



HHS Public Access

Author manuscript

Rehabil Psychol. Author manuscript; available in PMC 2019 April 14.

Published in final edited form as:

Rehabil Psychol. 2017 November ; 62(4): 413–424. doi:10.1037/rep0000174.

Using the NIH Toolbox Cognition Battery (NIHTB-CB) in Individuals with Traumatic Brain Injury

DS Tulsy¹, NE Carlozzi², J Holdnack¹, RK Heaton³, A Wong⁴, A Goldsmith⁵, and AW Heinemann⁵

¹Center for Health Assessment Research and Translation, University of Delaware.

²Center for Clinical Outcomes, Development and Application (CODA), University of Michigan.

³Department of Psychiatry, University of California, San Diego.

⁴Program in Occupational Therapy and Department of Neurology, Washington University in St. Louis.

⁵Center for Rehabilitation Outcomes Research, Shirley Ryan AbilityLab.

Abstract

Purpose/Objective: The NIH Toolbox for the Assessment of Neurological Behavior and Function Cognition Battery (NIHTB-CB) is a common data element (CDE) for use in individuals with traumatic brain injury (TBI). This study evaluates its sensitivity and specificity in distinguishing individuals with complicated mild, moderate, or severe TBI and provides support for the construct validity of the NIHTB-CB in individuals with TBI.

Research Method: One-hundred-eighty-two individuals with TBI (n=83 complicated mild/moderate; n=99 severe) completed the NIHTB-CB and neuropsychological criterion measures and complete data were obtained on 158 participants. A control sample of 158 individuals without known neurological impairment was extracted from the NIHTB-CB normative sample. Multivariate analyses of variance (MANOVAs) determined the sensitivity of the NIHTB-CB measures to TBI and injury severity (complicated mild/moderate TBI, severe TBI, and controls) on the demographically corrected NIHTB-CB composite scores and 7 NIHTB-CB subtests. A descriptive analysis of the sensitivity of each subtest was conducted. Finally, correlations between NIHTB-CB measures and criterion tests assessed convergent and discriminant validity.

Results: Multivariate analyses indicated that there is a main effect for group (complicated mild/moderate vs. severe vs. controls for fluid scores in the Toolbox as opposed to only marginally significant results for the verbal scores. Moderate to strong relationships were found between the NIHTB-CB measures and their corresponding neuropsychological measures (convergent validity) while much smaller correlations were found between measures of different cognitive domains (discriminant validity).

Conclusions: Findings provide initial evidence of construct validity and the clinical utility of the NIHTB-CB in individuals with TBI.

Keywords

TBI; Traumatic Brain Injury; cognitive assessment; neuropsychological assessment; psychometrics

Introduction

Several large-scale initiatives are underway to develop health outcome measures that utilize common data elements (CDEs) and allow within and cross-disease comparison in National Institutes of Health (NIH)-sponsored research. The value of utilizing CDEs has been highlighted by the NIH Interagency Common Data Elements Project, which was designed to identify appropriate CDEs in neurological research. As part of these efforts, the NIH Blueprint for Neuroscience Research (Baughman, 2006) supported the development of the Toolbox for the Assessment of Neurological and Behavioral Functioning (NIHTB) (Gershon et al., 2010; Gershon, Wagster, et al., 2013). The NIHTB is a set of integrated and co-normed tools to assess cognitive, emotional, motor, and sensory health across the lifespan. Since the NIHTB was developed for use in the general population, its construct validity and clinical utility for assessing neurocognitive impairments in clinical disorders or individuals with disabilities has not been established. Validation is especially relevant in traumatic brain injury (TBI), where several efforts have been devoted to establishing standardized definitions and protocols for TBI research through collaboration of multiple agencies and data sharing (Thurmond et al., 2010). Although validation research for the NIHTB in individuals with TBI is limited, the NIH Interagency Common Data Elements Project included it as a recommended CDE for TBI studies (E.A. Wilde et al., 2010).

TBI is an important and costly health problem (Ma, Chan, & Carruthers, 2014), and the clinical validation of the NIHTB in individuals with TBI is particularly important. The TBI population includes many previously healthy young people. Neurocognitive deficits are among the many physiological, medical, psychological, and psychosocial problems affecting individuals with TBI (S. S. Dikmen et al., 2014a). Cognitive tests used with neurological populations should be sensitive to the impact of TBI on neurocognitive functioning.

The cognitive sequelae of TBI are well documented. Therefore, the performance of individuals with TBI on the NIHTB is an important indicator of the measures' construct validity and clinical utility. In particular, individuals with TBI are at risk for decrements in attention (Donders & Levitt, 2012; Sinclair, Ponsford, Rajaratnam, & Anderson, 2013), working memory (Tulsky et al., 2014; D. S. Tulsky et al., 2013), episodic memory (Carlozzi, Grech, & Tulsky, 2013; Carlozzi, Kirsch, Kisala, & Tulsky, 2015; Donders & Levitt, 2012; Wright et al., 2014), executive functioning (Crowe & Crowe, 2013; Donders & Levitt, 2012; Merkley, Larson, Bigler, Good, & Perlstein, 2013), fluid reasoning (Carlozzi, Kirsch, et al., 2015), and processing speed (Carlozzi, Beaumont, Tulsky, & Gershon, 2015; Donders & Strong, 2016; Madigan, DeLuca, Diamond, Tramontano, & Averill, 2000; Sinclair et al., 2013). Language abilities, in contrast, are generally spared or less affected (Carlozzi, Kirsch, et al., 2015). In general, improvements in cognition are typically seen within the first two years post-injury although, in many people, there is not a full return to baseline functioning (S. Dikmen, Machamer, Temkin, & McLean, 1990; S. Dikmen, Reitan, & Temkin, 1983; Schretlen & Shapiro, 2003).

The degree of cognitive change and level of impairment after TBI vary as a function of the severity of injury (Carlozzi, Tulsky, Kail, & Beaumont, 2013; Rohling, Meyers, & Millis,

2003). More severe injuries result in greater cognitive impairment (S. Dikmen et al., 1990; S. S. Dikmen, Machamer, Powell, & Temkin, 2003), while mild injuries result in small effects (S. Dikmen et al., 1990; Schretlen & Shapiro, 2003). In addition, individuals with complicated mild TBI perform similarly on neuropsychological tests to those with moderate severity (Kashluba, Hanks, Casey, & Millis, 2008b). However, differences among injury groups (e.g., those with complicated mild, moderate, and severe injuries) can be affected by other factors (e.g., time since injury and premorbid functional level (Novack, Bush, Meythaler, & Canupp, 2001)).

There is also significant variability in cognitive performance among individuals with similar injury severity (Iverson, Holdnack, & Lange, 2013). Some individuals with relatively severe injuries may have average or better test scores, whereas individuals with less severe injuries may have below expected cognitive performance. Furthermore, at the individual level, some examinees, at all injury severity levels, have cognitive difficulties while others do not.

The NIHTB was developed and normed in a large, United States (US) census matched sample taken from the general population who did not have known neurocognitive deficits. Research on the seven NIHTB Cognition Battery (NIHTB-CB) subtests (Picture Sequence Memory test (Bauer et al., 2013; S. S. Dikmen et al., 2014b), Picture Vocabulary Test (Bauer et al., 2013; S. S. Dikmen et al., 2014b; Gershon et al., 2014; Gershon, Slotkin, et al., 2013), Oral Reading Recognition Test (Gershon et al., 2014; Gershon, Slotkin, et al., 2013) Pattern Comparison Processing Speed Test (Carlozzi, Beaumont, et al., 2015; Carlozzi et al., 2014; Carlozzi, Tulsky, et al., 2013), List Sorting Working Memory Test (Tulsky et al., 2014; D.S. Tulsky et al., 2013)), Dimensional Change Card Sort test (Zelazo et al., 2013; Zelazo et al., 2014), and Flanker Inhibitory Control and Attention test (Zelazo et al., 2013; Zelazo et al., 2014)), as well as 3 summary scores (Fluid Cognition, Crystallized Cognition, and Total Cognition Composite scores) (Akshoomoff et al., 2013; R. K. Heaton et al., 2014) have established reliability and validity for use in individuals across the lifespan. Table 1 lists the NIHTB-CB measures and their “gold standard” counterparts that serve as criterion measures to establish construct validity. The initial validation work also supported a multidimensional structure of cognition with episodic memory, vocabulary, reading, working memory, and a processing speed/executive functioning abilities, suggesting that NIHTB-CB subtests assess different domains of cognitive functioning, and provided initial evidence of convergent and discriminant validity among NIHTB-CB subtests (Mungas et al., 2014; Mungas et al., 2013). However, the clinical utility of the NIHTB for use with individuals with TBI has yet to be established.

