

Interpreting Patterns of Low Scores on the NIH Toolbox Cognition Battery

James A. Holdnack^{1,*}, David S. Tulsky², Brian L. Brooks^{3,4,5}, Jerry Slotkin¹, Richard Gershon^{6,7},
Allen W. Heinemann⁸, Grant L. Iverson^{9,10,11,12}

¹Center on Assessment Research and Translation, University of Delaware, USA

²Center on Assessment Research and Translation & Departments of Physical Therapy, Psychological and Brain Sciences,
University of Delaware, USA

³Neurosciences Program (Brain Injury and Rehabilitation), Alberta Children's Hospital, Alberta, Canada

⁴Departments of Paediatrics, Clinical Neurosciences, and Psychology, University of Calgary, Alberta, Canada

⁵Alberta Children's Hospital Research Institute, University of Calgary, Calgary, Alberta, Canada

⁶Department of Medical Social Sciences

⁷Division of Biomedical Informatics, Department of Preventive Medicine, Northwestern University Feinberg School of Medicine, USA

⁸Northwestern University Feinberg School of Medicine, Rehabilitation Institute of Chicago, USA

⁹Department of Physical Medicine and Rehabilitation, Harvard Medical School, USA

¹⁰Spaulding Rehabilitation Hospital, USA

¹¹MassGeneral Hospital for Children™ Sports Concussion Program, USA

¹²Home Base, A Red Sox Foundation and Massachusetts General Hospital Program, USA

*Corresponding author at: 5 Rose Hill Drive, Bear, DE 19701, USA. Tel.: +1-302-312-4077; fax: +1-302-836-4756.
E-mail address: jhold@udel.edu (J.A. Holdnack).

Editorial Decision 21 March 2017; Accepted 29 March 2017

Abstract

Introduction: The National Institutes of Health Toolbox for Assessment of Neurological and Behavioral Function Cognition Battery is comprised of seven cognitive tests, including two tests measuring crystallized cognitive ability (i.e., vocabulary and reading) and five tests measuring fluid cognitive functioning (i.e., working memory, memory, speed of processing, and executive functioning). This study presents comprehensive base rate tables for the frequency of low scores in adults and older adults from the normative sample.

Methods: Participants were 843 adults, ages 20–85, from the NIH Toolbox standardization sample who completed all seven cognition tests. Rates of low scores were derived for standard age-adjusted and fully-demographically-adjusted scores at multiple cut-scores. Base rates were stratified by education, crystallized intellectual ability, and cognitive domain.

Results: Using the five demographically-adjusted fluid cognitive test scores, 45.9% of adults obtained one or more scores at or below the 16th percentile, and 16.8% obtained one or more score at or below the 5th percentile, which is consistent with findings from other neurocognitive test batteries.

Discussion: Based on the study findings, nearly 50% of adults in the general population would meet psychometric criteria for a diagnosis of the Diagnostic and Statistical Manual of Mental Disorders-Fifth Edition (DSM-5) Mild Neurocognitive Disorder (MND). We developed new psychometric criteria for identifying MND using the NIH Toolbox Cognition Battery that reduce the false positive rate. Knowing these multivariate normative base rates will help researchers and clinicians interpret NIH Toolbox scores in people with neurodevelopmental, psychiatric, medical, neurological, and neurodegenerative disorders that affect cognitive functioning.

Keywords: Multivariate base rates; Cognition; NIH Toolbox; Traumatic brain injury; Cognitive impairment

Introduction

The National Institutes of Health Toolbox for the Assessment of Neurological and Behavioral Function Cognition Battery (NIHTB-CB; Gershon et al., 2010, 2013; Weintraub et al., 2013) was designed to be used in research and clinical trials with children, adolescents, adults, and older adults. It can be used with those who have neurodevelopmental disorders,

psychiatric disorders, general medical conditions, neurological conditions, and neurodegenerative diseases. Reliability and validity studies illustrate that the NIHTB-CB correlates well with established measures of verbal, working memory, memory, processing speed, and executive functioning in healthy adults (Heaton et al., 2014; Mungas et al., 2014) and clinical validation studies are in progress (Holdnack, Iverson, Silverberg, Tulsy, & Heinemann, under review; Tulsy et al., under review). Given its potential wide applicability, it is essential for users of this cognition battery to understand how to interpret patterns of performance. For example, if the battery is used to measure cognitive deficits associated with medical, psychiatric, or neurological conditions, it is important to know how common it is for a normative sample to obtain certain combinations of low scores on these tests.

It is common for individuals in the general population to obtain one or more low scores when administered a battery of cognitive tests (Axelrod & Wall, 2007; Brooks, Iverson, Holdnack, & Feldman, 2008; Brooks, Strauss, Sherman, Iverson, & Slick, 2009; Crawford, Garthwaite, & Gault, 2007; Ingraham & Aiken, 1996; Palmer, Boone, Lesser, & Wohl, 1998; Schretlen, Testa, Winicki, Pearlson, & Gordon, 2008). The more tests that are administered, the greater the probability that an individual will have one or more low scores (Brooks, Iverson, & Holdnack, 2013; Crawford et al., 2007). For example, 62.8% of subjects from the Wechsler Adult Intelligence Scale-Fourth Edition (WAIS-IV) standardization sample will have one or more scores at or below a scaled score of 7 (e.g., univariate base rate of 16%) when all 10 subtests are considered simultaneously. Considering the two visual working memory subtests from the Wechsler Memory Scale-Fourth Edition (WMS-IV), 22.1% of individuals in the standardization sample obtain one or more scaled scores of 7 or less (Brooks et al., 2013). Univariate and multivariate base rates can differ with as few as two tests considered simultaneously, although there is a strong relationship between the number of tests administered and the increased rate of observed low scores.

This multivariate base rate phenomenon is not specific to a particular model for normative data collection, norm development procedure, or a specific battery of tests, but occurs on a broad range of measures (Brooks, Holdnack, & Iverson, 2011; Brooks, Iverson, & White, 2007; Palmer et al., 1998; Schretlen et al., 2008) and ages (Brooks et al., 2011; Brooks, Iverson, Sherman, & Holdnack, 2009). Moreover, the probability of obtaining low scores varies based on the person's level of intelligence. Those with above average intelligence have a lower probability and those with below average intelligence having a much higher probability of obtaining low scores (Binder, Iverson, & Brooks, 2009; Brooks et al., 2008, 2011).

