the effects of medication use, psychiatric history, or learning disability on test performance; more work is needed to determine the effects that these variables have on NIHTB-CB performance.

Despite these limitations, this study provides support for the NIHTB-CB as a valid assessment of cognitive functioning in individuals with stroke. Future work should examine longitudinal NIHTB-CB test performance in order to document the typical recovery of function that is characteristic of individuals with stroke during the first three to six poststroke when the largest gains in cognitive function typically occurs (Hochstenbach et al., 2003; Sonoda, Chino, Domen, & Saitoh, 1997; Wade, Wood, & Hewer, 1985). More research is needed to understand how medication use, litigation status, psychiatric history, and history of learning disability affect NIHTB test performance. In addition, future studies should examine acute recovery following stroke onset (i.e., within 3 to 12 months).

Acknowledgments

References

American Psychiatric Association. (2013). Diagnostic and statistical manual of mental disorders (5th ed., DSM–5). Washington, DC: Author.

Bailey IL, & Lovie JE (1976). New design principles for visual acuity letter charts. American Journal of Optometry and Physiological Optics, 53, 740–745. http://dx.doi.org/10.1097/00006324-197611000-00006 [PubMed: 998716]

Barker-Collo S, & Feigin V (2006). The impact of neuropsychological deficits on functional stroke outcomes. Neuropsychology Review, 16, 53–64. http://dx.doi.org/10.1007/s11065-006-9007-5 [PubMed: 16967344]

Bates B, Choi JY, Duncan PW, Glasberg JJ, Graham GD, Katz RC, … the U.S. Department of Defense, & the Department of Veterans Affairs. (2005). Veterans Affairs/Department of Defense clinical practice guideline for the management of adult stroke rehabilitation care: Executive summary. Stroke, 36, 2049–2056. http://dx.doi.org/10.1161/01.STR.0000180432.73724.AD [PubMed: 16120847]

Bauer PJ, Dikmen SS, Heaton RK, Mungas D, Slotkin J, & Beaumont JL (2013). III. NIH Toolbox Cognition Battery (CB): Measuring episodic memory. Monographs of the Society for Research in Child Development, 78, 34–48. http://dx.doi.org/10.1111/mono.12033 [PubMed: 23952201]

Benedict RHB (1997). Brief Visuospatial Memory Test–Revised: Professional manual. Lutz, FL: Psychological Assessment Resources, Inc.

Benedict RHB, Schretlen D, Groninger L, Dobraski M, & Shpritz B (1996). Revision of the brief visuospatial memory test: Studies of normal performance, reliability, and validity. Psychological Assessment, 8, 145–153. http://dx.doi.org/10.1037/1040-3590.8.2.145

Black SE (2011). Vascular cognitive impairment: Epidemiology, sub-types, diagnosis and management. The Journal of the Royal College of Physicians of Edinburgh, 41, 49–56. http://dx.doi.org/10.4997/JRCPE.2011.121 [PubMed: 21365068]

Campbell DT, & Fiske DW (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. Psychological Bulletin, 56, 81–105. http://dx.doi.org/10.1037/h0046016 [PubMed: 13634291]

Caplan LR, Schmahmann JD, Kase CS, Feldmann E, Baquis G, Greenberg JP, … Hier DB (1990). Caudate infarcts. Archives of Neurology, 47, 133–143. http://dx.doi.org/10.1001/archneur.1990.00530020029011 [PubMed: 2405818]

Carlozzi NE, Goodnight S, Casaletto KB, Goldsmith A, Heaton RK, Wong AWK, … Tulsky DS (2017). Validation of the NIH Toolbox (NIHTB) in individuals with neurologic disorders. Archives of Clinical Neuropsychology, 32, 555–573. [PubMed: 28334392]

Carlozzi NE, Kirsch NL, Kisala PA, & Tulsky DS (2015). An examination of the Wechsler Adult Intelligence Scales, Fourth Edition (WAIS-IV) in individuals with complicated mild, moderate and Severe traumatic brain injury (TBI). Clinical Neuropsychology, 29, 21–37.