The current study evaluated the sensitivity of the NIHTB-CB in distinguishing individuals with medically confirmed diagnoses of complicated mild, moderate, or severe TBI. We examined the severity of cognitive deficits by comparing cognitive performance between individuals with TBI and matched controls. We hypothesized that individuals with TBI would perform worse than matched controls on all NIHTB-CB measures EXCEPT those that examine language/crystallized abilities (Picture Vocabulary Test, Oral Reading Recognition Test); we expected NIHTB-CB measures of crystallized abilities to show much less (or no) impairment, thereby serving as “hold” measures (Larrabee, Largen, & Levin, 1985). We hypothesized that, at a group level, individuals with severe TBI would

demonstrate greater cognitive impairments than individuals with less severe injuries (e.g., those with complicated mild/moderate TBI). We examined the relationship between the NIHTB-CB and neuropsychological measures to evaluate convergent and discriminant validity in individuals with TBI. The overall goals of this paper are to describe the NIHTB-CB sensitivity to TBI severity and to evaluate the suitability of its use in the assessment of individuals with TBI.

Methods

Sample Population/Participant Recruitment

One-hundred-eighty-two individuals with a medically documented complicated mild/moderate TBI (N=83) or severe TBI (N=99) were examined as part of a larger study examining cognitive functioning in individuals with disabilities. We combined complicated mild TBI with moderate TBI (see below for definitions of severity of injury) on the basis of previous work suggesting that long-term outcomes for individuals with complicated mild more closely resemble individuals with moderate TBI than they do mild TBI (Kashluba, Hanks, Casey, & Millis, 2008a). Participants were recruited from three rehabilitation centers: Rehabilitation Institute of Chicago (RIC), Washington University in St Louis (WU), and the University of Michigan (UM) and required to come into the center on one occasion to complete a series of cognitive tests and patient reported outcomes measures. Each institution received approval from their local Institutional Review Boards that provided oversight for the research. To be eligible for the study, participants were at least 18 years old, admitted to a hospital within 24 hours of a TBI, able to comprehend and speak English at a 5th grade level, and at least one year since their brain injury. Individuals with additional cognitive impairments due to other conditions such as a psychiatric disorder, Alzheimer's disease and other dementing illnesses were excluded.

Participants were recruited according to their injury severity. TBI severity was classified according to the lowest emergency department GCS score within the first 24 hours after injury (not due to intubation, sedation, or intoxication); a GCS score of 8 was classified as severe injury, and a score of 9–12 was classified as moderate injury (Teasdale & Jennett, 1974). Complicated mild cases had a GCS score of 13–15 with neuroimaging showing acute brain abnormality consistent with TBI, such as subarachnoid hemorrhaging or cortical contusions (Williams, Levin, & Eisenberg, 1990). If no GCS was available, TBI cases were classified based on the detailed description of their injury and confirmed by a neuropsychologist independent of performance on the NIHTB-CB or neurocognitive tests (see Figure 1).

We also examined a demographically matched control sample of 158 adults from the NIHTB standardization sample (which is comprised of 1038 English-speaking adults); participants were matched on key demographic variables. The NIHTB standardization sample was weighted to be demographically representative of the 2010 US Census data. Our subset of these participants was matched to the total TBI group on key demographic variables (i.e., age, education level, ethnicity, and gender).

Measures

Participants completed a larger eight-hour battery of assessments including the NIH Toolbox (including all four domains: Cognition, Sensory, Motor, and Emotional functioning) and neuropsychological tests. An examiner who was trained and certified according to a well-established protocol (described below) administered all cognitive tests.

The NIH Toolbox of Neurological and Behavioral Functioning is a set of assessment tools for measuring cognitive, emotional, motor, and sensory health which are appropriate for a wide range of ages (3–85 years), and settings (Gershon et al., 2010; Gershon, Wagster, et al., 2013). The NIHTB-CB contains subtests to measure episodic memory, language, processing speed, working memory, and executive function (Table 1) (Weintraub et al., 2013; Weintraub et al., 2014). Three composite measures were derived from combinations of individual subtests: overall cognition, crystallized cognition, and fluid cognition (R. K. Heaton et al., 2014). The Crystallized and Fluid Cognition Composite scores were derived from subsets of the battery; Picture Vocabulary and Oral Reading comprised the Crystallized Cognition Composite and Picture Sequence Memory, List Sorting, Pattern Comparison, Flanker Inhibitory Control and Attention Test, and Dimensional Change Card Sort comprised the Fluid Cognition Composite, respectively. The Overall Cognition Composite was the average of the Crystallized and Fluid Composites. Test administration time is 30 minutes.

“Gold Standard” cognitive and neuropsychological tests include common data element (CDE) tools that have widespread use in the TBI clinical and research applications. These measures were used in the original NIHTB validation efforts (Weintraub et al., 2014) and were selected based on their ability to inform diagnosis, outcome measurement and prediction, and treatment effectiveness research (E. A. Wilde et al., 2010). Table 1 includes a detailed description of these tests.

Data Collection Training and Assurance to Standardized Procedures

This multisite study required data collectors to follow standardized protocols. Each examiner practiced the administration procedures by administering them to a minimum of 5 cases. After the initial training, they were observed and certified by one of two authors of this paper (NC and DT) in a live testing session. We recertified examiners annually to ensure that tests continued to be administered in a standardized manner. We also monitored test scoring. One of the authors (NC) reviewed 10 test protocols from each examiner, rescored the protocol, and provided feedback about deviations from standard procedures. In cases where scoring agreement did not exceed 95%, the examiner was retrained and the scoring for 10 new cases was reviewed. Scoring was reviewed annually as part of the recertification process. All protocols were double-scored to minimize scoring errors.

Normative Standards

For the NIH Toolbox Cognition Battery (NIHTB-CB), demographically-corrected normative standards were developed in a cohort of neurological healthy adults (N=1038) in order to determine deviations from expected levels of performances. Details regarding these norms are described 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, using multiple fractional polynomials to account for any non-linear effects). The residuals from these models were corrected in order to enhance the homogeneity of the variances across demographic characteristics (i.e., age, gender, education, race/ethnicity). The corrected residuals were standardized and rescaled as T-scores. The resulting T-score (mean=50, SD=10) for each test represents an individual's neurocognitive performance compared to age-, education-, gender-, and race/ethnicity-matched peers. Demographically-corrected norms demonstrate better accuracy with less demographic bias (Casaletto et al., 2015; R. K. Heaton, Ryan, L., & Grant, I., 2009; Taylor & Heaton, 2001) on cognitive measures than age-only adjusted scores in neurological populations. The use of demographic adjustments also minimizes effects of slight differences in demographic characteristic on group comparisons.

Data Analysis

Analyses focused on participants with complete data on all seven NIHTB-CB tests, slightly reducing the sample size. The total number of missing tests did not differ by group (e.g., Mild Complicated/Moderate, Severe, Controls), the exclusion rate for missing data was identical for control (13.2%) and TBI samples (13.2%), and there was no effect of injury severity ($\chi^2(2, N = 364) = 0.73, p > .05$) on missing data. Moreover, there was no difference in missing values by age, race, or gender; however, there was a significant association between missing data and education ($\chi^2(3, N = 364) = 9.74, p < .05$). Missing data occurred at a rate of 20% for individuals with 12 or fewer years of education and 9% for those with at least some college education. Picture Sequence Memory (7.2%) and Vocabulary (6.3%) were the most frequently missing data and Oral Reading the least frequent (1.9%). Rates of missing test data were significantly higher in the TBI sample for Dimensional Change Card Sort (6.6% vs. 0.0%; $\chi^2(1) = 10.36, p < .01$); Flanker Inhibitory Control and Attention Test (6.0% vs. 0.0%; $\chi^2(1) = 9.32, p < .01$), and Pattern Comparison (7.1% vs. 0.6%; $\chi^2(1) = 8.93, p < .01$). Results of univariate ANOVAs evaluating the interaction effect of group (i.e., Mild Complicated/Moderate, Severe, Controls) by exclusion status did not reveal a significant interaction for any of the individual tests. The presence of missing data was not specific to TBI (relative to the control sample) nor more impaired individuals within the TBI groups.

