



OPEN ACCESS

EDITED BY

Sara Palermo,
University of Turin, Italy

REVIEWED BY

Eric E. Pierson,
Ball State University, United States
Daryaneh Badaly,
Child Mind Institute, United States

*CORRESPONDENCE

Nathan E. Cook
✉ necook@mgh.harvard.edu

RECEIVED 30 July 2024

ACCEPTED 18 March 2025

PUBLISHED 06 June 2025

CITATION

Cook NE, Iverson GL and Karr JE (2025)
Cognitive weaknesses or impairments on the
NIH toolbox cognition battery in children and
adolescents: base rates in a normative sample
and proposed methods for classification.
Front. Psychol. 16:1473095.
doi: 10.3389/fpsyg.2025.1473095

COPYRIGHT

© 2025 Cook, Iverson and Karr. This is an
open-access article distributed under the
terms of the [Creative Commons Attribution
License \(CC BY\)](#). The use, distribution or
reproduction in other forums is permitted,
provided the original author(s) and the
copyright owner(s) are credited and that the
original publication in this journal is cited, in
accordance with accepted academic
practice. No use, distribution or reproduction
is permitted which does not comply with
these terms.

Cognitive weaknesses or impairments on the NIH toolbox cognition battery in children and adolescents: base rates in a normative sample and proposed methods for classification

Nathan E. Cook^{1,2,3*}, Grant L. Iverson^{1,2,3,4} and Justin E. Karr⁵

¹Department of Physical Medicine and Rehabilitation, Harvard Medical School, Boston, MA, United States, ²Spaulding Rehabilitation Hospital, Charlestown, MA, United States, ³Mass General for Children Sports Concussion Program, Boston, MA, United States, ⁴Schoen Adams Research Institute at Spaulding Rehabilitation, Charlestown, MA, United States, ⁵Department of Psychology, University of Kentucky, Lexington, KY, United States

Introduction: The National Institutes of Health Toolbox for Assessment of Neurological and Behavioral Function Cognition Battery (NIHTB-CB) is a brief neuropsychological battery for the assessment of crystallized (i.e., vocabulary and word reading) and fluid cognition (i.e., working memory, visual episodic memory, processing speed, and executive functions). This study examined the frequency of low NIHTB-CB scores and proposes flexible algorithms for identifying cognitive weaknesses and impairment among youth.

Methods: Participants were 1,269 youth from the NIHTB-CB normative sample who did not have a neurodevelopmental, psychiatric, or medical problem that might be associated with cognitive difficulties (53% boys and 47% girls; $M = 11.8$ years-old, $SD = 3.0$, range 7–17). The sample included the following racial and ethnic composition: 58.1% White, 17.8% Black or African American, 16.8% Hispanic, 1.7% Asian, 3.1% multiracial and ethnic identities, and 2.6% not provided. The frequency of obtaining low scores falling at or below several cutoffs were calculated and stratified by gender, age, and crystallized intellectual ability.

Results: Considering the five fluid tests, nearly two-thirds of children and adolescents obtained one or more scores $\leq$ 25th percentile, half obtained one or more scores $\leq$ 16th percentile, between a third and a fourth obtained one or more scores $\leq$ 9th percentile, and nearly a fifth obtained one or more scores $\leq$ 5th percentile. We propose flexible, psychometrically derived criteria for identifying a cognitive weakness or impairment.

Discussion: Referencing the base rates of low scores will help researchers and clinicians enhance the interpretation of NIHTB-CB performance among children with cognitive weakness or impairments that are neurodevelopmental or acquired.