Carlozzi NE, Tulsky DS, Chiaravalloti ND, Beaumont JL, Weintraub S, Conway K, & Gershon RC (2014). NIH Toolbox Cognitive Battery (NIHTB-CB): The NIHTB Pattern Comparison Processing Speed Test. Journal of the International Neuropsychological Society, 20, 630–641. http://dx.doi.org/10.1017/S1355617714000319 [PubMed: 24960594]

Carlozzi NE, Tulsky DS, Kail RV, & Beaumont JL (2013). VI. NIH Toolbox Cognition Battery (CB): Measuring processing speed. Monographs of the Society for Research in Child Development, 78, 88–102. http://dx.doi.org/10.1111/mono.12036 [PubMed: 23952204]

Casaletto KB, Umlauf A, Beaumont J, Gershon R, Slotkin J, Akshoomoff N, & Heaton RK (2015). Demographically corrected normative standards for the English version of the NIH Toolbox Cognition Battery. Journal of the International Neuropsychological Society, 21, 378–391. http://dx.doi.org/10.1017/S1355617715000351 [PubMed: 26030001]

Cohen J (1977). Statistical power analysis for the behavioral sciences. New York, NY: Academic Press.

Cohen J (1992). A power primer. Psychological Bulletin, 112, 155–159. http://dx.doi.org/10.1037/0033-2909.112.1.155 [PubMed: 19565683]

Crowe SF (2000). Does the letter number sequencing task measure anything more than digit span? Assessment, 7, 113–117. http://dx.doi.org/10.1177/107319110000700202 [PubMed: 10868248]

de Bruijn MA, Synhaeve NE, van Rijsbergen MW, de Leeuw FE, Jansen BP, & de Kort PL (2014). Long-term cognitive outcome of ischaemic stroke in young adults. Cerebrovascular Diseases, 37, 376–381. http://dx.doi.org/10.1159/000362592 [PubMed: 24970603]

Delis DC, Kaplan E, & Kramer JH (2001). Delis Kaplan Executive Function System (D-KEFS). San Antonio, TX: The Psychological Corporation.

DeLuca J, Chelune GJ, Tulsky DS, Lengenfelder J, & Chiaravalloti ND (2004). Is speed of processing or working memory the primary information processing deficit in multiple sclerosis? Journal of Clinical and Experimental Neuropsychology, 26, 550–562. http://dx.doi.org/10.1080/13803390490496641 [PubMed: 15512942]

Desmond DW (2004). The neuropsychology of vascular cognitive impairment: Is there a specific cognitive deficit? Journal of the Neurological Sciences, 226, 3–7. http://dx.doi.org/10.1016/j.jns.2004.09.002 [PubMed: 15537510]

Desrosiers J, Demers L, Robichaud L, Vincent C, Belleville S, Ska B, & the BRAD Group. (2008). Short-term changes in and predictors of participation of older adults after stroke following acute care or rehabilitation. Neurorehabilitation and Neural Repair, 22, 288–297. http://dx.doi.org/10.1177/1545968307307116 [PubMed: 17916657]

Dikmen SS, Bauer PJ, Weintraub S, Mungas D, Slotkin J, Beaumont JL, … Heaton RK (2014). Measuring episodic memory across the lifespan: NIH Toolbox Picture Sequence Memory Test. Journal of the International Neuropsychological Society, 20, 611–619. http://dx.doi.org/10.1017/S1355617714000460 [PubMed: 24960230]

Donders J, Tulsky DS, & Zhu J (2001). Criterion validity of new WAIS-II subtest scores after traumatic brain injury. Journal of the International Neuropsychological Society, 7, 892–898. [PubMed: 11771633]

Dunn LM, & Dunn DM (2007). Peabody Picture Vocabulary Test(4th ed.). Minneapolis, MN: NCS Pearson.