We conducted two multivariate analyses to determine the sensitivity of the NIHTB-CB measures to TBI and injury severity. The first multivariate analysis evaluated group (complicated mild/moderate TBI, severe TBI and controls) differences on the two demographically adjusted NIHTB-CB composite scores (Fluid Cognition and Crystallized Cognition). The second analysis evaluated group (complicated mild/moderate TBI, severe TBI and controls) differences on the seven NIHTB-CB subtests (Picture Sequence Memory Test, Picture Vocabulary, Oral Reading Recognition, Pattern Comparison, List Sorting, the Dimensional Change Card Sort (DCCS); and the Flanker Inhibitory Control and Attention Test).

A descriptive analysis of the sensitivity of each subtest is presented to provide clinicians and researchers information regarding impairment rates in the clinical samples compared to the matched control sample. We applied a cut-off of one standard deviation unit on the demographically-corrected t-scores (e.g., 40) to detect “abnormal” performance. These scores represent the probability of an examinee showing a performance deficit based on injury status.

Pearson correlations were used to examine relationships among the uncorrected scores on the NIHTB measures and established neuropsychological measures to assess convergent and discriminant validity. To determine convergent validity, the uncorrected scores for the NIHTB-CB subtests were correlated with the scores on the corresponding “gold standard” neuropsychological measure of the same construct (e.g., NIHTB-CB Oral Reading Recognition with the Wide Range Achievement Test-4th edition); moderate to high correlations would provide evidence of convergent validity. Additional evidence of convergent validity was obtained by examining the correlation of the NIHTB-CB test with established measures from the same cognitive domain (e.g., NIHTB Oral Reading Test correlated with other measures within the language domain). Moderate to high correlation coefficients would provide evidence of convergent validity. Divergent validity was evaluated by correlating the NIHTB-CB measures with established tests that measure different cognitive domains (e.g., Picture Sequence Memory or other fluid cognition tests with the Wide Range Achievement Test-4th edition) where low correlation coefficients were expected. All domain correlations were created by converting the correlation of the NIH Toolbox test with each criterion measure within a domain into a Fisher’s z' -score (Cohen, 2003; Fisher, 1958). We then averaged the z' -scores within domain and then converting the z' -scores back into r equivalents. Using Fisher’s z' reduces estimation bias of the population correlation compared to averaging Pearson r 's (Corey, 1998). Correlations less than 0.4 were considered poor evidence of convergent validity, 0.4 – 0.6 were adequate evidence of convergent validity, and 0.6 or greater were good evidence of convergent validity; correlations less than 0.3 between measures of different constructs was evidence of discriminant validity (Campbell & Fiske, 1959). For group comparisons, we report effect sizes (Cohen’s D), with cutoffs of .20, .50, and .80 indicating small, medium, and large effects, respectively.

Results

The demographic characteristics of the TBI subgroups and the NIHTB-CB standardization matched control sample are presented in Table 2. Groups did not differ on gender, $\chi^2(2, N=318) = 0.124, p > .05$, race, $\chi^2(12, N=318) = 1.99, p > .05$, ethnicity, $\chi^2(2, N=317) = 2.28, p > .05$, or education, $F(2, 313) = 1.64, p > .05$. There were significant group differences on age, $F(2, 313) = 6.39, p < .01$ and time since injury, $F(1, 156) = 6.46, p < .05$. The complicated mild/moderate TBI subgroup was significantly older (mean = 43.5; range = 18–78) than the severe TBI subgroup (mean = 34.3; range = 18–67), and the complicated mild/moderate TBI subgroup (mean = 4.7; range=1 to 21 years) had significantly less time since injury than the severe TBI subgroup (mean = 6.8; range= 1to 29 years).

Since both composite scores and subtest scores yield clinically meaningful information, we report two sets of analyses, one with composite scores and one with subtest scores, recognizing that these analyses are not independent. We also report analyses separately for complicated mild/moderate TBI vs. controls and severe TBI vs. controls.

Multivariate Analyses (Complicated Mild/Moderate TBI vs. Severe TBI vs. Matched Controls)

The first multivariate analysis examined group (complicated mild/moderate vs. severe vs. controls) and the two NIHTB demographically-corrected composite scores as the dependent variables. Pillai's Trace indicated a statistically significant main effect of group, $F(4, 626) = 17.04$, $p < .0001$, Pillai's Trace = .197, partial $\eta^2 = .098$. Univariate analyses and *post hoc* comparisons (Scheffe') indicated that the complicated mild/moderate TBI and severe TBI group performed worse than controls on the Fluid Cognition Composite, and that the severe TBI group performed worse than the complicated mild/moderate TBI group on the Fluid Cognition Composite (See Table 3). While the Crystallized Cognition Composite was marginally significantly different overall between the groups ($F(2,313) = 4.89$, $p < .01$), *post hoc* analysis did not reveal statistically significant differences between groups.

The second multivariate analysis examined group (complicated mild/moderate vs. severe vs. controls) and seven NIHTB demographically-corrected subtest scores as the dependent variables (Picture Sequence Memory Test, Picture Vocabulary, Oral Reading Recognition, Pattern Comparison, List Sorting, DCCS, and the Flanker Inhibitory Control and Attention test). Pillai's Trace indicated a statistically significant main effect of group, $F(14, 614) = 5.57$, $p < .0001$, Pillai's Trace = .225, partial $\eta^2 = .113$. Univariate analyses indicated that there were significant group differences on all the fluid subtests (Picture Sequence Memory Test, Pattern Comparison, List Sorting, Flanker Inhibitory Control and Attention Test and DCCS); the severe TBI group performed worse than controls in all comparisons, and the complicated mild/moderate group performed worse than controls on Picture Sequence Memory Test, List Sorting, and Flanker Inhibitory Control and Attention Test and DCCS. There were also group differences between the two TBI groups for Picture Sequence Memory Test, Pattern Comparison, and Flanker Inhibitory Control and Attention Test, with the severe group performing worst (See Table 3). On crystallized subtests, which typically are the tests more resilient and less sensitive to TBI, the Oral Reading Test ($F(2,313) = 4.66$, $p < .05$) and Picture Vocabulary Test ($F(2,313) = 3.45$, $p < .05$) were marginally significantly different at the group level. Regardless, *post hoc* analysis did not reveal statistically significant differences between specific groups.

Table 3 presents the effect sizes comparing each clinical group with the matched control group for each demographically-corrected subtest and composite score. There are large effects between the severe TBI group and matched controls for the Fluid Composite, Picture Sequence Memory, and the Flanker Inhibitory Control and Attention Test. The complicated mild/moderate group had moderate effect sizes vs. controls on these variables and did not have any large effects.

Identification of Cognitive Impairment.

Table 4 presents base rates of examinees obtaining low scores at a cut off which is roughly equivalent to a percentile of 16th (*i.e.*, T-score = 40) for complicated mild/moderate TBI, severe TBI, and control groups. Sensitivity and specificity values are presented for each test and for the Crystallized and Fluid Composite scores. The highest sensitivity is observed for Picture Sequence Memory at a cut-off of T-score = 40, .36 for complicated mild/moderate vs. controls and .52 for severe TBI vs. controls at a 13% false positive rate. The Fluid composite shows similar values to Picture Sequence Memory. As expected, language based measures are minimally sensitive to TBI compared to expected levels of performance. The severe TBI sample shows a greater proportion of low scores on the Oral Reading and Picture Vocabulary measures compared to controls indicating that a subset of examinees with severe injuries may have language impairments. Applying a T-score cut-off of = 40, yields false positive rates in an acceptable range of 11–16%. These sensitivity and specificity results reflect effect sizes of 1 or less.