KEYWORDS

cognition, neuropsychological tests, psychometrics, cognitive dysfunction, cognitive assessment, base rates, fluid intelligence

1 Introduction

Families often receive neuropsychology services for their child during challenging and trying times. For example, consider a child who sustains a traumatic brain injury during a tragic accident, such as being struck by a car while crossing the street. Thankfully, given advances in modern medicine, the child survives, though neuroimaging reveals extensive structural injury to the brain including contusions and hemorrhages. A brief neuropsychological evaluation early in the rehabilitation process could be pursued to document the extent of cognitive problems and in a serial fashion to monitor outcome. To this end, the current study joins considerable prior work aiming to refine the scientific approach to interpreting performances across tests in a neuropsychological battery and determine whether results indicate cognitive impairment or neurodevelopmental weakness. We contend that examining individual test scores without considering multivariate base rates (as described further below) may increase the risk of interpretive and diagnostic errors, which might negatively impact the quality of care clinical neuropsychology can offer patients and their families.

When administered a cognitive test battery, children and adolescents in the general population frequently obtain one or more low scores (Brooks and Iverson, 2010, 2012; Brooks et al., 2013a; Cook et al., 2019). When considered in the context of a test battery, obtaining one or more low scores, even at or below the 5th percentile (which might reasonably be interpreted as occurring relatively infrequently in isolation), is surprisingly common among healthy individuals who have no known clinical conditions. The number of low test scores healthy children obtain varies based on the length of the battery (i.e., the more tests administered and interpreted, the greater the base rate of low scores), the cutoff score used to define a *low* or *impaired* score (i.e., higher cut-offs, such as 1 SD below the mean result in greater base rate of low scores compared to lower cutoffs, such as 1.5 SDs below the mean), and the examinee's overall level of intellectual functioning (e.g., higher levels of intelligence are associated with fewer obtained low scores) (Brooks, 2010; Brooks and Iverson, 2012; Brooks et al., 2009, 2010b; Cook et al., 2019).

The National Institutes of Health Toolbox for Assessment of Neurological and Behavioral Function Cognition Battery (NIHTB-CB) is a brief cognitive test battery (Weintraub et al., 2013). The battery is administered in an iPad-assisted format and includes seven tests assessing attention, working memory, visual episodic memory, language, processing speed, and executive functions. Tests are organized into composites scores of crystallized and fluid cognitive abilities as well as an overall total composite score. The developers evaluated and described the psychometric soundness of each test, including test-retest reliability (intra-class correlation coefficients ranging from 0.76 to 0.97) as well as convergent and discriminant validity (Bauer et al., 2013; Carlozzi et al., 2013; Gershon et al., 2013; Tulskey et al., 2013; Zelazo et al., 2013). Research on NIHTB-CB performances among children and adolescents has supported (i) a two-factor model of crystallized and fluid abilities (Akshoomoff et al., 2018), (ii) the convergent validity of fluid measures with traditional executive function tests (Kavanaugh et al., 2019), (iii) the usefulness of the NIHTB-CB as an outcome in clinical trials (Gladstone et al., 2019), and (iv) the feasibility of using the NIHTB-CB in pediatric rehabilitation settings (Watson et al., 2020).

The NIHTB-CB has been used extensively to examine the neuropsychological functioning of children and adolescents with diverse characteristics and clinical conditions. Examples of such conditions include preterm birth (Joseph et al., 2022); autism spectrum disorders (Jones et al., 2022; Solomon et al., 2021); fragile X syndrome, Down syndrome, and idiopathic intellectual disability (Dakopoulos et al., 2024; Shields et al., 2023); congenital heart defects (Schmithorst et al., 2022; Siciliano et al., 2021; Wallace et al., 2024); diabetes (Shapiro et al., 2021); human immunodeficiency virus (Molinario et al., 2021); epilepsy (Thompson et al., 2020); and mild traumatic brain injury (Chadwick et al., 2021). Moreover, researchers have also found that social/economic factors are associated with cognitive test performances, such that children from families of lower socioeconomic status obtain lower NIHTB-CB scores, on average (Hackman et al., 2021; Isaiah et al., 2023; Taylor et al., 2020). Thus, researchers and clinicians investigating and treating a broad array of clinical conditions as well as those investigating social and economic factors associated with pediatric health might consider using the NIHTB-CB. As the use of the NIHTB-CB continues to increase, investigators and clinicians alike would benefit from scientifically based methods for identifying cognitive weakness or impairment on the NIHTB-CB among children and adolescents that go beyond interpreting individual scores and that hold promise for reducing interpretive errors.