To use the NIHTB-CB more effectively in research (and potentially in clinical practice), it is essential to know the prevalence of low scores in the normative sample. The NIHTB-CB is comprised of seven performance-based neuropsychological tests measuring attention, working memory, language, processing speed, and executive functioning in two broad domains of functioning (crystallized and fluid abilities). The crystallized tests are Picture Vocabulary and Oral Reading Recognition. They yield two scores and are combined into a Crystallized Cognition Composite Score. The fluid cognition tests are Picture Sequence Memory, List Sorting Working Memory, Pattern Comparison Processing Speed, Flanker Inhibitory Control and Attention, and Dimensional Change Card Sort (Weintraub et al., 2014). These two domains differ in their sensitivity to cognitive impairment in that verbal abilities are less likely to be affected by brain injury compared to fluid measures (Carlozzi, Grech, & Tulsy, 2013; Carlozzi, Kirsch, Kisala, & Tulsy, 2015; Donders & Strong, 2015; Madigan, DeLuca, Diamond, Tramontano, & Averill, 2000; Sinclair, Ponsford, Rajaratnam, & Anderson, 2013). The current study provides general population base rates of low scores for the NIHTB-CB for adults. We report base rates for the five fluid cognition scores using cut-off scores at the 25th, 16th, 9th, and 5th percentiles.

Additionally, we propose new psychometric criteria for defining cognitive impairment on the NIHTB-CB. Previously, we recommended a psychometric model for identifying memory deficits that may be associated with mild cognitive impairment (MCI) or dementia using the Wechsler Adult Intelligence Scale-Third Edition and Wechsler Memory Scale-Third Edition (Brooks, Iverson, Feldman, & Holdnack, 2009). This model focused primarily on memory deficits, which are considered a core feature of MCI (Petersen et al., 1999) and an early sign of Alzheimer's disease (Grober & Kawas, 1997). In this study, 97.1% of individuals identified with probable Alzheimer's disease met criteria for possible memory impairment and 94.1% met criteria for probable memory impairment; the specificity for these criteria was 0.80 and 0.90, respectively (Brooks, Iverson, Feldman, & Holdnack, 2009). A more general cognitive impairment model was developed for the combined WAIS-IV/WMS-IV core battery and applied in a sample of individuals with moderate to severe traumatic brain injury (Brooks et al., 2011). In the TBI sample, 81.5% met criteria for below expected cognitive performance and 63% met criteria for well below expected cognitive performance with specificity values of 0.75 and 0.90 for these classifications (Brooks et al., 2011). Based on this prior work, we propose a new psychometric model for identifying cognitive impairment that may support a diagnosis mild neurocognitive disorder for the NIHTB-CB.

Methods

Participants

The NIHTB-CB normative sample (Beaumont et al., 2013) was used to estimate the population prevalence of low scores. We were interested in the performance of adults 20–85 years of age ($N = 1,021$). We selected examinees who completed all seven of the NIHTB-CB tests, resulting in 843 individuals. For demographically-adjusted norms, 793 examinees had complete performance data and complete demographic data required for derivation of demographic norms. Demographically-adjusted norms adjust obtained raw scores by race/ethnicity (e.g., White, Black, and Hispanic), age, sex, and years of education (Casaletto et al., 2015).

Identification and testing of examinees was completed through a marketing research firm from locations throughout the US. For the purposes of data collection, the normative sample was stratified by age and sex and while education level, race, and ethnicity were not stratification variables. Population based targets were set for these demographic factors within each age group (Beaumont et al., 2013). Exclusion criteria included: capacity to following instructions in English and adequate visual, auditory, vestibular, and motor functioning to complete all items in the full Toolbox test battery, or availability of assistance or assistive devices to complete tasks (Beaumont et al., 2013). Participants were not screened for medical or psychiatric conditions and examinees were not evaluated or excluded based on performance validity measures. Examiners were trained and monitored by NIH Toolbox investigators. Examinees were compensated for participation. Table 1 presents demographic data for the age-only and full demographically-adjusted (e.g., age, education, sex, and race/ethnicity) normative samples.

Measures

The NIHTB-CB contains measures of episodic memory, language, processing speed, working memory, executive function, and attention. Three composite measures are derived from combinations of individual subtests. The Total Composite is calculated using all seven subtests of the NIHTB-CB. The crystallized composite score is derived from two subsets of the battery: Picture Vocabulary and Oral Reading Recognition. For individuals with acquired brain injury, the Crystallized Cognition Composite Score can serve as an estimate of premorbid or pre-injury general cognitive ability. This composite correlates strongly with other known “hold” tests (Gershon et al., 2014) and is resistant to the effects of traumatic brain injury (TBI) (Tulsky et al., under review). The fluid composite score is derived from five subtests of the battery: Picture Sequence Memory, List Sorting Working Memory, Pattern Comparison Processing Speed, Flanker Inhibitory Control and Attention, and Dimensional Change Card Sort. The Fluid Composite scores were derived by averaging the normalized scores of each of the fluid tests, and then deriving scale scores based on this new distribution of averaged normalized scores. The age-adjusted

Table 1. Demographic characteristics of the normative sample

Variable	Age norms	Demographic norms
Sample size	843	793
Age (years)		
M (SD)	47.4 (17.4)	47.7 (17.1)
Gender (%)		
Male	34.4	34.3
Female	65.6	65.7
Race (%)		
Caucasian	64.5	68.6
African American	16.4	17.3
Hispanic	10.7	10.5
Other	6.6	3.6
Not provided	1.8	0.0
Education (%)		
Less than 12 years	9.6	9.7
12 years	26.3	26.5
13–15 years	23.8	23.1
16 or more years	39.4	40.7
Unknown	0.8	0.0
Education (years)		
M (SD)	14.2 (2.5)	14.2 (2.5)

scale scores have a mean of 100 and a standard deviation (*SD*) of 15. The demographically-adjusted *T* scores are adjusted for age, gender, race (white, black, other), ethnicity (Hispanic vs. non-Hispanic), and education; they have a mean of 50 and a *SD* of 10 (Casaletto et al., 2015). A description of these tests is provided in Table 2. Total test administration requires approximately 30 minutes.

Results

Table 3 presents multivariate base rates for the five fluid tests of the NIHTB-CB for age- and demographically-adjusted scores. Rates of low age-adjusted scores (i.e., ≤ 16 th percentile) on the fluid tests were significantly correlated with the NIHTB-CB Crystallized composite ($r = -0.41, p < .001$), Picture Vocabulary Test ($r = -0.39, p < .001$), and Oral Reading Recognition Test ($r = -0.37, p < .001$). This relationship was also observed for demographically-adjusted scores: Crystallized composite ($r = -0.58, p < .001$), Picture Vocabulary Test ($r = -0.54, p < .001$), and Oral Reading Recognition Test ($r = -0.49, p < .001$). Therefore, Table 3 provides base rates by crystallized ability level, in addition to overall base rates of low scores.

The multivariate base rates of obtaining one or more low demographically-adjusted scores in the total normative sample and stratified by intellectual ability (i.e., the Crystallized Composite) are illustrated visually in Fig. 1. As seen in this figure, it is common to obtain one or more low scores across this 5-test battery, and the base rate of low scores declines in tandem with increasing levels of intelligence.