Convergent and Discriminant Validity

Table 5 displays Pearson correlation coefficients for NIH Toolbox core measures compared to analogous established neuropsychological tests and by domain in the TBI sample. The largest correlations (.83 and .80) were observed with Oral Reading Recognition and WRAT-4 Reading and with Picture Vocabulary and PPVT, providing evidence of convergent validity. These measures also correlate highly with and averaged correlation created using the verbal gold standard measures (PPVT and WRAT-4 Reading; see Table 5 for a detailed description of the average language and executive function correlations). The lowest correlations (.13 and .10) were observed between the Pattern Comparison and Dimensional Change Card Sorting Test with language measures. The NIHTB-CB measures of Fluid Cognition are highly correlated in this sample. There were moderate correlations between List Sorting and the gold standard working memory measures, as well as Picture Sequence Memory and processing speed measures. Similarly, the Flanker Inhibitory Control and Attention Test and DCCS tests have moderate correlations with both the established executive functioning measures as well as tests of processing speed, working memory, and episodic memory. Of particular note, the Flanker Inhibitory Control and Attention Test and DCCS have higher correlations with processing speed, which may reflect the importance of timed performance in the NIHTB-CB measures of executive functioning. Finally, the Pattern Comparison and Picture Sequence Memory are correlated moderately with measures across most domains; however, the highest correlations were with the gold standard measures that measure the same constructs of processing speed and episodic memory, respectively.

Discussion

This study provides evidence of validity of the NIHTB-CB in adults with TBI. The strengths of the study include a large clinical sample of well-defined TBI severities. Multi-center data collection insured greater representation of individuals from different rehabilitation centers. The study also included rigorous training of psychometricians for proper administration and data integrity.

The primary goal of this study was to establish the validity of the NIHTB-CB in a sample of individuals with documented TBI. Validation is necessary to provide empirical support for the use of this battery of cognitive tests in research and clinical applications. In order to use the NIHTB-CB in clinical settings, the specific tests and composite indices should measure the appropriate constructs (e.g., language/crystallized abilities, working memory, episodic memory, processing speed, and executive functioning). Moreover, the sensitivity of cognitive tests to the effects of TBI is an important litmus test, as the impact of TBI on cognition is well documented. At the same time, it is known that a significant subset of the TBI population recovers quite well.

Comparison of performance on the NIH Toolbox tests and composite scores in a sample of individuals with documented complicated mild/moderate TBI, severe TBI, and a control sample, provide strong support for the NIHTB-CB battery to detect impairment. We expected that individuals with TBI would have lower scores on the Picture Memory, Pattern Comparison, Flanker Inhibitory Control and Attention Test, Dimensional Change Card Sort, and List Sorting Task compared to controls (Donders & Levitt, 2012; Donders, Tulsky, & Zhu, 2001; Willmott, Ponsford, Hocking, & Schonberger, 2009), and that most individuals with TBI would not have poor scores on tests of language and crystalized abilities. We also analyzed the data using fully demographically-corrected scores (Casaletto et al., 2015). The results are consistent with the expectation of differences on NIHTB-CB variables when compared with controls or by injury severity.

Memory complaints and impairment are hallmark symptoms of TBI and for this reason, a new memory measure should demonstrate sensitivity to the effects of TBI to establish construct validity and clinical sensitivity. As expected, the Picture Sequence Memory Test was the most sensitive measure of the NIHTB-CB to the effects of TBI and TBI severity. Over 50% of the severe and 30% of the complicated mild/moderate TBI groups had scores in the impaired range. These results are similar to those obtained in other tests of episodic memory such as the fourth version of the Wechsler Memory Scale (WMS-IV) (Carlozzi, Grech, et al., 2013; Iverson et al., 2013). The comparisons of performance between individuals with complicated mild/moderate, severe, and controls provides evidence that the Picture Sequence Memory Test functions in a manner similar to other widely used memory measures.

On the test of working memory (List Sorting), we observed lower scores for the complicated mild and severe TBI groups compared to controls with moderate and large effect sizes, respectively. In the control group, 11% had low scores, compared to 27% and 32% of the complicated mild/moderate and severe TBI, respectively. The level of performance is again consistent to the results found with other tests of working memory such as the WAIS-IV and WMS-IV Working Memory Indices (Carlozzi, Grech, et al., 2013; Carlozzi, Kirsch, et al., 2015; Donders & Strong, 2015).

Processing speed impairment represents another domain of cognitive functioning that is a well-documented sequela of TBI and, along with memory functioning, often yields the largest effect sizes (Iverson et al., 2013). The Pattern Comparison test is a choice reaction time test which significantly differentiates TBI from control groups and was sensitive to

severity of injury. However, while the Pattern Comparison test is sensitive to the effects of TBI, individuals in the current sample had higher performance scores compared to what would have been expected with the WAIS-III (Donders et al., 2001) and the WAIS-IV Processing Speed Indices (Carlozzi, Beaumont, et al., 2015; Carlozzi, Kirsch, et al., 2015; Donders & Strong, 2015). Unfortunately, the legacy measures were not administered in the normative study, so no direct comparison is possible.

Executive functioning is a broad category of behaviors defined by behavioral and cognitive changes often observed after brain injury, and the NIHTB-CB provides two measures of executive functioning, the Flanker Inhibitory Control and Attention Test and Dimensional Change Card Sorting tests. These tests use a choice reaction response process and response time is a significant component of the scores. Both tests differentiated the TBI from control groups; however, only the Flanker test was sensitive to severity effects. Both tests were sensitive to TBI vs. control, second only to Picture Sequence Memory. Unlike Picture Sequence Memory, both Flanker Inhibitory Control and Attention Test and Dimensional Change Card Sort tests require a fast performance and processing speed.

Finally, on tests of language and crystallized abilities, the mild/moderate TBI group was not significantly different from the control group while the severe TBI group only showed a trend for lower reading scores (Carlozzi, Kirsch, et al., 2015; Iverson et al., 2013). Base rate tables show that 16% of controls will have low scores on the Oral Reading Test; by comparison, 25% of individuals with severe injuries have low scores at the 1 SD cutoff. These results illustrate that only a small subgroup of severely injured examinees will exhibit language/reading problems greater than 1 year post injury. While Picture Vocabulary did not show statistically significant differences between the severe TBI and control groups, the level of performance relative to Oral Reading is nearly identical (48.3 vs. 48.4). The rates of low scores are similar as well (25% vs. 20%). This lack of difference between the tests has clinical relevance as clinicians should anticipate that the Oral Reading Test and the Picture Vocabulary test will yield similar estimates of premorbid abilities in the severe TBI group. The results observed in this study are consistent with previous studies using similar, language-based “crystallized” measures (Carlozzi, Kirsch, et al., 2015).

This study also sought to establish the convergent and discriminant validity of the NIHTB-CB with established neuropsychological measures in a clinical sample. Cognitive tests are comprised of related constructs (Holdnack & Weiss, 2006; Tulsky & Price, 2003) and many of the subtests will have moderate to high correlations with each other. Convergent and discriminant validity studies are necessary to understand the degree to which a given test measures the specific construct of interest and the degree that non-construct related variance (e.g., processing speed and executive functioning) affects test performance. The analyses comparing most NIHTB-CB tests to “gold standard” measures provides strong evidence of convergent and discriminant validity in most instances. Each NIHTB-CB test correlated with similar measures. The correlations in the second column in Table 5 show the convergent validity coefficient and range from 0.42 to 0.83 (median = 0.68). As expected, we observe high correlations between language tests (e.g., 0.80 and 0.83) which is consistent with previous findings showing correlations between language measures. High within-construct correlations were observed for episodic memory and processing speed tests (e.g., 0.68 and

0.69, respectively). Significant correlations in the moderate range were obtained between the working memory tasks (0.56) and between the executive functioning tests (0.46 and 0.42 in magnitude). Correlations between executive functioning tasks in other studies also have tended to be in the moderate range, around the magnitude of the coefficients obtained in this study. While the NIHTB-CB Executive Functioning tests appear to be tapping the construct for which they were developed, they also reflect processing speed. Considering that processing speed is a frequent deficit in individuals with TBI, it is not surprising that executive function tests based on reaction time also would have a high correlation with measures of processing speed. Collectively, most NIHTB-CB subtests are correlated with similar measures suggesting convergent validity of the NIHTB-CB tests in a TBI sample. The columns on the right of Table 5 display correlations with the construct (as measured in an aggregate of tests of this construct). As described above, the individual test generally has a relatively high correlation with the aggregate score and lower correlations with tests of different constructs.