This study sought to examine the prevalence of low NIHTB-CB scores among youth from the battery's normative sample to provide psychometrically based guidance for identifying potential cognitive impairment or neurodevelopmental weakness. The frequency of low scores among children and adolescents has been reported for other test batteries, such as those measuring intelligence (Brooks, 2010), executive functions (Brooks et al., 2013b; Cook et al., 2019), and memory (Brooks et al., 2009). Further, the base rates of low scores have been reported among children and adolescents on multidomain batteries, both using traditional paper-and-pencil neuropsychological tests (Brooks et al., 2010b) as well as computerized tests (Brooks et al., 2010a). For the NIHTB-CB, base rates of low scores among adults have been reported (Holdnack et al., 2017a; Holdnack et al., 2017b; Iverson and Karr, 2021; Karr et al., 2022a,b), but base rates have not been reported among children and adolescents. The purpose of this study was to (a) provide base rates for low scores among healthy youth on the NIHTB-CB, and (b) provide flexible algorithms for identifying cognitive weakness or impairment.

2 Methods

2.1 Participants

The NIHTB-CB normative sample (Gershon, 2016) includes 2,413 youth ranging in age from 7 to 17 years, of which 1,722 completed all seven tests. Participants were excluded if they had a self- or proxy-reported diagnosis or history of preexisting health conditions that may affect cognitive functioning including: (a) attention-deficit/hyperactivity disorder ($n = 145$); (b) specific learning disability ($n = 51$); (c) autism, Asperger's syndrome, pervasive developmental disorder, or other autism spectrum disorder ($n = 17$); (d) developmental delay ($n = 13$); (e) mental health problems, including depression or anxiety ($n = 53$), bipolar disorder or schizophrenia

($n = 3$), a serious emotional disturbance ($n = 5$), or being hospitalized for emotional problems ($n = 12$); (f) alcohol or drug abuse ($n = 2$); (g) epilepsy/seizures ($n = 12$); (h) cerebral palsy ($n = 2$); (i) traumatic brain injury ($n = 2$); (j) head injury with loss of consciousness ($n = 19$) or head injury with amnesia ($n = 3$); (k) stroke or transient ischemic attack ($n = 1$); (l) brain surgery ($n = 4$); (m) diabetes ($n = 11$); (n) heart problems (e.g., heart attack, angina) ($n = 12$); (o) thyroid problems (e.g., Graves' disease) ($n = 9$); or (p) hypertension ($n = 6$). Participants were also excluded if they had a self-or proxy-report of (q) being on an individualized education program ($n = 76$), (r) attending special education classes ($n = 41$), or (s) receiving private tutoring or schooling for learning problems ($n = 27$). A total of 453 children had one or more of the above exclusion criteria.

The final sample ($N = 1,269$) included 53.0% girls and 47.0% boys with a mean age of 11.8 years ($SD = 3.0$) and a mean maternal education of 12.6 years ($SD = 2.5$). The sample included the following racial and ethnic composition: 58.1% White, 17.8% Black or African American, 16.8% Hispanic, 1.7% Asian, 3.1% multiracial and ethnic identities, and 2.6% not provided. A subsample ($N = 1,233$) had sufficient data to calculate demographically adjusted scores, including 53.1% girls and 46.9% boys with a mean age of 11.8 years ($SD = 3.0$) and a mean maternal education of 12.7 years ($SD = 2.4$). The sample included the following racial and ethnic composition: 59.8% White, 18.2% Black or African American, 17.2% Hispanic, 1.6% Asian, 3.2% multiracial and ethnic identities, and 1.7% not provided. This study includes secondary analyses of deidentified data and was approved by the Institutional Review Board of Mass General Brigham (Protocol #: 2020P000504).