Edwards DF, Hahn M, Baum C, & Dromerick AW (2006). The impact of mild stroke on meaningful activity and life satisfaction. Journal of Stroke and Cerebrovascular Diseases, 15, 151–157. http://dx.doi.org/10.1016/j.jstrokecerebrovasdis.2006.04.001 [PubMed: 17904068]

Enderby P, & Crow E (1996). Frenchay Aphasia Screening Test: Validity and comparability. Disability and Rehabilitation: An International, Multidisciplinary Journal, 18, 238–240. http://dx.doi.org/10.3109/09638289609166307

Enderby PM, Wood VA, Wade DT, & Hewer RL (1987). The Frenchay Aphasia Screening Test: A short, simple test for aphasia appropriate for non-specialists. International Rehabilitation Medicine, 8, 166–170. http://dx.doi.org/10.3109/03790798709166209 [PubMed: 2440825]

Ferris FL, III, & Bailey I (1996). Standardizing the measurement of visual acuity for clinical research studies: Guidelines from the Eye Care Technology Forum. Ophthalmology, 103, 181–182. http://dx.doi.org/10.1016/S0161-6420(96)30742-2 [PubMed: 8628551]

Gershon RC, Cella D, Fox NA, Havlik RJ, Hendrie HC, & Wagster MV (2010). Assessment of neurological and behavioural function: The NIH Toolbox. The Lancet Neurology, 9, 138–139. http://dx.doi.org/10.1016/S1474-4422(09)70335-7 [PubMed: 20129161]

Gershon RC, Cook KF, Mungas D, Manly JJ, Slotkin J, Beaumont JL, & Weintraub S (2014). Language measures of the NIH Toolbox Cognition Battery. Journal of the International Neuropsycho-logical Society, 20, 642–651. http://dx.doi.org/10.1017/S1355617714000411

Gershon RC, Slotkin J, Manly JJ, Blitz DL, Beaumont JL, Schnipke D, … Weintraub S (2013). IV. NIH Toolbox Cognition Battery (CB): Measuring language (vocabulary comprehension and reading decoding). Monographs of the Society for Research in Child Development, 78, 49–69. http://dx.doi.org/10.1111/mono.12034 [PubMed: 23952202]

Go AS, Mozaffarian D, Roger VL, Benjamin EJ, Berry JD, Borden WB, … the American Heart Association Statistics Committee and Stroke Statistics Subcommittee. (2013). Heart disease and stroke statistics—2013 update: A report from the American Heart Association. Circulation, 127, e6–e245. http://dx.doi.org/10.1161/CIR.0b013e31828124ad [PubMed: 23239837]

Godefroy O, Fickl A, Roussel M, Auribault C, Bugnicourt JM, Lamy C, … Petitnicolas G (2011). Is the Montreal Cognitive Assessment superior to the Mini-Mental State Examination to detect poststroke cognitive impairment? A study with neuropsychological evaluation. Stroke, 42, 1712–1716. http://dx.doi.org/10.1161/STROKEAHA.110.606277 [PubMed: 21474808]

Gold JM, Carpenter C, Randolph C, Goldberg TE, & Weinberger DR (1997). Auditory working memory and Wisconsin Card Sorting Test performance in schizophrenia. Archives of General Psychiatry, 54, 159–165. http://dx.doi.org/10.1001/archpsyc.1997.01830140071013 [PubMed: 9040284]

Goldstein LB, Bertels C, & Davis JN (1989). Interrater reliability of the NIH stroke scale. Archives of Neurology, 46, 660–662. http://dx.doi.org/10.1001/archneur.1989.00520420080026 [PubMed: 2730378]

Gontkovsky ST, & Beatty WW (2006). Practical methods for the clinical assessment of information processing speed. International Journal of Neuroscience, 116, 1317–1325. http://dx.doi.org/10.1080/00207450500516537 [PubMed: 17000532]

Haut MW, Kuwabara H, Leach S, & Arias RG (2000). Neural activation during performance of number-letter sequencing. Applied Neuropsychology, 7, 237–242. http://dx.doi.org/10.1207/S15324826AN0704_5 [PubMed: 11296686]

Hawkins KA (1998). Indicators of brain dysfunction derived from graphic representations of the WAIS-III/WMS-III Technical Manual clinical samples data: A preliminary approach to clinical utility. Clinical Neuropsychologist, 12, 535–551. http://dx.doi.org/10.1076/clin.12.4.535.7236

Heaton RK, Akshoomoff N, Tulsky D, Mungas D, Weintraub S, Dikmen S, … 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. http://dx.doi.org/10.1017/S1355617714000241 [PubMed: 24960398]

Heaton RK, Miller SW, Taylor JT, & Grant I (2004). Revised comprehensive norms for an expanded Halstead-Reitan Battery: Demographically adjusted neuropsychological norms for African American and Caucasian adults. Lutz, FL: Psychological Assessment Resources, Inc.