Study Limitations

While the presence of a TBI in this study was confirmed with medical records, making this a well worked up sample, examinees had to be willing and able to complete extensive testing that lasted 8-hours and spanned two days (to complete the NIHTB, criteria measures, and a series of other tests). As a result, the sample may not be representative of the population of individuals living with TBI. We did not exclude volunteers if they had participated in other clinical or research evaluations which conceivably may have included neuropsychological tests; exposure to tests may affect their performance. Additionally, we used the original dual-screen monitor format. It is unclear if the results will completely generalize to the newer iPad version. Finally, the study did not include a control sample that completed the established neuropsychological measures as did the clinical sample. Rather, the control sample was derived from the NIHTB-CB standardization study sample and these individuals did not complete most of the concurrent validity measures. Having parallel research methodology would have allowed us to compare the sensitivity of NIHTB-CB measures along with the established gold standards. Such data would have strengthened this study. In the absence of these data, we are not able to estimate the degree to which correlations between the measures were inflated due to injury severity effects or attenuated due to range restriction.

Application of NIHTB-CB

This study provides evidence that the NIHTB-CB tests are sensitive to the effects of TBI and severity of injury. The level of performance is consistent with that reported in studies of established measures, though the NIHTB-CB Pattern Comparison Processing Speed test may slightly underestimate impairment. The crystallized and fluid composites provide good estimates of intact vs. impaired functions post-injury. With multiple measures in each domain (Crystallized vs Fluid), provide a robust estimate of cognitive functioning (Brooks, Holdnack, & Iverson, 2011). Furthermore, with only single measures of key cognitive functions, it may be insufficient to identify deficits related to specific cognitive domains (e.g. executive dysfunction compared to the more general “fluid” category) or by skills within a domains.

We obtained moderate rates of missing data that were not random but associated with education level. The Picture Sequence Memory and Picture Vocabulary tests were the most often missing measures. The types of tests that were missing differed by TBI group: tests requiring rapid responses were most likely to be missing. The impact of missing data in research and clinical settings must be considered when using the NIHTB-CB (Magasi, Harniss, Tulsky, Cohen, & Heaton, Under Review).

Of critical importance in clinical and research settings evaluating individuals with TBI is the need to identify invalid responses (Heilbronner et al., 2009; Larrabee, 2012). In its current configuration, the NIHTB-CB does not have methods for estimating performance invalidity. The absence of invalidity indicators might result in scores below expected for individuals in which secondary gain may be an issue (Slick, Sherman, & Iverson, 1999).

One of the primary advantages of using the NIHTB-CB is its relatively short administration time. The battery requires only 30 minutes to complete. Brevity allows evaluation of a range of abilities without fatiguing examinees. This strength may be a limitation when researchers or clinicians require information about detailed aspects of cognition, such as identifying retrieval vs. encoding deficits or differentiating slow vs. errors in performance.

In conclusion, findings support the construct validity of the NIHTB-CB in community-dwelling adults with TBI. We observed group differences and patterns of correlations that are consistent with previous studies and these results provide evidence of convergent-discriminant validity for the NIHTB-CB language measures and good convergent validity for the List Sorting, Pattern Comparison and Picture Memory tests. Most participants could complete most tests. The NIHTB-CB is a promising CDE candidate.

References

- Akshoomoff N, Beaumont JL, Bauer PJ, Dikmen SS, Gershon RC, Mungas D, ... Heaton RK (2013). VIII. NIH Toolbox Cognition Battery (CB): composite scores of crystallized, fluid, and overall cognition. *Monogr Soc Res Child Dev*, 78(4), 119–132. doi:10.1111/mono.12038 [PubMed: 23952206]
- Bauer PJ, Dikmen SS, Heaton RK, Mungas D, Slotkin J, & Beaumont JL (2013). III. NIH Toolbox Cognition Battery (CB): measuring episodic memory. *Monogr Soc Res Child Dev*, 78(4), 34–48. doi:10.1111/mono.12033 [PubMed: 23952201]
- Baughman R, Farkas R, Guzman M, & Huerta MF. (2006). The National Institutes of Health Blueprint for Neuroscience Research. *J Neurosci*, 26(41), 10329–10331. [PubMed: 17035514]
- Brooks BL, Holdnack JA, & Iverson GL (2011). Advanced clinical interpretation of the WAIS-IV and WMS-IV: prevalence of low scores varies by level of intelligence and years of education. *Assessment*, 18(2), 156–167. doi:10.1177/1073191110385316 [PubMed: 20947705]
- Campbell DT, & Fiske DW (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, 56(2), 81–105. [PubMed: 13634291]
- Carlozzi NE, Beaumont JL, Tulsky DS, & Gershon RC (2015). The NIH Toolbox Pattern Comparison Processing Speed Test: Normative Data. *Arch Clin Neuropsychol*, 30(5), 359–368. doi:10.1093/arclin/acv031 [PubMed: 26025230]
- Carlozzi NE, Grech J, & Tulsky DS (2013). Memory functioning in individuals with traumatic brain injury: an examination of the Wechsler Memory Scale-Fourth Edition (WMS-IV). *J Clin Exp Neuropsychol*, 35(9), 906–914. doi:10.1080/13803395.2013.833178 [PubMed: 24033318]
- 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). *Clin Neuropsychol*, 29(1), 21–37. doi: 10.1080/13854046.2015.1005677 [PubMed: 25646823]
- 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. *J Int Neuropsychol Soc*, 20(6), 630–641. doi:10.1017/s1355617714000319 [PubMed: 24960594]
- Carlozzi NE, Tulsky DS, Kail RV, & Beaumont JL (2013). VI. NIH Toolbox Cognition Battery (CB): measuring processing speed. *Monogr Soc Res Child Dev*, 78(4), 88–102. doi: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. *J Int Neuropsychol Soc*, 21(5), 378–391. doi:10.1017/S1355617715000351 [PubMed: 26030001]
- Cohen J, Cohen P, West SG, & Aiken LS . (2003). *Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences* (3rd ed.). Mahweh, NJ: Lawrence Earlbaum Associates, Publishers.
- Corey DM, Dunlap WP ... & Burke MJ (1998). Averaging correlations: Expected values and bias in combined Pearson rs and Fisher's z transformations. *Journal of General Psychology*, 125, 245–261.
- Crowe SF, & Crowe LM (2013). Does the presence of posttraumatic anosmia mean that you will be disinhibited? *J Clin Exp Neuropsychol*, 35(3), 298–308. doi:10.1080/13803395.2013.771616 [PubMed: 23432111]
- Dikmen S, Machamer J, Temkin N, & McLean A (1990). Neuropsychological recovery in patients with moderate to severe head injury: 2 year follow-up. *J Clin Exp Neuropsychol*, 12(4), 507–519. doi:10.1080/01688639008400997 [PubMed: 2211973]
- Dikmen S, Reitan RM, & Temkin NR (1983). Neuropsychological recovery in head injury. *Arch Neurol*, 40(6), 333–338. [PubMed: 6847436]
- Dikmen SS, Bauer PJ, Weintraub S, Mungas D, Slotkin J, Beaumont JL, ... Heaton RK (2014a). Measuring episodic memory across the lifespan: NIH Toolbox Picture Sequence Memory Test. *Journal of the International Neuropsychological Society*, 20(6), 611–619. doi:10.1017/S1355617714000460 [PubMed: 24960230]
- Dikmen SS, Bauer PJ, Weintraub S, Mungas D, Slotkin J, Beaumont JL, ... Heaton RK (2014b). Measuring episodic memory across the lifespan: NIH Toolbox Picture Sequence Memory Test. *J Int Neuropsychol Soc*, 20(6), 611–619. doi:10.1017/S1355617714000460 [PubMed: 24960230]
- Dikmen SS, Machamer JE, Powell JM, & Temkin NR (2003). Outcome 3 to 5 years after moderate to severe traumatic brain injury. *Arch Phys Med Rehabil*, 84(10), 1449–1457. [PubMed: 14586911]
- Donders J, & Levitt T (2012). Criterion validity of the neuropsychological assessment battery after traumatic brain injury. *Arch Clin Neuropsychol*, 27(4), 440–445. doi:10.1093/arclin/acs043 [PubMed: 22532153]
- Donders J, & Strong CA (2015). Clinical utility of the Wechsler Adult Intelligence Scale-Fourth Edition after traumatic brain injury. *Assessment*, 22(1), 17–22. doi:10.1177/1073191114530776 [PubMed: 24752385]
- Donders J, & Strong CA (2016). Latent Structure of the Behavior Rating Inventory of Executive Function-Adult Version (BRIEF-A) After Mild Traumatic Brain Injury. *Arch Clin Neuropsychol*, 31(1), 29–36. doi:10.1093/arclin/acv048 [PubMed: 26209247]
- Donders J, Tulsky DS, & Zhu J (2001). Criterion validity of new WAIS-II subtest scores after traumatic brain injury. *J Int Neuropsychol Soc*, 7(7), 892–898. [PubMed: 11771633]
- Fisher RA (1958). *Statistical Methods for Research Workers* (13th ed.). Edinburgh: Oliver & Boyd.
- Gershon RC, Cella D, Fox NA, Havlik RJ, Hendrie HC, & Wagster MV (2010). Assessment of neurological and behavioural function: the NIH Toolbox. *Lancet Neurol*, 9(2), 138–139. doi: 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. *J Int Neuropsychol Soc*, 20(6), 642–651. doi:10.1017/S1355617714000411 [PubMed: 24960128]