2.2 Measures

There are seven NIHTB-CB tests (Akshoomoff et al., 2013). The crystallized composite is based on the Picture Vocabulary and Oral Reading Recognition tests (Gershon et al., 2013). The fluid composite is based on the other five tests: List Sorting Working Memory (Tulsky et al., 2013), Picture Sequence Memory (Bauer et al., 2013), Pattern Comparison Processing Speed (Carlozzi et al., 2013), Flanker Inhibitory Control and Attention (Zelazo et al., 2013), and the Dimensional Change Card Sort (Zelazo et al., 2013). Detailed descriptions of all tests are provided elsewhere (Holdnack et al., 2017b) and can also be accessed at: <https://www.nihtoolbox.org/domain/cognition/>. Age-adjusted scores are normed as Standard Scores (SS; $M = 100$, $SD = 15$) and demographically adjusted scores are normed as T-scores ($M = 50$, $SD = 10$), adjusting for age, gender, maternal education, and race/ethnicity based on published formulas (Casaletto et al., 2015).

2.3 Statistical analyses

The crystallized composite score, which consists of a receptive vocabulary test and a single word reading test, was used as a proxy for overall intellectual ability consistent with prior research on the NIHTB-CB (Holdnack et al., 2017b). This composite was either age-or demographic-adjusted, depending on this analysis, meaning the score reflected the expected language ability for a typically developing child of similar age or demographic characteristics. Tests of vocabulary and

verbal abilities have been examined as potential predictors of children's overall intellectual functioning or general cognitive ability (Schoenberg et al., 2007). Further, single word reading tests have long been used as estimates of overall intellectual abilities in adults, though less attention has been paid to children (Davis et al., 2018). In a sample of healthy children, the NIHTB-CB crystallized composite score correlated highly with composite scores of language abilities ($r = 0.90$) and fluid cognitive abilities (e.g., tests of attention, processing speed, working memory, and executive functions) ($r = 0.71$), showing a strong correspondence between this score and performances on several cognitive domains (Akshoomoff et al., 2013). Given prior evidence that the base rate of low cognitive test scores is strongly associated with level of intellectual ability (Brooks, 2010; Brooks and Iverson, 2012; Brooks et al., 2009; Brooks et al., 2010b; Cook et al., 2019), we examined the base rates of low test scores on the five fluid subtests stratified by crystallized composite standard scores. Specifically, the base rates of low scores among the five fluid scores was calculated across the following cutoffs: (a) $\leq 25th$ percentile was defined as $SS \leq 90$ and $T \leq 43$; (b) $\leq 16th$ percentile was defined as $SS \leq 85$ and $T \leq 40$; (c) $\leq 9th$ percentile was defined as $SS \leq 80$ and $T \leq 36$ (N.B., there is no whole number T score that corresponds to the 9th percentile, and a T score of 36 was selected as this cutoff, because it corresponds to the lowest whole number T score typically interpreted as unusually low in clinical practice); (d) $\leq 5th$ percentile was defined as $SS \leq 76$ and $T \leq 34$; and (e) $\leq 2nd$ percentile was defined as $SS \leq 70$ and $T \leq 30$. These percentile cutoffs for defining low scores are consistent with previous research (Brooks et al., 2013a; Cook et al., 2019; Karr et al., 2017, 2018). The frequencies of age-adjusted low scores were stratified by gender, age bands (i.e., 7–8, 9–10, 11–12, 13–14, 15–17), and crystallized composite standard scores (i.e., ≤ 89 , 90–99, 100–109, ≥ 110). The frequencies of demographically adjusted low scores were stratified by crystallized composite T scores (i.e., ≤ 43 , 44–49, 50–56, ≥ 57).

Descriptive interpretive labels were provided for the frequencies of low scores to aide the translation to clinical practice. The number of low scores obtained by the top performing 75% of the standardization sample reflected *Broadly Normal* performance; the number of low scores obtained by fewer than 25% of the standardization sample reflected *Below Average* performance; the number of low scores obtained by fewer than 10% of the standardization sample reflected *Uncommon* performance; and the number of low scores obtained by less than 3% of the standardization sample reflected *Very Uncommon* performance. For age-adjusted norms, these qualitative ranges are provided for the full sample and stratifications by gender, age, maternal education, and crystallized ability. For demographically adjusted norms, these descriptive labels are provided for the full sample and stratifications by crystallized ability.