Hochstenbach JB, den Otter R, & Mulder TW (2003). Cognitive recovery after stroke: A 2-year follow-up. Archives of Physical Medicine and Rehabilitation, 84, 1499–1504. http://dx.doi.org/10.1016/S0003-9993(03)00370-8 [PubMed: 14586918]

Hochstenbach J, Mulder T, van Limbeek J, Donders R, & Schoonderwaldt H (1998). Cognitive decline following stroke: A comprehensive study of cognitive decline following stroke. Journal of Clinical and Experimental Neuropsychology, 20, 503–517. http://dx.doi.org/10.1076/jcen.20.4.503.1471 [PubMed: 9892054]

Holdnack JA, Tulsky DS, Brooks BL, Slotkin J, Gershon R, Heinemann AW, & Iverson GL (2017). Interpreting patterns of low scores on the NIH Toolbox Cognition Battery. Archives of Clinical Neuropsychology, 32, 574–584. http://dx.doi.org/10.1093/arclin/acx032 [PubMed: 28419177]

Kwah LK, & Diong J (2014). National Institutes of Health Stroke Scale (NIHSS). Journal of Physiotherapy, 60, 61 http://dx.doi.org/10.1016/j.jphys.2013.12.012 [PubMed: 24856948]

National Institutes of Health & Northwestern University. (2017). NIH Toolbox for Assessment of Neurological and Behavioral Function Administrator's Manual. Retrieved from http://assistly-production.s3.amazonaws.com/228622/kb_article_attachments/111250/NIH_Toolbox_App_Administrator%27s_Manual_v1.11_original.pdf?AWSAccessKeyId=AKIAJNSFWOZ6ZS23BMKQ&Expires=1499103624&Signature=bTL4D7os3ZJHNc0CcqCeyFfWL9g%3D&response-content-disposition=filename%3D%22NIH_Toolbox_App_Administrator%27s_Manual_v1.11.pdf%22&response-content-type=application%2Fpdf

Reuben DB, Magasi S, McCreath HE, Bohannon RW, Wang YC, Bubela DJ, … Gershon, R. C. (2013). Motor assessment using the NIH Toolbox. Neurology, 80(11, Suppl. 3), S65–S75. http://dx.doi.org/10.1212/WNL.0b013e3182872e01 [PubMed: 23479547]

Salter K, Jutai J, Foley N, Hellings C, & Teasell R (2006). Identification of aphasia post stroke: A review of screening assessment tools. Brain Injury, 20, 559–568. http://dx.doi.org/10.1080/02699050600744087 [PubMed: 16754281]

Salter K, Teasell R, Bitensky J, Foley N, & Bhogal S (2008). Evidence-based review of stroke rehabilitation: 12. Cognitive disorders and apraxia. Retrieved from http://www.ebrsr.com/uploads/cognition-SREBR-13.pdf

Schendel K, Dronkers NF, & Turken AU (2016). Not just language: Persisting lateralized visuospatial impairment after left hemisphere stroke. Journal of the International Neuropsychological Society, 22, 695–704. http://dx.doi.org/10.1017/S1355617716000515 [PubMed: 27329557]

Sonoda S, Chino N, Domen K, & Saitoh E (1997). Changes in impairment and disability from the third to the sixth month after stroke and its relationship evaluated by an artificial neural network. American Journal of Physical Medicine & Rehabilitation, 76, 395–400. http://dx.doi.org/10.1097/00002060-199709000-00010 [PubMed: 9354494]

Sörös P, Harnadek M, Blake T, Hachinski V, & Chan R (2015). Executive dysfunction in patients with transient ischemic attach and minor stroke. Journal of the Neurological Sciences, 354, 17–20. http://dx.doi.org/10.1016/j.jns.2015.04.022 [PubMed: 25979637]

Strauss E, Sherman EMS, & Spreen O (2006). A compendium of neuropsychological tests: Administration, norms, and commentary (Third Edition). New York, NY: Oxford University Press.