As seen in Table 3, when considering the entire normative sample, it is common for people to obtain one or more scores ≤ 16 th percentile [age norms base rate (BR) = 42.7%, demographic norms base rate = 45.9%], but it is uncommon to obtain three or more scores ≤ 16 th percentile (age norms BR = 9.5%, demographic norms BR = 9.8%). People with less than 12 years of education are more likely to obtain 2 or more scores ≤ 16 th percentile (BR = 25.9%) than people with 16+ years of education (BR = 17.2%). Obtaining 3 or more scores ≤ 16 th percentile is fairly common for those with below average intellectual abilities (Crystallized Composite $SS = 90/T = 43$; age norms BR = 22.4%, demographic norms BR = 22.9%), but this many low scores is rare for those with above average intellectual abilities (Crystallized Composite $SS = 110+/T = 58+$;

Table 2. Descriptions of the seven tests of the NIH Toolbox Cognition Battery.

Picture Vocabulary (crystallized)

A computer adaptive test where the examinee hears a word and sees four images on the screen, and then is asked to choose the picture that most closely matches the meaning of the word.

Oral Reading Recognition (crystallized)

A computer adaptive test pronouncing letters and words.

List Sorting Working Memory (fluid)

Pictures of foods (e.g., “hamburger”) and animals (e.g., “elephant”) are displayed with their names presented visually (in text) and auditorily (by a recording). The examinee is asked to repeat all the items in order of size, from smallest to largest. For List 1, there is only one dimension (foods or animals). For List 2, the examinee must first repeat the foods by size and then repeat the animals by size.

Picture Sequence Memory (fluid)

There are two learning trials for this test. A sequence of pictures of objects and activities are presented. Audio-recorded phrases representing the pictures are presented simultaneously. The examinee is asked to recall the sequence of pictures in order. The sequence length varies from 6 to 18 pictures, depending on age. Examinees get points for each adjacent pair of pictures correctly recalled, to a maximum of one minus the total number of pictures (e.g., $18 - 1 = 17$).

Pattern Comparison Processing Speed (fluid)

The examinee is presented with two side-by-side pictures on the computer screen and asked if they are the same or not. The pictures are designed to be simple. The maximum time to complete the test is 90 seconds. The highest possible score is 130.

Flanker Inhibitory Control and Attention (fluid)

The examinee must focus on a stimulus in the middle of the screen while inhibiting attention to stimuli flanking it. The flanking stimuli are either congruent (same direction) or incongruent (opposite direction) with the middle stimulus. The examinee must indicate with the arrow key the left-right direction of the centrally presented stimulus and ignore the stimuli on either side (the flankers). The stimuli are fish for children and arrows for adults. The test consists of 20 trials.

Dimensional Change Card Sort (fluid)

The examinee is presented with two target pictures that vary in two dimensions, such as shape and color (e.g., a white rabbit and a green boat). The test consists of four blocks: practice, pre-switch, post-switch, and mixed. During the practice block, the examinee is shown a picture in the middle of the screen and asked to match (i.e., sort) it to a target based on either being the same shape (or object, such as a rabbit) or the same color (such as green). For “mixed” trials, the examinee must change the dimension that is being matched (i.e., shape or color). For example, after several trials matching on color the examinee will be asked to match on shape, and then the next trial might go back to color, requiring shifting back and forth between these dimensions.

Note: These descriptions were adapted from the NIH Toolbox Scoring and Interpretation Guide (Slotkin et al., September 18, 2012, available for download at www.nihtoolbox.org).

Table 3. Base rates of low scores on five NIHTB-CB fluid measures overall and by ability level in the normative sample

Low scores	Age norms	Age norms by education level				Age norms by crystallized composite				Demographic norms	Demographic norms by crystallized composite			
		12	12	13–15	16+	<90	90–99	100–109	110 +		<43	43–49	50–57	58+
<i>N</i>	843	81	222	201	332	201	201	213	228	793	78	204	229	194
≤25th percentile														
5	3.0	3.7	4.1	2.5	2.4	8.5	3.0	0.9	0.0	2.4	6.6	2.9	0.9	0.0
4+	8.8	17.3	9.5	8.0	6.9	21.4	10.9	3.8	0.4	7.7	16.3	12.3	3.1	1.0
3+	18.0	30.9	21.6	16.4	13.9	40.3	22.4	8.5	3.5	15.0	33.1	19.1	8.7	2.6
2+	35.1	53.1	40.1	32.3	28.9	61.2	42.8	26.8	13.2	32.8	60.8	37.3	24.9	13.4
1+	62.8	82.7	67.1	62.7	55.4	85.6	68.7	53.5	46.1	62.2	82.5	66.2	59.8	43.3
None	37.2	17.3	32.9	37.3	44.6	14.4	31.3	46.5	53.9	37.8	17.5	33.8	40.2	56.7
≤16th percentile														
5	0.7	2.5	0.5	1.0	0.3	3.0	0.0	0.0	0.0	0.9	3.6	0.5	0.0	0.0
4+	3.4	7.4	3.6	3.0	2.7	11.9	1.5	0.5	0.4	4.2	10.8	4.9	1.3	1.0
3+	9.5	16.0	9.5	8.5	8.7	22.4	10.9	4.2	1.8	9.8	22.9	12.7	5.2	1.0
2+	19.8	25.9	23.4	17.9	17.2	37.3	26.9	13.1	4.4	20.7	41.0	23.5	14.0	8.2
1+	42.7	56.8	49.5	38.8	37.3	67.7	49.3	34.7	22.4	45.9	74.7	48.0	39.7	26.3
None	57.3	43.2	50.5	61.2	62.7	32.3	50.7	65.3	77.6	54.1	25.3	52.0	60.3	73.7
≤9th percentile														
5	0.2	2.5	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.3	0.6	0.5	0.0	0.0
4+	1.2	4.9	0.9	0.5	0.9	4.5	0.5	0.0	0.0	1.4	6.0	1.0	0.4	0.0
3+	4.2	11.1	4.1	4.0	2.7	12.9	3.5	0.5	0.4	2.4	11.4	3.9	1.3	1.0
2+	11.2	19.8	12.6	11.9	7.5	26.4	13.9	4.7	1.3	8.1	27.7	14.7	6.1	3.1
1+	28.2	38.3	32.4	27.4	23.8	48.8	32.8	22.1	11.8	31.9	57.2	33.8	25.8	15.5
None	71.8	61.7	67.6	72.6	76.2	51.2	67.2	77.9	88.2	68.1	42.8	66.2	74.2	84.5
≤5th percentile														
5	0.1	1.2	0.0	0.0	0.0	0.5	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
4+	0.4	2.5	0.5	0.0	0.0	1.5	0.0	0.0	0.0	0.4	1.8	0.0	0.0	0.0
3+	1.7	4.9	1.4	1.5	1.2	6.0	1.0	0.0	0.0	1.8	6.0	1.5	0.4	0.0
2+	5.3	12.3	5.0	4.5	4.2	15.4	6.0	0.5	0.4	4.7	14.5	4.4	0.9	1.0
1+	16.6	25.9	17.1	16.4	14.2	31.8	17.9	12.7	5.3	16.8	33.7	18.6	10.9	7.2
None	83.4	74.1	82.9	83.6	85.8	68.2	82.1	86.9	94.7	83.2	66.3	81.4	89.1	92.8

Note: The five tests are: Picture Sequence Memory, List Sorting Working Memory, Pattern Comparison Processing Speed, Flanker Inhibitory Control and Attention, and Dimensional Change Card Sort. Standard scores represent age-only adjustments. T-scores represent scores adjusted by age, sex, education level, and race/ethnicity.