Flexible algorithms were developed to identify patterns of performances on the NIHTB-CB involving two or more low scores at various cutoffs that may be interpreted as a cognitive weakness or cognitive impairment. They were influenced by previously published studies of the NIHTB-CB adult normative sample (Holdnack et al., 2017b; Iverson and Karr, 2021; Karr et al., 2022b). Algorithms based on patterns of performances that occur in 16% or fewer (i.e., 1 SD below the mean) and 7% or fewer (i.e., 1.5 SD s below the mean) were created. A pattern that occurred in roughly 16% of the normative sample was selected based on DSM-5 psychometric diagnostic criteria for mild neurocognitive disorder, which requires a standardized cutoff

of approximately 1 *SD* below the mean on neuropsychological testing for defining cognitive impairment (American Psychiatric Association, 2013). Because the frequencies of low scores varies by crystallized ability, separate algorithms for defining cognitive impairment were developed for different stratifications of crystallized composite scores.

3 Results

When considering the five fluid tests, a substantial percentage of healthy children and adolescents from the NIHTB-CB normative sample obtained low scores based on various cutoffs (see Table 1). Roughly two-thirds of children and adolescents obtained one or more scores $\leq$ 25th percentile (age-adjusted norms: 62.4%; demographically adjusted norms: 65.2%), about half obtained one or more scores $\leq$ 16th percentile (age-adjusted norms: 45.1%; demographically adjusted norms: 47.1%), between a third and a fourth obtained one or more scores $\leq$ 9th percentile (age-adjusted norms: 29.1%; demographically adjusted norms: 24.9%), and nearly a fifth obtained one or more scores $\leq$ 5th percentile (age-adjusted norms: 15.9%; demographically adjusted norms: 16.3%).

3.1 Low scores stratified by crystallized ability

Youth with low crystallized composite scores are more likely to obtain low fluid test scores (see Table 1 and Figure 1). For example, considering age-adjusted norms, one or more scores $\leq$ 25th percentile was obtained by 76.1% of youth with a below average crystallized composite (i.e., $SS \leq 89$), compared to 51.9% of youth with an above average crystallized composite (i.e., $SS \geq 110$). Similarly, considering demographically adjusted norms, one or more scores $\leq$ 25th percentile were obtained by 75.6% of youth with a below average crystallized composite (i.e., $T \leq 43$) compared to 56.6% of youth with an above average crystallized composite (i.e., $T \geq 57$). The clinical implications of this association are more notable at more stringent cutoffs. For example, only about 1 in 10 youth with above average crystallized composites obtained any scores $\leq$ 5th percentile (age-adjusted norms: 9.0%; demographically adjusted norms: 10.2%), whereas nearly 1 in 3 youth with below average crystallized composites obtained at least one score this low (age norms: 29.2%; demographically adjusted norms: 24.8%).

3.2 Classification ranges for low scores

Classification ranges for the expected number of low scores are provided (see Table 2), which can serve as an interpretive guide for the overall pattern of performances. For instance, using age-adjusted norms, obtaining two or more scores $\leq$ 16th percentile is considered *Broadly Normal* for youth with below average crystallized composite scores (i.e., $SS \leq 89$), *Below Average* for youth with average crystallized composite scores (i.e., $SS = 90-109$), and *Uncommon* for youth with above average crystallized composite scores (i.e., $SS \geq 110$). Similarly, it would be considered *Broadly Normal* for a healthy adolescent with a demographically adjusted crystallized composite in the average range (i.e., $T = 44-56$) to obtain two low scores $\leq$ 25th percentile, but this

would be considered *Below Average* for an adolescent with a crystallized composite score in the above average range (i.e., $T \geq 57$).