Su CY, Wuang YP, Lin YH, & Su JH (2015). The role of processing speed in post-stroke cognitive dysfunction. Archives of Clinical Neuropsychology, 30, 148–160. http://dx.doi.org/10.1093/arclin/acu057 [PubMed: 25618135]

Tulsky DS, & Heinemann AW (2017). The clinical utility and construct validity of the NIH Toolbox Cognition Battery (NIHTB-CB) in individuals with disabilities. Rehabilitation Psychology, 62, 409–412. [PubMed: 29265861]

Tulsky DS, Carlozzi NE, Chevalier N, Espy KA, Beaumont JL, & Mungas D (2013). V. NIH Toolbox Cognition Battery (CB): Measuring working memory. Monographs of the Society for Research in Child Development, 78, 70–87. http://dx.doi.org/10.1111/mono.12035 [PubMed: 23952203]

Tulsky DS, Carlozzi N, Chiaravalloti ND, Beaumont JL, Kisala PA, Mungas D, … 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. http://dx.doi.org/10.1017/S135561771400040X [PubMed: 24959983]

Tulsky DS, Saklofske DH, & Zhu J (2003). Revising a standard: An evaluation of the origin and development of the WAIS-III In Tulsky DS, Saklofske DH, Chelune GJ, Heaton RK, Ivnik RJ, Bornstein R, … Ledbetter MF (Eds.), Clinical interpretation of the WAIS-III and WMS-III (pp. 43–92). San Diego, CA: Academic Press http://dx.doi.org/10.1016/B978-012703570-3/50006-7

Wade DT, Wood VA, & Hewer RL (1985). Recovery after stroke: The first 3 months. Journal of Neurology, Neurosurgery, & Psychiatry, 48, 7–13. http://dx.doi.org/10.1136/jnnp.48.1.7

Wall KJ, Isaacs ML, Copland DA, & Cumming TB (2015). Assessing cognition after stroke. Who misses out? A systematic review. International Journal of Stroke, 10, 665–671. http://dx.doi.org/10.1111/ijs.12506 [PubMed: 25964002]

Wang YC, Magasi SR, Bohannon RW, Reuben DB, McCreath HE, Bubela DJ, … Rymer WZ (2011). Assessing dexterity function: A comparison of two alternatives for the NIH Toolbox. Journal of Hand Therapy, 24, 313–320. http://dx.doi.org/10.1016/j.jht.2011.05.001 [PubMed: 21798715]

Wechsler D (2008). Wechsler Adult Intelligence Scale IV. San Antonio, TX: Harcourt Assessment Inc.

White JH, Magin P, & Pollack MR (2009). Stroke patients' experience with the Australian health system: A qualitative study. Canadian Journal of Occupational Therapy, 76, 81–89. http://dx.doi.org/10.1177/000841740907600205

Wilkinson GS, & Robertson GJ (2006). Wide Range Achievement Test 4 professional manual. Lutz, FL: Psychological Assessment Resources.

Zelazo PD, Anderson JE, Richler J, Wallner-Allen K, Beaumont JL, Conway KP, … Weintraub S (2014). NIH Toolbox Cognition Battery (CB): Validation of executive function measures in adults. Journal of the International Neuropsychological Society, 20, 620–629. http://dx.doi.org/10.1017/S1355617714000472 [PubMed: 24960301]

Zelazo PD, Anderson JE, Richler J, Wallner-Allen K, Beaumont JL, & Weintraub S (2013). II. NIH Toolbox Cognition Battery (CB): Measuring executive function and attention. Monographs of the Society for Research in Child Development, 78, 16–33. http://dx.doi.org/10.1111/mono.12032 [PubMed: 23952200]

Impact and Implications

Although the National Institutes of Health (NIH) Toolbox Cognition Battery (NIHTB-CB) is useful across the life span (ages 3–85), little has been published to support its utility in clinical populations. This article highlights the clinical utility of the NIHTB-CB in individuals with stroke. This article includes data to support both the reliability and validity of the NIHTB-CB in individuals with stroke. This data suggest that the NIHTB-CB provides a reliable and valid assessment of the cognitive challenges that are associated with mild, moderate, and severe stroke. Such data are necessary precursors to utilizing the NIHTB-CB in clinical practice with individuals with stroke.