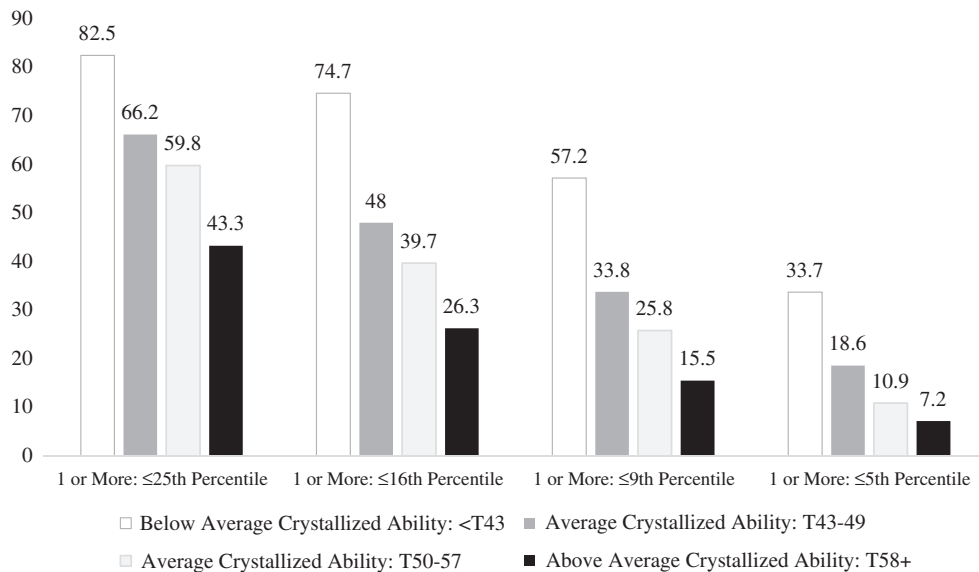


Fig. 1. Base rates of adults in the normative sample obtaining low demographically-adjusted fluid test scores stratified by crystallized composite score ability level (a total of 5 scores were considered).

age norms BR = 1.8%, demographic norms BR = 1.0%). Having a single score \leq 5th percentile occurred in 16.6% of the normative sample using age norms and 16.8% using demographic norms. Having two or more scores at or below the 5th percentile is uncommon for most examinees (see Table 3).

“One-Size-Fits-All” Diagnostic and Statistical Manual of Mental Disorders-Fifth Edition (DSM-5) Criteria for Mild Neurocognitive Disorder

The DSM-5 (American Psychiatric Association, 2013) recommends that mild neurocognitive disorder be defined, psychometrically, as cognitive performance in at least one cognitive domain that is at least one *SD* below average using age- and education-adjusted normative data. If that criterion is applied to age-adjusted or demographically-adjusted scores then 42.7% or 45.9% of the general population would be identified with MND (see Table 3). An alternative approach is to apply algorithms with a specific number of low scores based on a specific level of performance that takes into account a person’s level of intellectual ability as assessed by the Crystallized Composite score.

When applying a “one-size-fits-all” model as presented in Table 4, if a person with a traumatic brain injury obtained 4+ scores \leq 25th percentile (BR = 7.7%), or 3+ scores \leq 16th percentile (BR = 9.8%) or 2+ scores \leq 9th percentile (BR = 8.1%), the overall false positive rate for that *clinical algorithm* is 14.2% (less than 1 *SD* below the mean), using demographically-adjusted scores. That same algorithm appears to be perfect when using age-adjusted normative data, too (overall BR = 14.9%). However, as seen in the last 4 rows of Table 4, problems arise when considering a person’s level of intellectual functioning, as measured by the Crystallized Composite score. Using demographically-adjusted normative data (i.e., adjusted for age, education, sex, and race), the false positive rate for this one-size-fits-all algorithm ranges from 3.1% to 29.5%.

Table 4. Applying a standardized “One-Size-Fits-All” algorithm for defining DSM-5 mild neurocognitive disorder on the NIH Toolbox Cognition Battery

Comparison group	Age-adjusted scores (percentile rank cut-off scores)	Cut-off base rate	Overall base rate	Demographically-adjusted scores (percentile rank cut-off scores)	Cut-off base rate	Overall base rate
Adult general population	4+ scores \leq 25th, or 3+ scores \leq 16th, or 2+ scores \leq 9th	8.8% 9.5% 11.2%	14.9%	4+ scores \leq 25th, or 3+ scores \leq 16th, or 2+ scores \leq 9th	7.7% 9.8% 8.1%	14.2%
Education <12	4+ scores \leq 25th, or 3+ scores \leq 16th, or 2+ scores \leq 9th	17.3% 16.0% 19.8%	25.9%	NA	—	—
Education = 12	4+ scores \leq 25th, or 3+ scores \leq 16th, or 2+ scores \leq 9th	9.5% 9.5% 12.6%	16.2%	NA	—	—
Education = 13–15	4+ scores \leq 25th, or 3+ scores \leq 16th, or 2+ scores \leq 9th	8.0% 8.5% 11.9%	13.9%	NA	—	—
Education = 16+	4+ scores \leq 25th, or 3+ scores \leq 16th, or 2+ scores \leq 9th	6.9% 8.7% 7.5%	12.0%	NA	—	—
Crystallized Composite: <SS90/<T43 (Below Average)	4+ scores \leq 25th, or 3+ scores \leq 16th, or 2+ scores \leq 9th	21.4% 22.4% 26.4%	31.8%	4+ scores \leq 25th, or 3+ scores \leq 16th, or 2+ scores \leq 9th	16.3% 22.9% 27.7%	29.5%
Crystallized Composite: SS90-99/T43-49 (Average)	4+ scores \leq 25th, or 3+ scores \leq 16th, or 2+ scores \leq 9th	10.9% 10.9% 13.9%	19.4%	4+ scores \leq 25th, or 3+ scores \leq 16th, or 2+ scores \leq 9th	12.3% 12.7% 14.7%	20.1%
Crystallized Composite: SS100-109/T50-57 (Average)	4+ scores \leq 25th, or 3+ scores \leq 16th, or 2+ scores \leq 9th	3.8% 4.2% 4.7%	8.9%	4+ scores \leq 25th, or 3+ scores \leq 16th, or 2+ scores \leq 9th	3.1% 5.2% 6.1%	7.4%
Crystallized Composite: SS110+/T58+ (Above Average)	4+ scores \leq 25th, or 3+ scores \leq 16th, or 2+ scores \leq 9th	0.4% 1.8% 1.3%	1.8%	4+ scores \leq 25th, or 3+ scores \leq 16th, or 2+ scores \leq 9th	1.0% 1.0% 3.1%	3.1%

Note: The cut-off base rates represent the percentages of adults in the normative sample, from each comparison group, who show *that specific pattern* of test performance across the five fluid cognition tests. The overall base rates represent the percentages of adults in the normative sample, from each comparison group, who show *one or more* of the patterns of test performance across the five fluid cognition tests. For some groups, the standardized algorithms yield high rates of false positives that can be reduced using a more flexible approach (see Table 5).