3.3 Algorithms for defining cognitive weakness or impairment

Flexible algorithms for identifying a developmental cognitive weakness or acquired cognitive impairment on the fluid tests were developed (Table 3). Some algorithms consist of a specific pattern of low scores, whereas other algorithms refer to whether a child or adolescent displays one of multiple possible patterns of low scores. For example, a potential algorithm to identify a developmental weakness or acquired cognitive impairment for youth with average crystallized ability ($SS = 100-109/T = 50-56$) is having two or more scores $\leq$ 16th percentile. The base rate for this pattern of low scores is 14.2% for age-adjusted norms and 12.6% for demographically adjusted norms. Another potential algorithm to identify a cognitive weakness or impairment for youth with the same crystallized ability is whether an examinee obtains three or more scores $\leq$ 25th percentile or two or more scores $\leq$ 9th percentile. The base rate for having either of these patterns of low scores is 11.5% for age-adjusted norms and 10.6% for demographically adjusted norms.

4 Discussion

It was common for children and adolescents in the standardization sample to obtain at least one low score across the five NIHTB-CB fluid subtests. Across the NIHTB-CB fluid subtests, nearly two-thirds of children and adolescents obtained one or more scores $\leq$ 25th percentile, half obtained one or more scores $\leq$ 16th percentile, between a third and a fourth obtained one or more scores $\leq$ 9th percentile, and nearly a fifth obtained one or more low scores $\leq$ 5th percentile. This finding is consistent with numerous prior studies on pediatric normative and standardization samples for various test batteries (Brooks, 2010; Brooks and Iverson, 2012; Brooks et al., 2009, 2010a, 2010b, 2013b; Cook et al., 2019). It is well established that low scores are more frequent when more tests are administered and interpreted. In the full sample, the base rate of obtaining one or more low scores $\leq$ 5th percentile on the five fluid NIHTB-CB tests is about 16% (age-adjusted norms: 15.9%; demographically adjusted norms: 16.3%). In comparison, the frequency of obtaining one or more low scores $\leq$ 5th percentile considering six subtests from the Children's Memory Scale is 22.4% (Brooks et al., 2009) and considering the four index scores from the Wechsler Intelligence Scale, Fourth edition (WISC-IV) is 14.6% (Brooks, 2010). Also consistent with prior studies, the number of low scores obtained is related to the cutoff used to define a low score. For example, when low scores were defined as falling $\leq$ 25th percentile, 62.4% of youth obtained at least one low age-adjusted NIHTB-CB fluid score, but only 7.1% obtained one or more low scores $\leq$ 2nd percentile.

It is important to note that we computed the base rates of low scores and the algorithms using only youth who had no reported neurodevelopmental, psychiatric, or medical conditions, and thus they represent the healthiest sample of youth included in the overall NIHTB normative sample. Given that we analyzed data from the normative sample, the percentage of youth obtaining a low score on

TABLE 1 Base rates of low scores on five NIHTB-CB fluid subtests stratified by gender, age, and crystalized ability level.