Table 1

National Institutes of Health Toolbox (NIHTB) Measure and Its Associated Standard Neuropsychological Comparison Measure

Cognitive subdomains	NIHTB measure	Construct validation measure
Episodic memory	Picture sequence memory	Auditory Verbal Learning Test (Rey), BVMT-R
Language	Picture vocabulary	PPVT-IV
	Oral Reading Recognition	WRAT-IV
Processing speed	Pattern comparison processing speed	WAIS-IV Digit Symbol, WAIS-IV Symbol Search
Working memory	List sorting working memory	WAIS-IV Letter-Number Sequencing
Executive function	Flanker inhibitory control and attention	DKEFS Interference
	Dimensional change card sort	

Note. BVMT-R = Brief Visuospatial Memory Test–Revised; PPVT-IV = Peabody Picture Vocabulary Test, Fourth Edition; WRAT-IV = Word Reading Achievement Test, Fourth Edition; WAIS-IV SS = Wechsler Adult Intelligence Scale Symbol Search, Fourth Edition; WAIS-IV LN = Wechsler Adult Intelligence Scale Letter Number Sequencing, Fourth Edition; DKEFS = Delis Kaplan Executive Functioning System.

Table 2

Missing Data

Variable	Stroke ($N = 131$)	Statistic	p
Missing 1 NIHTB-CB Score: %	14.5		
Mild ($N = 71$)	11.4	$\chi^2(1) = 1.55$	*ns*
Moderate/severe ($N = 60$)	19.2		
Number of missing scores, M (SD)	.39 (1.27)		
Mild ($N = 71$)	.33 (1.23)	$T(129) = -.67$	*ns*
Moderate/severe ($N = 60$)	.48 (1.34)		
% Participants with missing scores			
Picture vocabulary			
Mild	2.5	$\chi^2(1) = .05$	*ns*
Moderate/severe	1.9		
Oral reading			
Mild	5.1	$\chi^2(1) = .11$	*ns*
Moderate/severe	3.8		
Picture Sequence Memory			
Mild	7.6	$\chi^2(1) = 1.21$	*ns*
Moderate/severe	13.5		
List Sorting			
Mild	6.3	$\chi^2(1) = 1.92$	*ns*
Moderate/severe	13.5		
Pattern comparison			
Mild	6.3	$\chi^2(1) = .02$	*ns*
Moderate/severe	5.8		
Flanker			
Mild	2.5	$\chi^2(1) = .18$	*ns*
Moderate/severe	3.8		
Dimensional Change Card Sort			
Mild	2.5	$\chi^2(1) = .90$	*ns*
Moderate/severe	5.8		

Note. NIHTB-CB = National Institutes of Health Toolbox—Cognition Battery; *ns* = not significant ($p > .05$).

Table 3

Demographic Characteristics for Stroke Groups

Variable	Mild stroke ($N = 71$)	Moderate/severe stroke ($N = 60$)
Age (years)		
$M(SD)$	57.59 (13.09)	57.37 (11.82)
Time since injury (months)		
$M(SD)$	31.30 (11.96)	31.69 (11.57)
Gender (%)		
Male	49.4	53.8
Female	50.6	46.2
Race (%)		
Caucasian	53.2	36.5
African American	46.8	63.5
Ethnicity (%)		
Not Hispanic or Latino	97.4	96.2
Hispanic or Latino	2.6	3.8
Not Provided	2.5	.0
Education (%)		
$M(SD)$	13.47 (2.40)	13.10 (2.64)
Work status (%)[a]		
Full-time	38.2	25.0
Part-time	11.8	31.3
Volunteer	5.9	.0
Not employed	44.1	43.8

Note. No group differences were found among/between groups for all variables (all $p > .05$); categorical variables were examined by using chi-squared tests; continuous variables were examined by using independent sample t tests.

[a] This information was only consistently collected at a single study site.