- 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). *Monogr Soc Res Child Dev*, 78(4), 49–69. doi:10.1111/mono.12034 [PubMed: 23952202]
- Gershon RC, Wagster MV, Hendrie HC, Fox NA, Cook KF, & Nowinski CJ (2013). NIH Toolbox for Assessment of Neurological and Behavioral Function. *Neurology*, 80(Suppl 3), S2–S6. [PubMed: 23479538]
- 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. *J Int Neuropsychol Soc*, 20(6), 588–598. doi:10.1017/S1355617714000241 [PubMed: 24960398]
- Heaton RK, Ryan L, & Grant I (2009). Demographic influences and use of demographically corrected norms in neuropsychological assessment. In I. G. K. A. (Eds.) (Ed.), *Neuropsychological assessment of neuropsychiatric and neuromedical disorders* (pp. 127–155). New York, NY: Oxford University Press.
- Heilbronner RL, Sweet JJ, Morgan JE, Larrabee GJ, Millis SR, & Conference, P. (2009). American Academy of Clinical Neuropsychology Consensus Conference Statement on the neuropsychological assessment of effort, response bias, and malingering. *Clin Neuropsychol*, 23(7), 1093–1129. doi:10.1080/13854040903155063 [PubMed: 19735055]
- Holdnack J, & Weiss LW (2006). WISC-IV integrated: Beyond the essentials. In Weiss LW, Prifitera A, Saklofske DH, & Holdnack J (Eds.), *WISC-IV: Advanced Clinical Interpretation* San Diego: Academic Press, Inc.
- Iverson GL, Holdnack J, & Lange RT (2013). Using the WAIS-IV/WMS-IV/ACS following moderate-severe traumatic brain injury. In Holdnack J, Drozdick L, Weiss LG, & Iverson GL (Eds.), *WAIS-IV/WMS-IV/ACS: Advanced Clinical Interpretation* San Diego, CA: Elsevier Science.
- Kashluba S, Hanks RA, Casey JE, & Millis SR (2008a). Neuropsychologic and functional outcome after complicated mild traumatic brain injury. *Archives of Physical Medicine & Rehabilitation*, 89(5), 904–911. [PubMed: 18452740]
- Kashluba S, Hanks RA, Casey JE, & Millis SR (2008b). Neuropsychologic and functional outcome after complicated mild traumatic brain injury. *Arch Phys Med Rehabil*, 89(5), 904–911. [PubMed: 18452740]
- Larrabee GJ (2012). Performance validity and symptom validity in neuropsychological assessment. *J Int Neuropsychol Soc*, 18(4), 625–630. [PubMed: 23057079]
- Larrabee GJ, Largen JW, & Levin HS (1985). Sensitivity of age-decline resistant (“hold”) WAIS subtests to Alzheimer’s disease. *J Clin Exp Neuropsychol*, 7(5), 497–504. doi: 10.1080/01688638508401281 [PubMed: 4066906]
- Ma VY, Chan L, & Carruthers KJ (2014). Incidence, prevalence, costs, and impact on disability of common conditions requiring rehabilitation in the United States: stroke, spinal cord injury, traumatic brain injury, multiple sclerosis, osteoarthritis, rheumatoid arthritis, limb loss, and back pain. *Arch Phys Med Rehabil*, 95(5), 986–995 e981. doi:10.1016/j.apmr.2013.10.032 [PubMed: 24462839]
- Madigan NK, DeLuca J, Diamond BJ, Tramontano G, & Averill A (2000). Speed of information processing in traumatic brain injury: modality-specific factors. *J Head Trauma Rehabil*, 15(3), 943–956. [PubMed: 10785624]
- Magasi M, Harniss M, Tulsky DS, Cohen ML, & Heaton RK (Under Review). Test Accommodations for Individuals with Neurological Conditions Completing the NIH Toolbox - Cognition Battery: An Evaluation of the Frequency and Appropriateness. *Rehabilitation Psychology*
- Merkley TL, Larson MJ, Bigler ED, Good DA, & Perlstein WM (2013). Structural and functional changes of the cingulate gyrus following traumatic brain injury: relation to attention and executive skills. *J Int Neuropsychol Soc*, 19(8), 899–910. doi:10.1017/S135561771300074X [PubMed: 23845701]
- Mungas D, Heaton R, Tulsky D, Zelazo PD, Slotkin J, Blitz D, ... Gershon R (2014). Factor structure, convergent validity, and discriminant validity of the NIH Toolbox Cognitive Health Battery (NIHTB-CHB) in adults. *J Int Neuropsychol Soc*, 20(6), 579–587. doi:10.1017/S1355617714000307 [PubMed: 24960474]