New Criteria for Defining Cognitive Impairment

There are no established criteria for defining cognitive impairment on the NIHTB-CB. Therefore, the base rates presented in Table 3 were used as the foundation for creating flexible algorithms based on examinee characteristics for defining cognitive impairment. As seen in Fig. 1 and Table 3, obtaining a single low fluid cognition score is common, especially in those with average or below average crystallized ability, so it might be preferred to require two or more low scores as the psychometric criterion for cognitive impairment. If a clinician or researcher set the criterion for cognitive impairment as having two or more demographically-adjusted scores \leq 16th percentile, then the false positive rates for cognitive impairment, stratified by crystallized ability, would be as follows: (a) High Average Crystallized Composite ($T = 58+$), BR = 8.2%; (b) Average Crystallized Composite ($T = 50-57$), BR = 14.0%; (c) Average Crystallized Composite ($T = 43-49$), BR = 23.5%; and Low Average Crystallized Composite ($T < 43$), BR = 41.0%. Applying this “one-size-fits-all” criterion to the entire sample, there is a low false positive rate in those with high average crystallized ability (8.2%), an acceptable rate in those with Crystallized Composite scores between T scores of 50 and 57 (14.0%), but an unacceptably high false positive rate of cognitive impairment for those with Crystallized Composite scores less than $T = 50$ (23.5–41.0%). Therefore, more accurate criteria for cognitive impairment need to reflect estimated pre-injury or pre-disease level of crystallized cognitive functioning.

As seen in Tables 3 and 5, if a researcher or clinician wanted to require the presence of at least 2 low scores as the definition of having a problem with cognitive functioning, and wanted to standardize the cut-off for the pattern of impairment across the 5 fluid cognitive tests as being approximately one SD below the mean (e.g., per DSM-5 mild neurocognitive disorder psychometric criteria), then the criteria for impairment using demographically-adjusted T scores would need to vary in relation to the person’s level of crystallized intellectual ability. For example, criteria for cognitive impairment, stratified by pre-injury or pre-disease level of cognitive functioning, could be as follows: (a) High Average Crystallized Composite ($T = 58+$), 2 or more scores \leq 25th percentile, BR = 13.4%; (b) Average Crystallized Composite ($T = 50-57$), 2 or more scores \leq 16th percentile, BR = 14.0%; (c) Average Crystallized Composite ($T = 43-49$), 2 or more scores \leq 9th percentile, BR = 14.7%; and Low Average Crystallized Composite ($T < 43$), 2 or more scores \leq 5th percentile, BR = 14.5%. Notice that these ability-stratified criteria for cognitive impairment hold the false positive rate (i.e., the base rate) constant.

New criteria for defining cognitive impairment on the NIHTB-CB are shown in Table 5. These criteria were selected for both the age-adjusted normative data and the demographically-adjusted normative data. The clinician or researcher can compare to the adult general population or to people from specific subgroups, such as subgroups stratified by education (for the age norms) or subgroups stratified by intellectual ability (for the age or demographic norms). The cut-off base rates represent the percentages of adults in the normative sample, from each comparison group, who show *that specific pattern* of test performance across the five fluid cognition tests. The *overall base rates* represent the percentages of adults in the normative sample, from each comparison group, who show *one or more* of the patterns of test performance across the five fluid cognition tests (for those rows in which 2 or more score patterns are listed). As such, the overall base rate represents the false positive rate for cognitive impairment when one considers *all of the patterns of performance simultaneously*. Numerous algorithms are provided so that researchers can test combinations of these algorithms in clinical samples, with the goal of identifying those with the greatest accuracy for identifying acquired cognitive impairment.

Discussion

This study provides new and important data for researchers and clinicians on how to interpret low scores on the Cognition Battery of the NIH Toolbox. It is common for adults in the general population to obtain one or more low scores on the NIHTB-CB using age-adjusted and demographically-adjusted normative scores. The probability of obtaining a low score varies based on the *a priori* cut-off for defining a low score (e.g., the 25th, 16th, 9th, or 5th percentile), and the likelihood that a person will obtain a low score does not correspond to the percentile rank for that score (see Fig. 1). Obtaining a score corresponding to the 16th percentile on one of the fluid cognition tests means that only 16% of people score at that level or lower. However, obtaining one or more scores \leq 16th percentile on the Cognition Battery of the Toolbox occurs in 42.7% of people when considering the five fluid tests. The results presented here are similar to prior studies illustrating that a substantial percentage of people without injury or disease will obtain one or more low test scores when administered a battery of cognitive tests (Axelrod & Wall, 2007; Binder et al., 2009; Brooks et al., 2007, 2008; Crawford et al., 2007; Heaton, Grant, & Matthews, 1991; Heaton, Miller, Taylor, & Grant, 2004; Ingraham & Aiken, 1996; Iverson, Brooks, & Holdnack, 2008; Iverson, Brooks, White, & Stern, 2008; Palmer et al., 1998; Schretlen et al., 2008).

The prevalence of low scores is strongly related to a person’s estimated intellectual ability, the Crystallized Composite score from the Toolbox. Using demographically-adjusted normative scores does not mitigate the effect of intelligence, as seen in

Table 5. New flexible algorithms for identifying cognitive impairment on the NIH Toolbox Cognition Battery