Table 4

National Institutes of Health (NIH) Toolbox (NIHTB) Scores and Univariate Analyses for Individuals With Mild Versus Moderate/Severe Stroke

NIHTB scores	N	Mild stroke M (SD)	N	Moderate/severe stroke M (SD)	F	η_p^2
Composite scores						
Fluid	71	42.71 (12.64)	42	34.00 (9.57)	14.90**(3.25)	.12 (.03)
Crystallized	75	50.54 (11.73)	50	45.72 (10.85)	5.39*	.04
Subtest scores						
Picture vocabulary	77	50.65 (12.77)	51	47.08 (11.38)	2.62	.02
Oral Reading Recognition	75	50.10 (10.18)	50	45.22 (10.76)	6.58*	.05
Picture Sequence Memory Test	73	45.86 (12.96)	45	39.58 (11.34)	7.61**	.06
Pattern comparison	74	45.10 (11.25)	49	38.05 (9.13)	13.38**(1.76)	.10 (.02)
List sorting	74	45.70 (10.26)	45	42.21 (10.86)	3.09	.03
Flanker	77	44.74 (10.89)	50	37.76 (9.53)	13.73**(1.05)	.10 (.01)
DCCS	77	44.31 (10.11)	49	38.94 (8.83)	9.32**(1.52)	.07 (.01)

Note. DCCS = Dimensional Change Card Sort. Values provided in parentheses were based on analyses that included motor function as a covariate.

*
p < .05.

**
p < .01.

Table 5

Descriptive Statistics for Comparison Neuropsychological Measures

Standard neuropsychological measures	Mild stroke			Moderate/severe stroke			F	η_p^2
	N	M	SD	N	M	SD		
AVLT (Rey)	71	98.06	17.58	45	88.37	14.83	9.41**	.08
BVMT-R	69	36.43	12.17	44	30.77	10.58	6.42*	.05
WRAT-IV	77	96.04	14.95	51	88.25	14.04	8.72**	.06
PPVT-IV	77	97.99	14.86	52	90.90	15.75	6.72*	.05
WAIS-IV Coding	77	8.10	3.19	50	5.74	2.36	20.19** (2.07)	.14 (.02)
WAIS-IV SS	77	8.34	3.27	50	5.74	2.84	21.15** (2.03)	.14 (.02)
WAIS-IV LN	64	17.94	4.36	44	14.05	5.93	15.44**	.13
DKEFS Interference	75	8.63	3.94	48	5.38	3.40	22.17**	.15

Note. AVLT (Rey) = Auditory Verbal Learning Test (Rey); BVMT-R = Brief Visuospatial Memory Test-Revised; PPVT-IV = Peabody Picture Vocabulary Test, Fourth Edition; WRAT-IV = Word Reading Achievement Test, Fourth Edition; WAIS-IV SS = Wechsler Adult Intelligence Scale Symbol Search, Fourth Edition; WAIS-IV Coding = Wechsler Adult Intelligence Scale Digit Symbol Coding, Fourth Edition; WAIS-IV LN = Wechsler Adult Intelligence Scale Letter Number Sequencing, Fourth Edition; DKEFS = Delis Kaplan Executive Functioning System; DCCS = Dimensional Change Card Sort. Values provided in parentheses were based on analyses that included motor function as a covariate.

*
$p < .05$.

**
$p < .01$.

Table 6

Correlations Between National Institutes of Health (NIH) Toolbox (NIHTB) Subtests and Its Relative Standard Neuropsychological Test

Cognitive domain of functioning	NIHTB score	Corresponding Standard Neuropsychological Test	N	Pearson r
Episodic memory	Picture Sequence Memory	Auditory Verbal Learning Test (Rey)	107	.52
		BVMT-R	104	.65
Language	Picture Vocabulary	PPVT-IV	126	.87
	Oral Reading Recognition	WRAT-IV	124	.88
Processing speed	Pattern Comparison	WAIS-IV Coding	121	.59 (.58)
		WAIS-IV SS	121	.67 (.66)
Working memory	List Sorting	WAIS-IV LN	101	.66
Executive functioning	Flanker	DKEFS Interference	119	.46 (.50)
	DCCS	DKEFS Interference	118	.54 (.40)

Note. BVMT-R = Brief Visuospatial Memory Test-Revised; PPVT-IV = Peabody Picture Vocabulary Test, Fourth Edition; WRAT-IV = Word Reading Achievement Test, Fourth Edition; WAIS-IV SS = Wechsler Adult Intelligence Scale Symbol Search, Fourth Edition; WAIS-IV Coding = Wechsler Adult Intelligence Scale Digit Symbol Coding, Fourth Edition; WAIS-IV LN = Wechsler Adult Intelligence Scale Letter Number Sequencing, Fourth Edition; DKEFS = Delis Kaplan Executive Functioning System; DCCS = Dimensional Change Card Sort. Correlations provided in parentheses included motor function as a covariate; all $p < .01$ unless noted.