- Mungas D, Widaman K, Zelazo PD, Tulsky D, Heaton RK, Slotkin J, ... Gershon RC (2013). VII. NIH Toolbox Cognition Battery (CB): factor structure for 3 to 15 year olds. *Monogr Soc Res Child Dev*, 78(4), 103–118. doi:10.1111/mono.12037 [PubMed: 23952205]
- Novack TA, Bush BA, Meythaler JM, & Canupp K (2001). Outcome after traumatic brain injury: pathway analysis of contributions from premorbid, injury severity, and recovery variables. *Arch Phys Med Rehabil*, 82(3), 300–305. doi:10.1053/apmr.2001.18222 [PubMed: 11245749]
- Rohling ML, Meyers JE, & Millis SR (2003). Neuropsychological impairment following traumatic brain injury: a dose-response analysis. *Clin Neuropsychol*, 17(3), 289–302. doi:10.1076/clin.17.3.289.18086 [PubMed: 14704881]
- Schretlen DJ, & Shapiro AM (2003). A quantitative review of the effects of traumatic brain injury on cognitive functioning. *Int Rev Psychiatry*, 15(4), 341–349. doi:10.1080/09540260310001606728 FAW1HLRL4E1J0FFL [pii] [PubMed: 15276955]
- Sinclair KL, Ponsford JL, Rajaratnam SM, & Anderson C (2013). Sustained attention following traumatic brain injury: use of the Psychomotor Vigilance Task. *J Clin Exp Neuropsychol*, 35(2), 210–224. doi:10.1080/13803395.2012.762340 [PubMed: 23391455]
- Slick DJ, Sherman EM, & Iverson GL (1999). Diagnostic criteria for malingered neurocognitive dysfunction: proposed standards for clinical practice and research. *Clin Neuropsychol*, 13(4), 545–561. doi:10.1076/1385-4046(199911)13:04;1-Y;FT545 [PubMed: 10806468]
- Taylor MJ, & Heaton RK (2001). Sensitivity and specificity of WAIS-III/WMS-III demographically corrected factor scores in neuropsychological assessment. *J Int Neuropsychol Soc*, 7(7), 867–874. [PubMed: 11771630]
- Teasdale G, & Jennett B (1974). Assessment of Coma and Impaired Consciousness - Practical Scale. *Lancet*, 2(7872), 81–84. [PubMed: 4136544]
- Thurmond VA, Hicks R, Gleason T, Miller AC, Szuflita N, Orman J, & Schwab K (2010). Advancing integrated research in psychological health and traumatic brain injury: common data elements. *Arch Phys Med Rehabil*, 91(11), 1633–1636. doi:S0003-9993(10)00672-6 [pii] 10.1016/j.apmr.2010.06.034 [PubMed: 21044705]
- 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. *J Int Neuropsychol Soc*, 20(6), 599–610. doi:10.1017/S135561771400040X [PubMed: 24959983]
- Tulsky DS, Carlozzi NE, Chevalier N, Espy K, Beaumont J, & Mungas D (2013). NIH Toolbox Cognitive Function Battery (CFB): Measuring Working Memory. *Society For Research In Child Development, Monograph*, 78(4), 70–87.
- Tulsky DS, Carlozzi NE, Chevalier N, Espy KA, Beaumont JL, & Mungas D (2013). V. NIH Toolbox Cognition Battery (CB): measuring working memory. *Monogr Soc Res Child Dev*, 78(4), 70–87. doi:10.1111/mono.12035 [PubMed: 23952203]
- Tulsky DS, & Price LR (2003). The joint WAIS-III and WMS-III factor structure: development and cross-validation of a six-factor model of cognitive functioning. *Psychol Assess*, 15(2), 149–162. [PubMed: 12847775]
- Weintraub S, Dikmen SS, Heaton RK, Tulsky DS, Zelazo PD, Bauer PJ, ... Gershon RC (2013). Cognition assessment using the NIH Toolbox. *Neurology*, 80(11 Suppl 3), S54–64. doi:10.1212/WNL.0b013e3182872ded [PubMed: 23479546]
- Weintraub S, Dikmen SS, Heaton RK, Tulsky DS, Zelazo PD, Slotkin J, ... Gershon R (2014). The cognition battery of the NIH toolbox for assessment of neurological and behavioral function: validation in an adult sample. *J Int Neuropsychol Soc*, 20(6), 567–578. doi:10.1017/S1355617714000320 [PubMed: 24959840]
- Wilde EA, Whiteneck GG, Bogner J, Bushnik T, Cifu D, Dikmen S, ... von Steinbuechel N (2010). Recommendations for the use of common outcome measures in traumatic brain injury research. *Archives of Physical Medicine and Rehabilitation*, 91(11), 1650–1660.e1617. [PubMed: 21044708]
- Wilde EA, Whiteneck GG, Bogner J, Bushnik T, Cifu DX, Dikmen S, ... von Steinbuechel N (2010). Recommendations for the use of common outcome measures in traumatic brain injury research. *Arch Phys Med Rehabil*, 91(11), 1650–1660 e1617. doi:S0003-9993(10)00656-8 [pii] 10.1016/j.apmr.2010.06.033 [PubMed: 21044708]

- Williams DH, Levin HS, & Eisenberg HM (1990). Mild Head-Injury Classification. *Neurosurgery*, 27(3), 422–428. [PubMed: 2234336]
- Willmott C, Ponsford J, Hocking C, & Schonberger M (2009). Factors contributing to attentional impairments after traumatic brain injury. *Neuropsychology*, 23(4), 424–432. doi:10.1037/a0015058 [PubMed: 19586207]
- Wright MJ, Wong AL, Obermeit LC, Woo E, Schmitter-Edgecombe M, & Fuster JM (2014). Memory for performed and observed activities following traumatic brain injury. *J Clin Exp Neuropsychol*, 36(3), 268–277. doi:10.1080/13803395.2014.884543 [PubMed: 24524393]
- Zelazo PD, Andersen J, Richler J, Wallner-Allen K, Beaumont J, & Weintraub S (2013). NIH Toolbox Cognitive Function Battery (CFB): Measuring Executive Function and Attention. Society For Research In Child Development, Monograph, 78(4), 16–33.
- 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. *J Int Neuropsychol Soc*, 20(6), 620–629. doi:10.1017/S1355617714000472 [PubMed: 24960301]

Impact

1. Although, the NIH Toolbox Cognition Battery (NIHTB-CB) was designed to be a common data element to be utilized across NIH studies, the NIHTB-CB had not been validated for use in clinical populations nor had it been tested, previously, in individuals with Traumatic Brain Injury (TBI). This study addresses this research gap.
2. The strength of this study is that the NIHTB-CB was administered to a relatively large sample of community dwelling individuals with TBI (confirmed by medical records) along with a series of criterion measurement scales.
3. The manuscript presents the clinical utility, sensitivity and specificity, and construct validity of the NIHTB-CB. Additionally, base rate data are provided that can be used by rehabilitation psychologists and/or neuropsychologists to identify clinical impairments in cognition in clinical practice.
4. The data presented in this manuscript are essential to clinicians who wish to use the NIHTB-CB in clinical practice when testing individuals with TBI.

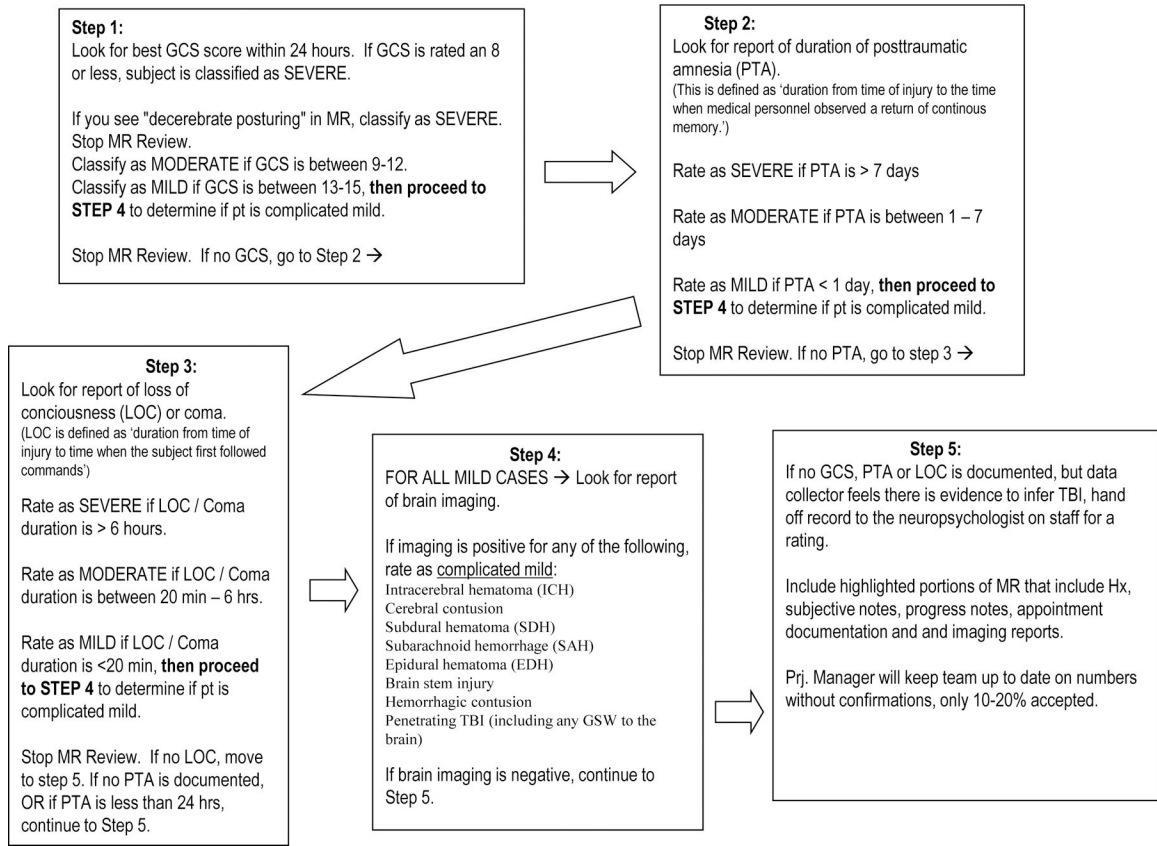


Figure 1. Classification Process of Severity of Traumatic Brain Injury

Table 1.