	Age-adjusted norms												Demographically-adjusted norms				
	Total Sample	Boys	Girls	Age					Crystalized composite				Total Sample	Crystalized composite			
				7–8	9–10	11–12	13–14	15–17	≤89	90–99	100–109	≥110		≤43	44–49	50–56	≥57
Sample size	1,269	597	672	220	260	258	235	296	301	305	339	324	1,233	307	300	301	325
≤25th percentile																	
5 low scores	0.9	1.5	0.4	0.5	1.2	1.9	0.9	0.3	2.0	1.3	0.6	–	0.8	1.3	1.3	0.7	–
4 or more	4.4	5.7	3.3	2.3	5.8	6.6	3.4	3.7	8.3	5.6	2.9	1.2	4.3	6.5	5.7	3.7	1.5
3 or more	13.3	13.7	12.9	7.3	20.0	14.7	11.1	12.5	20.6	15.7	10.6	7.1	14.9	22.5	18.0	10.3	9.2
2 or more	33.0	36.7	29.8	32.7	40.0	32.6	27.2	32.1	48.2	35.4	27.1	22.8	33.5	45.6	38.7	27.2	23.1
1 or more	62.4	65.2	60.0	67.7	67.3	64.7	54.9	58.1	76.1	64.3	58.7	51.9	65.2	75.6	69.3	59.8	56.6
No low scores	37.6	34.8	40.0	32.3	32.7	35.3	45.1	41.9	23.9	35.7	41.3	48.1	34.8	24.4	30.7	40.2	43.4
≤16th percentile																	
5 low scores	0.2	0.3	0.1	–	0.4	0.8	–	–	0.7	0.3	–	–	0.2	0.3	0.3	–	–
4 or more	1.3	1.5	1.0	0.5	0.4	2.3	0.9	2.0	3.0	1.3	0.6	0.3	1.5	2.3	2.7	1.3	–
3 or more	5.0	5.2	4.8	2.7	3.1	7.8	4.3	6.4	10.0	4.3	3.5	2.5	5.4	8.5	7.3	4.3	1.5
2 or more	16.8	17.8	15.9	13.2	15.8	16.7	15.3	21.6	26.9	17.7	14.2	9.3	17.1	25.4	19.0	12.6	11.7
1 or more	45.1	48.1	42.4	45.5	41.9	47.7	42.1	47.6	62.8	45.6	39.8	33.6	47.1	56.4	53.0	39.9	39.7
No low scores	54.9	51.9	57.6	54.5	58.1	52.3	57.9	52.4	37.2	54.4	60.2	66.4	52.9	43.6	47.0	60.1	60.3
≤9th percentile																	
5 low scores	0.1	0.2	–	–	–	0.4	–	–	–	0.3	–	–	0.1	–	0.3	–	–
4 or more	0.3	0.3	0.3	0.5	–	0.4	0.4	0.3	1.0	0.3	–	–	0.2	–	0.3	0.3	–
3 or more	1.7	1.5	1.8	0.5	1.2	2.7	2.1	1.7	3.7	1.0	1.2	0.9	1.2	2.0	2.0	0.7	0.3
2 or more	6.8	7.7	6.0	4.5	6.2	6.2	10.2	6.8	13.3	6.9	5.0	2.5	4.4	9.1	4.3	2.7	1.5
1 or more	29.1	31.7	26.8	29.1	27.3	29.1	32.3	28.0	44.9	29.5	23.3	20.1	24.9	36.2	27.7	19.9	16.3
No low scores	70.9	68.3	73.2	70.9	72.7	70.9	67.7	72.0	55.1	70.5	76.7	79.9	75.1	63.8	72.3	80.1	83.7
≤5th percentile																	
5 high scores	0.1	0.2	–	–	–	0.4	–	–	–	0.3	–	–	–	–	–	–	–
4 or more	0.2	0.2	0.1	–	–	0.4	–	0.3	0.3	0.3	–	–	0.1	–	0.3	–	–
3 or more	0.3	0.3	0.3	–	–	0.8	0.4	0.3	1.0	0.3	–	–	0.4	0.3	1.0	–	0.3
2 or more	2.4	2.7	2.2	1.4	1.2	3.5	3.4	2.7	5.0	2.3	2.1	0.6	2.3	4.2	3.0	1.0	0.9

(Continued)

TABLE 1 (Continued)

	Age-adjusted norms										Demographically-adjusted norms				
	Total Sample		Boys		Girls		Age				Crystallized composite				
							7–8	9–10	11–12	13–14	15–17	≤89	90–99	100–109	≥110
Sample size	1,269		597		672		220	260	258	235	296	301	305	339	324
1 or more	15.9		17.3		14.7		17.7	16.9	18.6	13.6	13.2	29.2	14.8	11.8	9.0
No low scores	84.1		82.7		85.3		82.3	83.1	81.4	86.4	86.8	70.8	85.2	88.2	91.0
≤2nd percentile															
3 or more	0.1		–		0.1		–	–	–	–	0.3	0.3	–	–	–
2 or more	0.6		0.7		0.4		–	–	0.4	0.9	1.4	1.3	0.3	0.6	–
1 or more	7.1		9.2		5.2		7.7	6.5	8.1	6.4	6.8	14.3	3.9	8.0	2.5
No low scores	92.9		90.8		94.8		92.3	93.5	91.9	93.6	93.2	85.7	96.1	92.0	97.5