Table 7

Convergent and Discriminant Validity for National Institutes of Health (NIH) Toolbox (NIHTB) Scores for Combined Stroke Sample

Cognitive domain of functioning	Episodic memory				Language				Processing speed				Working memory (LN)		Executive functioning (DKEFS)	
	(AVLT)		(BVMT)		(WRAT)		(PPVT)		(Coding)		(SS)					
	n	r	n	r	n	r	n	r	n	r	n	r	n	r	n	r
Episodic memory																
Picture sequencing	107	.45	104	.62	117	.29	117	.36	118	.50 (.48)	118	.54 (.52)	101	.38	113	.30
Language																
Oral Recognition Reading	112	.32	108	.54	124	.88	124	.77	122	.47 (.48)	122	.48 (.48)	106	.66	120	.48
Picture vocabulary	113	.40	110	.61	125	.74	126	.87	124	.47 (.45)	124	.52 (.51)	106	.57	120	.48
Processing speed																
Pattern comparison	109	.28 (.23)	107	.38 (.36)	121	.25 (.24)	122	.38 (.38)	121	.59 (.57)	121	.67 (.66)	103	.35 (.32)	116	.47 (.44)
Working memory																
List sorting	107	.40	105	.60	119	.61	118	.58	119	.50 (.49)	119	.54 (.54)	102	.64	114	.47
Executive functioning																
DCCS	111	.45 (.41)	109	.57 (.55)	123	.51 (.50)	124	.55 (.55)	123	.60 (.57)	123	.65 (.64)	104	.49 (.47)	118	.54 (.51)
Flanker	112	.33 (.28)	110	.50 (.48)	124	.37 (.36)	125	.50 (.50)	124	.61 (.58)	124	.61 (.59)	105	.41 (.38)	119	.46 (.43)

Note. AVLT = Auditory Verbal Learning Test (Rey); BVMT = Brief Visuospatial Memory Test, Fourth Edition; WRAT = Word Reading Achievement Test; WRAT = Word Reading Achievement Test-Revised; WRAT = Word Reading Achievement Test; PPVT = Peabody Picture Vocabulary Test, Fourth Edition; Coding = Wechsler Adult Intelligence Scale Digit Symbol Coding, Fourth Edition; SS = Wechsler Adult Intelligence Scale Symbol Search, Fourth Edition; LN = Wechsler Adult Intelligence Scale Letter Number Sequencing, Fourth Edition; DKEFS = Delis Kaplan Executive Functioning System; DCCS = Dimensional Change Card Sort. All $p < .01$; correlations provided in parentheses included motor function as a covariate.

Table 8

Effect Sizes (Cohen's d) and Rates of Clinical Impairment for Individuals With Stroke

NIHTB scores	Mild stroke	Moderate/severe stroke	Clinical impairment rates for combined sample[a]
NIHTB subtest score			
Fluid	−.64	−1.64	42
Crystallized	.05	−.41	—
NIHTB subtest score			
Picture vocabulary	.06	−.27	—
Oral Reading Recognition	.01	−.46	—
Picture sequence memory	−.36	−.98	31.0%
Pattern comparison	−.46	−1.25	43.0%
List sorting	−.42	−.75	32.2%
Flanker	−.50	−1.25	36.6%
DCCS	−.57	−1.18	37.4%

Note. NIH = National Institutes of Health; NIHTB = NIH Toolbox; DCCS = Dimensional Change Card Sort. For effect sizes, the group means for the composite and subtest scores were compared to the NIHTB normative sample ($N = 972$; $M = 50$; $SD = 10$).

[a] Clinical impairment rates are based on the approach outlined in Holdnack et al., (2017) for calculating base rates of clinical impairment using the NIHTB.