Description of NIH Toolbox Cognitive Domains/"Gold Standard" Neuropsychological Tests

Cognitive Subdomains	NIH Toolbox Task Name	"Gold Standard" Neuropsychological Tests	Description
Episodic Memory	Picture Sequence Memory	Rey Auditory Language Learning Test (RAVLT); Brief Visuospatial Memory Test-Revised (BVRT-R)	Mental processes involved in acquisition, learning, and retrieval of new information
Language	Picture Vocabulary; Oral Reading Recognition;	Peabody Picture Vocabulary Test- 4 th Edition (PPVT); Wide Range Achievement Test- 4 th Edition (WRAT-4) Reading Test	Mental processes that serve to translate thought into shared symbols (words, gestures) for the purpose of communication
Processing Speed	Pattern Comparison	Wechsler Adult Intelligence Scale, 4 th Edition Coding (CD) and Symbol Search (SS) subtests, Oral Symbol Digit Modalities Test	The amount of time to process a set of information or the amount of information processed within a certain unit of time
Working Memory	List Sorting	Wechsler Adult intelligence Scale, 4 th Edition Letter-Number Sequencing (LNS) Subtest; Paced Auditory Serial Addition Test (PASAT)	The capacity to 1) process information across a series of tasks and modalities; 2) hold information in a short-term buffer; 3) manipulate the information; and 4) hold the products in the same short-term buffer
Executive Function	Dimensional Change Card Sort; Flanker Inhibitory Control	Wisconsin Card Sorting Test- 64 (WCST); Delis- Kaplan Executive Function Scales Color/Word Interference Test (CWIT)	The capacities to plan, organize, and monitor the execution of behaviors that are strategically directed in a goal oriented manner

Table 2.

Demographic and TBI characteristics

Variable	Complicated Mild/Moderate (N=74)	Severe TBI (N=84)	Control (N=158)
Age(Years)			
M (SD)	43.5 (18.7)	34.3 (13.8)	39.2 (16.2)
Time Since Injury (years)			
M(SD)	4.7 (4.2)	6.8 (6.3)	
Gender(%)			
Male	60.8	63.1	61.4
Female	39.2	36.9	38.6
Race (%)			
Caucasian	78.4	75.0	76.6
African American	9.5	14.3	14.6
Other	12.1	10.7	8.2
Not Provided	0	0	0.6
Ethnicity (%)			
Not Hispanic or Latino	93.2	90.5	87.3
Hispanic or Latino	6.8	8.3	10.5
Not Provided	0	1.2	1.2
Education (%)			
Less than 12 years	8.1	19.0	14.6
12 years	18.9	17.9	15.8
13–15 years	35.1	34.5	35.4
16 or more years	37.8	28.6	34.2
Education (Years)			
M (SD)	14.2 (2.6)	13.5 (2.3)	14.0 (2.7)
Work Status (%)			
Full-Time	25.7	16.7	
Part-Time	21.6	23.8	
Volunteer	1.4	0.0	
Not Employed	50.0	54.8	
Unknown	1.4	4.8	100

Table 3.

NIHTB Full-Demographically adjusted scores and univariate analyses for individuals with complicated mild/moderate TBI, severe TBI, and matched controls

NIHTB Scores	Comp Mild/ Mod TBI	Severe TBI	Control	<i>F</i>	<i>p</i>	Comp Mild/ Mod TBI	Severe TBI
	Mean (SD)	Mean (SD)	Mean (SD)			<i>d</i>	<i>d</i>
NIHTB Composite Scores	N=74	N=84	N=158				
Crystalized	103.2 (15.9)	96.2 (14.7)	101.7 (15.2)	4.89	< .01	0.09	-0.35
Fluid ^{a,b,c}	92.2 (18.7)	82.4 (18.4)	101.6 (16.0)	34.44	< .001	-0.56	-1.14
NIHTB Subtest Scores	N=74	N=84	N=158				
Oral Reading Recognition ^b	103.2 (15.8)	96.4 (15.3)	102.0 (15.5)	4.69	< .05	0.08	-0.36
Picture Vocabulary ^b	102.9 (15.4)	97.1 (14.2)	101.5 (15.4)	3.45	< .05	0.09	-0.29
List Sorting ^{a,b}	96.3 (15.5)	91.2 (15.0)	102.1 (15.3)	14.52	< .001	-0.38	-0.71
Picture Sequence Memory ^{a,b,c}	92.5 (16.2)	84.2 (15.5)	100.2 (14.3)	31.64	< .001	-0.52	-1.08
Pattern Comparison ^{b,c}	96.7 (15.9)	89.8 (17.2)	101.8 (15.9)	15.02	< .001	-0.32	-0.75
Flanker ^{a,b,c}	94.0 (17.2)	86.2 (16.4)	100.8 (14.6)	24.11	< .001	-0.44	-0.94
DCCS ^{a,b}	94.9 (16.7)	90.1 (16.6)	101.1 (15.5)	13.48	< .001	-0.39	-0.69

note:

^a = complicated mild/moderate vs. control

^b =severe vs. control

^c =complicated mild/moderate vs. severe

^d = $(\text{mean}_{\text{TBI}} - \text{mean}_{\text{control}}) / \sqrt{((n_1 - 1) * (sd_1^2)) + ((n_2 - 1) * (sd_2^2)) / (n_1 + n_2 - 2)}$.

Table 4.
Base Rates of Low Scores and Sensitivity and Specificity for NIHTB-CB Tests

T-Score	40	Crystalized	Fluid	Oral Reading Recognition	Picture Vocabulary	List Sorting	Picture Sequence Memory	Pattern Comparison	Flanker	DCCS
Mild Complicated/Moderate (MCMO)	12.2	33.8	17.6	10.8	27.0	36.5	25.7	32.4	31.1	
Severe (SEV)	17.9	50.0	25.0	20.2	32.1	52.4	32.1	42.9	44.0	
Controls	12.0	12.7	16.5	13.9	11.4	12.7	12.7	15.2	12.0	
Sensitivity MCMO	0.12	0.34	0.18	0.11	0.27	0.36	0.26	0.32	0.31	
Sensitivity SEV	0.18	0.50	0.25	0.20	0.32	0.52	0.32	0.43	0.44	
Specificity	0.88	0.87	0.84	0.86	0.89	0.87	0.87	0.85	0.88	

Note: DCCS=Dimensional Category Card Sorting, MCMO=mild complicated/moderate TBI and SEV=severe TBI

Table 5: Pearson Correlations of NIH/TCB Tests with Established Neuropsychological Measures

NIH Tool/Box Test	r	Established Measure	Language	Working Memory	Episodic Memory	Processing Speed	Executive Functioning
Oral Reading Recognition	0.83	Wide Range Achievement Test-4th Edition Reading	0.75	0.47	0.27	0.30	-0.25
Picture Vocabulary	0.80	Peabody Picture Vocabulary Test-4th Edition	0.71	0.40	0.26	0.27	-0.20
List Sorting	0.56	WAIS-IV Letter Number Sequencing	0.28	0.54	0.58	0.59	-0.46
Picture Sequence Memory	0.68	Brief Visuospatial Memory Test-Revised	0.24	0.50	0.68	0.65	-0.40
Pattern Comparison	0.69	WAIS-IV Coding	0.13	0.45	0.49	0.69	-0.43
Flanker Inhibitory Control	-0.46	DKEFS Color-Word Interference - Inhibition	0.21	0.40	0.40	0.59	-0.44
DCCS	-0.42	Wisconsin Card Sort Test	0.10	0.48	0.48	0.55	-0.49

note: Sample size varies from 142 to 158. All correlations based on raw test scores. High scores on WCST perseverative errors and D-KEFS Inhibition time indicate poor performance. Correlations in last 5 columns are aggregates based on averages using Fisher's z' transformation. All correlations significant at $p < .05$ except Language domain with Pattern Comparison and Wisconsin Card Sort Test.

The way we averaged the two correlations is to transform each individual correlation into a z score using fisher's z-transformation. Then we added the 2 z-scores together and divided by 2. This gave us a new z-score. We then converted the new z-score back into the associated r value.