Comparison group	Age-adjusted scores (percentile rank cut-off scores)	Cut-off base rate	Overall base rate	Demographically-adjusted scores (percentile rank cut-off scores)	Cut-off base rate	Overall base rate
Adult general population	4+ scores \leq 25th, or	8.8%	14.9%	4+ scores \leq 25th, or	7.7%	14.2%
	3+ scores \leq 16th, or	9.5%		3+ scores \leq 16th, or	9.8%	
	2+ scores \leq 9th	11.2%		2+ scores \leq 9th	8.1%	
				4+ scores \leq 25th, or	7.7%	
				3+ scores \leq 16th, or	9.8%	
				2+ scores \leq 9th	8.1%	
Education <12	5+ scores \leq 25th, or	3.7%	16.0%	NA	—	—
	4+ scores \leq 16th, or	7.4%				
	3+ scores \leq 9th, or	11.1%				
	2+ scores \leq 5th	12.3%				
	4+ scores \leq 16th, or	7.4%	16.0%	NA	—	—
	3+ scores \leq 9th, or	11.1%				
	2+ scores \leq 5th	12.3%				
Education = 12	5+ scores \leq 25th, or	3.7%	14.8%	NA	—	—
	4+ scores \leq 16th, or	7.4%				
	2+ scores \leq 5th	12.3%				
	4+ scores \leq 25th, or	9.5%	16.2%	NA	—	—
	3+ scores \leq 16th, or	9.5%				
2+ scores \leq 9th	12.6%					
Education = 13–15	4+ scores \leq 25th, or	9.5%	13.5%	NA	—	—
	3+ scores \leq 16th, or	9.5%				
	2+ scores \leq 9th	12.6%				
	4+ scores \leq 25th, or	9.5%	13.9%	NA	—	—
3+ scores \leq 16th, or	8.5%					
2+ scores \leq 9th	11.9%					
Education = 16+	4+ scores \leq 25th, or	6.9%	12.0%	NA	—	—
	3+ scores \leq 16th, or	8.7%				
	2+ scores \leq 9th	7.5%				
Crystallized Composite: < SS90/ < T43 (Below Average)	4+ scores \leq 16th or	11.9%	15.9%	4+ scores \leq 16th or	10.8%	14.5%
	3+ scores \leq 9th	12.9%		3+ scores \leq 9th	11.4%	
	2+ scores \leq 5th	15.4%		2+ scores \leq 5th	14.5%	
Crystallized Composite: SS90-99/T43-49 (Average)	4+ scores \leq 25th, or	10.9%	15.9%	2+ scores \leq 9th	14.7%	14.7%
	3+ scores \leq 16th, or	10.9%				
	2+ scores \leq 5th	6.0%				
Crystallized Composite: SS100-109/T50-57 (Average)	2+ scores \leq 9th	13.9%	13.9%	3+ scores \leq 16th or	12.7%	14.2%
	2+ scores \leq 5th	13.9%		2+ scores \leq 5th	4.4%	
	3+ scores \leq 25th or	8.5%		2+ scores \leq 16th	14.0%	
Crystallized Composite: SS110+/T58+ (Above Average)	2+ scores \leq 16th	13.1%	13.1%	3+ scores \leq 25th or	8.7%	15.3%
	2+ scores \leq 16th	13.1%		2+ scores \leq 16th	14.0%	
	2+ scores \leq 25th or	13.2%		2+ scores \leq 25th or	13.4%	
Crystallized Composite: SS110+/T58+ (Above Average)	1+ score \leq 5th	5.3%	13.2%	1+ score \leq 5th	7.2%	13.4%
	2+ scores \leq 25th	13.2%		2+ scores \leq 25th	13.4%	
	1+ scores \leq 9th	11.8%		1+ scores \leq 9th	15.5%	

Note: The cut-off base rates represent the percentages of adults in the normative sample, from each comparison group, who show *that specific pattern* of test performance across the five fluid cognition tests. The overall base rates represent the percentages of adults in the normative sample, from each comparison group, who show *one or more* of the patterns of test performance across the five fluid cognition tests. The overall base rates using this approach range from 11.2% to 16.0%. For all groups, with the exception of those with Crystallized Composite scores in the high average range, the criteria for possible cognitive impairment require at least two low scores (out of 5).

Table 3 and Fig. 1. This finding illustrates the importance of considering longstanding intellectual ability when interpreting fluid cognitive test performance, such as performance on tests of memory, speed of processing, and executive functioning. Prior studies have also shown a strong association between a person's intelligence and the probability of obtaining low neuropsychological test scores (Binder et al., 2009, 2008, 2011; Horton, 1999; Steinberg, Bieliauskas, Smith, & Ivnik, 2005; Steinberg, Bieliauskas, Smith, Ivnik, & Malec, 2005; Tremont, Hoffman, Scott, & Adams, 1998; Warner, Ernst, Townes, Peel, & Preston, 1987).

Knowing multivariate base rates of low scores on this battery allows researchers to develop more refined and accurate criteria for cognitive impairment. The DSM-5 criteria for mild neurocognitive disorder indicates that cognitive test performance, in at least one cognitive domain, typically lies 1–2 standard deviations below the mean using age- and education-adjusted

normative data (between the 16th and 3rd percentiles). If we apply the DSM-5 criterion to the five fluid cognition measures from the NIHTB-CB (see Table 3), then 45.9% of the normative sample will meet this psychometric criterion for mild neurocognitive disorder using demographically-adjusted normative data, an unacceptable false positive rate. This finding is similar to previous studies using different test batteries (Brooks et al., 2007, 2011; Palmer et al., 1998; Schretlen et al., 2008) so is not specific to the psychometric qualities of the NIHTB-CB. Refined psychometric criteria for DSM-5 mild neurocognitive disorder that take into account multivariate base rates and intellectual ability are presented in Table 4. However, these criteria illustrate that a Procrustean “one-size-fits-all” approach for defining cognitive impairment in clinical practice and research will result in a greater than expected rate of false negative classifications in individuals with high premorbid functioning and false positive classifications in low-functioning people. By adjusting the criteria for cognitive impairment based on estimated intellectual ability, as seen in Table 5, the false positive rates can be held relatively constant. Several options for defining cognitive impairment, with comparable false positive rates, are provided in Table 5.

These multivariate base rates and criteria for cognitive impairment are designed to be selected *a priori* and used as a single algorithm. That is, if a researcher or clinician did not select the algorithm *a priori*, but instead examined a person’s specific pattern of results, considered multiple algorithms, and then selected one specific algorithm, the overall base rate would no longer be accurate because multiple combinations of scores were considered before selecting the final algorithm. An example of how base rates can be used incorrectly is as follows. Assume a patient with a history of moderate TBI obtained a Crystallized Composite score of 44. We estimate that his *pre-injury* Crystallized Composite was under 50, so we chose to use impairment criteria for people with Crystallized Composites between 43 and 49. As seen in Tables 3 and 4, for people with Crystallized Composites between 43 and 49, it is uncommon to obtain (a) 4+ scores \leq 25th percentile (BR = 12.3%), (b) 3+ scores \leq 16th percentile (BR = 12.7%), and (c) 2+ scores \leq 9th percentile (BR = 14.7%). However, if you consider all three of those criteria for impairment simultaneously, 1 in 5 adults from the normative sample with crystallized ability in that range will meet at least one criterion for impairment (BR = 20.1%). That is, if we did not make an *a priori* choice, and we considered all three of those patterns simultaneously and then chose one post hoc, the base rate would actually be 20.1% in the general population, not 12.3%–14.7%. Therefore, the more combinations of patterns of scores considered, after knowing the results of testing, the greater the probability of a false positive finding of acquired cognitive impairment.

The current sample included only examinees that completed the entire seven subtests of the NIHTB-CB. Excluded participants were not evaluated for any cognitive factors or technology related issues (e.g., technical problems or lack of computer experience) that may have contributed to a failure to complete all tasks. The goal of this study was to evaluate the rates of low scores in individuals that completed the entire battery and was not intended to evaluate characteristics of individuals that did not complete the full battery. Future studies are needed to determine if there are examinee characteristics that affect performance that may introduce non-construct related variance into test performance.

The current study does not assess the relative sensitivity of the NIHTB-CB tests nor the proposed algorithms to cognitive deficits in specific clinical conditions. Additionally, the algorithms provided are designed primarily for cognitive impairments affecting memory, executive functioning, and processing speed and presume verbal crystallized abilities are mostly intact such as in mild or moderate traumatic brain injury. Further, this study does not provide information regarding the sufficiency of domain coverage (e.g., comprehensiveness of memory tests to identify amnesic MCI) of the NIHTB-CB to identify all cognitive disorders that may be associated with the DSM-5 MND diagnosis.

In conclusion, if a clinician or researcher wanted to establish *a priori* criteria, and require a person to have two or more low scores in order to meet psychometric criteria for cognitive impairment, and control for a person’s level of intellectual ability, the following criteria could be used: (a) estimated premorbid high average Crystallized Composite ($T = 58+$), 2 or more scores \leq 25th percentile, BR = 13.4%; (b) estimated premorbid average Crystallized Composite ($T = 50–57$), 2 or more scores \leq 16th percentile, BR = 14.0%; (c) estimated premorbid average Crystallized Composite ($T = 43–49$), 2 or more scores \leq 9th percentile, BR = 14.7%; and (d) estimated premorbid low average Crystallized Composite ($T < 43$), 2 or more scores \leq 5th percentile, BR = 14.5%. Research is needed to inform how we estimate premorbid Crystallized Composite scores and to determine if these criteria are more accurate than the criteria for impairment described in Table 5.

Funding

Support for JAH was provided by an Institutional Development Award (IDeA) from the National Institute of General Medical Sciences of the National Institutes of Health under grant number U54-GM104941 (PI: Binder-Macleod). BLB acknowledges support from the Canadian Institutes of Health Research. AWH received funding for the Rehabilitation Research and Training Center on Measuring Rehabilitation Outcomes and Effectiveness (H133B040032) from the National Institute on Disability, Independent Living, and Rehabilitation Research. GLI acknowledges support from the Mooney-Reed Charitable

Foundation. GLI notes that this work is related in part to the TBI Endpoints Development Initiative and a grant titled Development and Validation of a Cognition Endpoint for Traumatic Brain Injury Clinical Trials.

Conflict of Interest

None declared.

Acknowledgments

Standardization data from NIH Toolbox used with permission.

Financial Disclosure

The authors have no commercial or proprietary financial interest in the NIH Toolbox.

General Disclosure

GLI has been reimbursed by the government, professional scientific bodies, and commercial organizations for discussing or presenting research at meetings, scientific conferences, and symposiums. He has a clinical practice in forensic neuropsychology involving individuals who have sustained mild TBIs. He has received honorariums for serving on research panels that provide scientific peer review of programs. He is a co-investigator, collaborator, or consultant on grants relating to mild TBI funded by several organizations. He has received research support from test publishing companies in the past (not in the past 5 years). He receives royalties for one neuropsychological test (Wisconsin Card Sorting Test-64 Card Version).

BLB receives royalties for the sales of the Pediatric Forensic Neuropsychology textbook (2012, Oxford University Press) and pediatric neuropsychological tests [Child and Adolescent Memory Profile (ChAMP, Sherman and Brooks, 2015, PAR Inc.), Memory Validity Profile (MVP, Sherman and Brooks, 2015, PAR Inc.), and Multidimensional Everyday Memory Ratings for Youth (MEMRY, Sherman and Brooks, 2017, PAR Inc.)]. He has previously received in-kind support (free test credits) from a computerized cognitive test publisher (CNS Vital Signs, Chapel Hill, North Carolina).

References

- American Psychiatric Association (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Washington, DC: American Psychiatric Association.
- Axelrod, B. N., & Wall, J. R. (2007). Expectancy of impaired neuropsychological test scores in a non-clinical sample. *International Journal of Neuroscience*, *117*, 1591–1602.
- Beaumont, J. L., Havlik, R., Cook, K. F., Hays, R. D., Wallner-Allen, K., Korper, S. P., et al. (2013). Norming plans for the NIH Toolbox. *Neurology*, *80*, S87–S92.
- Binder, L. M., Iverson, G. L., & Brooks, B. L. (2009). To err is human: “Abnormal” neuropsychological scores and variability are common in healthy adults. *Archives of Clinical Neuropsychology*, *24*, 31–46.
- Brooks, B. L., Holdnack, J. A., & Iverson, G. L. (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*, 156–167.
- Brooks, B. L., Iverson, G. L., Feldman, H. H., & Holdnack, J. A. (2009). Minimizing misdiagnosis: Criteria for possible or probable memory impairment. *Dementia and Geriatric Cognitive Disorders*, *27*, 439–450.
- Brooks, B. L., Iverson, G. L., & Holdnack, J. A. (2013). Understanding and using multivariate base rates with the WAIS-IV/WMS-IV. In J. A. Holdnack, L. W. Drozdzick, L. G. Weiss, & G. L. Iverson (Eds.), *WAIS-IV/WMS-IV/ACS: Advanced clinical interpretation* (pp. 75–102). San Diego, CA: Elsevier Science.
- Brooks, B. L., Iverson, G. L., Holdnack, J. A., & Feldman, H. H. (2008). Potential for misclassification of mild cognitive impairment: a study of memory scores on the Wechsler Memory Scale-III in healthy older adults. *Journal of the International Neuropsychological Society*, *14*, 463–478.
- Brooks, B. L., Iverson, G. L., Sherman, E. M., & Holdnack, J. A. (2009). Healthy children and adolescents obtain some low scores across a battery of memory tests. *Journal of the International Neuropsychological Society*, *15*, 613–617.
- Brooks, B. L., Iverson, G. L., & White, T. (2007). Substantial risk of “Accidental MCI” in healthy older adults: Base rates of low memory scores in neuropsychological assessment. *Journal of the International Neuropsychological Society*, *13*, 490–500.
- Brooks, B. L., Strauss, E., Sherman, E. M. S., Iverson, G. L., & Slick, D. J. (2009). Developments in neuropsychological assessment: Refining psychometric and clinical interpretive methods. *Canadian Psychology*, *50*, 196–209.
- Carlozzi, N. E., Grech, J., & Tulskey, D. S. (2013). Memory functioning in individuals with traumatic brain injury: An examination of the Wechsler Memory Scale-Fourth Edition (WMS-IV). *Journal of Clinical and Experimental Neuropsychology*, *35*, 906–914.

- Carlozzi, N. E., Kirsch, N. L., Kisala, P. A., & Tulskey, D. S. (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). *The Clinical Neuropsychologist*, *29*, 21–37.
- Casaletto, K. B., Umlauf, A., Beaumont, J., Gershon, R., Slotkin, J., Akshoomoff, N., et al. (2015). Demographically corrected normative standards for the English version of the NIH Toolbox Cognition Battery. *Journal of the International Neuropsychological Society*, *21*, 378–391.
- Crawford, J. R., Garthwaite, P. H., & Gault, C. B. (2007). Estimating the percentage of the population with abnormally low scores (or abnormally large score differences) on standardized neuropsychological test batteries: A generic method with applications. *Neuropsychology*, *21*, 419–430. <http://homepages.abdn.ac.uk/j.crawford/pages/dept/PercentAbnormKtests.htm>. Test Software.
- Donders, J., & Strong, C. A. (2015). Clinical utility of the Wechsler Adult Intelligence Scale-Fourth Edition after traumatic brain injury. *Assessment*, *22*, 17–22.
- Gershon, R. C., Cella, D., Fox, N. A., Havlik, R. J., Hendrie, H. C., & Wagster, M. V. (2010). Assessment of neurological and behavioural function: the NIH Toolbox. *Lancet Neurology*, *9*, 138–139.
- Gershon, R. C., Cook, K. F., Mungas, D., Manly, J. J., Slotkin, J., Beaumont, J. L., et al. (2014). Language measures of the NIH Toolbox Cognition Battery. *Journal of the International Neuropsychological Society*, *20*, 642–651.
- Gershon, R. C., Wagster, M. V., Hendrie, H. C., Fox, N. A., Cook, K. F., & Nowinski, C. J. (2013). NIH Toolbox for assessment of neurological and behavioral function. *Neurology*, *80*, S2–S6.
- Grober, E., & Kawas, C. (1997). Learning and retention in preclinical and early Alzheimer's disease. *Psychology and Aging*, *12*, 183–188.
- Heaton, R. K., Akshoomoff, N., Tulskey, D., Mungas, D., Weintraub, S., Dikmen, S., et al. (2014). Reliability and validity of composite scores from the NIH Toolbox Cognition Battery in adults. *Journal of the International Neuropsychological Society*, *20*, 588–598.
- Heaton, R. K., Grant, I., & Matthews, C. G. (1991). *Comprehensive norms for an extended Halstead-Reitan Battery: Demographic corrections, research findings, and clinical applications*. Odessa, FL: Psychological Assessment Resources, Inc.
- Heaton, R. K., Miller, S. W., Taylor, M. J., & Grant, I. (2004). *Revised comprehensive norms for an expanded Halstead-Reitan Battery: Demographically adjusted neuropsychological norms for African American and Caucasian adults professional manual*. Lutz, FL: Psychological Assessment Resources.
- Holdnack, J. A., Iverson, G. L., Silverberg, N. D., Tulskey, D. S., & Heinemann, A. W. (under review). NIH Toolbox cognition tests following traumatic brain injury: Frequency of low scores. *Rehabilitation Psychology*.
- Horton, A. M. Jr (1999). Above-average intelligence and neuropsychological test score performance. *International Journal of Neuroscience*, *99*, 221–231.
- Ingraham, L. J., & Aiken, C. B. (1996). An empirical approach to determining criteria for abnormality in test batteries with multiple measures. *Neuropsychology*, *10*, 120–124.
- Iverson, G. L., Brooks, B. L., & Holdnack, J. A. (2008). Misdiagnosis of cognitive impairment in forensic neuropsychology. In R. L. Heilbrunner (Ed.), *Neuropsychology in the courtroom: Expert analysis of reports and testimony* (pp. 243–266). New York: Guilford Press.
- Iverson, G. L., Brooks, B. L., White, T., & Stern, R. A. (2008). Neuropsychological assessment battery (NAB): introduction and advanced interpretation. *The Neuropsychology Handbook*, 279–343.
- Madigan, N. K., DeLuca, J., Diamond, B. J., Tramontano, G., & Averill, A. (2000). Speed of information processing in traumatic brain injury: Modality-specific factors. *Journal of Head Trauma Rehabilitation*, *15*, 943–956.
- Mungas, D., Heaton, R., Tulskey, D., Zelazo, P. D., Slotkin, J., Blitz, D., et al. (2014). Factor structure, convergent validity, and discriminant validity of the NIH Toolbox Cognitive Health Battery (NIHTB-CHB) in adults. *Journal of the International Neuropsychological Society*, *20*, 579–587.
- Palmer, B. W., Boone, K. B., Lesser, I. M., & Wohl, M. A. (1998). Base rates of “impaired” neuropsychological test performance among healthy older adults. *Archives of Clinical Neuropsychology*, *13*, 503–511.
- Petersen, R. C., Smith, G. E., Waring, S. C., Ivnik, R. J., Tangalos, E. G., & Kokmen, E. (1999). Mild cognitive impairment: Clinical characterization and outcome. *Archives of Neurology*, *56*, 303–308.
- Schretlen, D. J., Testa, S. M., Winicki, J. M., Pearlson, G. D., & Gordon, B. (2008). Frequency and bases of abnormal performance by healthy adults on neuropsychological testing. *Journal of the International Neuropsychological Society*, *14*, 436–445.
- Sinclair, K. L., Ponsford, J. L., Rajaratnam, S. M., & Anderson, C. (2013). Sustained attention following traumatic brain injury: Use of the Psychomotor Vigilance Task. *Journal of Clinical and Experimental Neuropsychology*, *35*, 210–224.
- Steinberg, B. A., Bieliauskas, L. A., Smith, G. E., & Ivnik, R. J. (2005). Mayo's older Americans normative studies: Age- and IQ-Adjusted Norms for the Trail-Making Test, the Stroop Test, and MAE Controlled Oral Word Association Test. *The Clinical Neuropsychologist*, *19*, 329–377.
- Steinberg, B. A., Bieliauskas, L. A., Smith, G. E., Ivnik, R. J., & Malec, J. F. (2005). Mayo's older Americans normative studies: Age- and IQ-Adjusted Norms for the Auditory Verbal Learning Test and the Visual Spatial Learning Test. *The Clinical Neuropsychologist*, *19*, 464–523.
- Tremont, G., Hoffman, R. G., Scott, J. G., & Adams, R. L. (1998). Effect of intellectual level on neuropsychological test performance: A response to Dodrill (1997). *The Clinical Neuropsychologist*, *12*, 560–567.
- Tulskey, D. S., Carlozzi, N. E., Holdnack, J. A., Heaton, R. K., Wong, A., Goldsmith, A., et al. (under review). Using the NIH Toolbox Cognition Battery (NIHTB-CB) in individuals with traumatic brain injury. *Rehabilitation Psychology*.
- Warner, M. H., Ernst, J., Townes, B. D., Peel, J., & Preston, M. (1987). Relationships between IQ and neuropsychological measures in neuropsychiatric populations: Within-laboratory and cross-cultural replications using WAIS and WAIS-R. *Journal of Clinical and Experimental Neuropsychology*, *9*, 545–562.
- Weintraub, S., Dikmen, S. S., Heaton, R. K., Tulskey, D. S., Zelazo, P. D., Bauer, P. J., et al. (2013). Cognition assessment using the NIH Toolbox. *Neurology*, *80*, S54–S64.
- Weintraub, S., Dikmen, S. S., Heaton, R. K., Tulskey, D. S., Zelazo, P. D., Slotkin, J., et al. (2014). The cognition battery of the NIH toolbox for assessment of neurological and behavioral function: Validation in an adult sample. *Journal of the International Neuropsychological Society*, *20*, 567–578.