NIHTB-CB, National Institutes of Health Toolbox for the Assessment of Neurological and Behavioral Function Cognition Battery. All values represent cumulative percentages except for the rows labeled “No low scores,” which provide the percentage of the normative sample with no scores falling below the low score cut-offs. Age norms are provided as age-corrected standard scores (M = 100, SD = 15) and fully demographic corrected norms are provided as T scores (M = 50, SD = 10) adjusted for age, sex, maternal education level, and race/ethnicity. The NIHTB-CB includes five tests of fluid abilities: Flanker Inhibitory Control and Attention Test, Picture Sequence Memory Test, List Sorting Working Memory Test, Dimensional Change Card Sort Test, and Pattern Comparison Processing Speed Test. Two tests measure crystallized abilities, from which the crystallized composite is calculated: Picture Vocabulary Test and Oral Reading Recognition Test.

each test would be relatively close to the percentile rank (i.e., approximately 16% would obtain a score at or below the 16th percentile). However, given that we excluded youth from the normative sample who had preexisting conditions, this would mean that the percentage of included participants scoring below the percentile cutoff for each test was somewhat lower than the percentile cutoff would suggest due to excluding children and adolescents who, on average, potentially scored lower on the NIHTB-CB. Nonetheless, our results highlight the important finding that the frequency of obtaining one or more low scores across a battery (so called multivariate base rates) differ markedly from the expected frequency for any given test considered in isolation.

Estimated overall intellectual ability (i.e., the crystallized composite scores) was strongly associated with the frequency of low NIHTB-CB fluid test scores. For example, nearly two-thirds of youth (i.e., 62.8%) with a low average crystallized composite (i.e., SS ≤ 89) obtained at least one score at or below the 16th percentile, compared to one-third of youth (i.e., 33.6%) with a high average crystallized composite (i.e., SS ≥ 110). Numerous prior studies have found that intellectual ability is associated with the likelihood of obtaining low scores across pediatric test batteries (Brooks, 2010, 2011; Brooks et al., 2009; Cook et al., 2019). Of note, in a study of low scores among an inpatient pediatric psychiatric sample, estimated intellectual ability was the only independent predictor of low scores (Gaudet et al., 2019).

4.1 Algorithms for identifying developmental cognitive weakness or acquired cognitive impairment

The flexible algorithms presented in Table 3 can be used to identify potential cognitive weakness or impairment. For example, when using age-corrected norms based on the full sample, there are two options that classify around 16% of the sample as having possible cognitive weakness or impairment. For the first option (Algorithm #1), there are multiple successive cutoffs. If a child obtains four or more scores ≤ 25th percentile, that would qualify as cognitive weakness or impairment. If not, a clinician or researcher would apply the next cutoff: three or more scores ≤ 16th percentile. If that criterion is met, that would qualify as cognitive weakness or impairment. If not, the third and final criterion would be applied: two or more scores ≤ 9th percentile. Among the children and adolescents from the NIHTB-CB standardization sample included in this study, only 9.7% met at least one of these criteria and thus could be classified showing cognitive weakness or impairment on the fluid tests using Algorithm #1 (see Table 3). For the second option (Algorithm #2), there is only one criterion: three or more scores ≤ 25th percentile, with a base rate of 13.3% of children and adolescents in this sample.

There are also algorithms provided for youth stratified by their NIHTB-CB crystallized composite. For example, there are two options (Algorithm #10 and Algorithm #11) provided for those with crystallized abilities in the average range (i.e., age-adjusted SS = 100–109 or demographically adjusted T = 50–56). Considering the algorithms for demographically adjusted scores, cognitive weakness or impairment could be identified if (a) the examinee obtains three or more scores ≤ 25th percentile or, if not, the examinee obtains two or more scores ≤ 9th percentile (Algorithm #10); or (b) the examinee obtains two or more scores ≤ 16th percentile (Algorithm

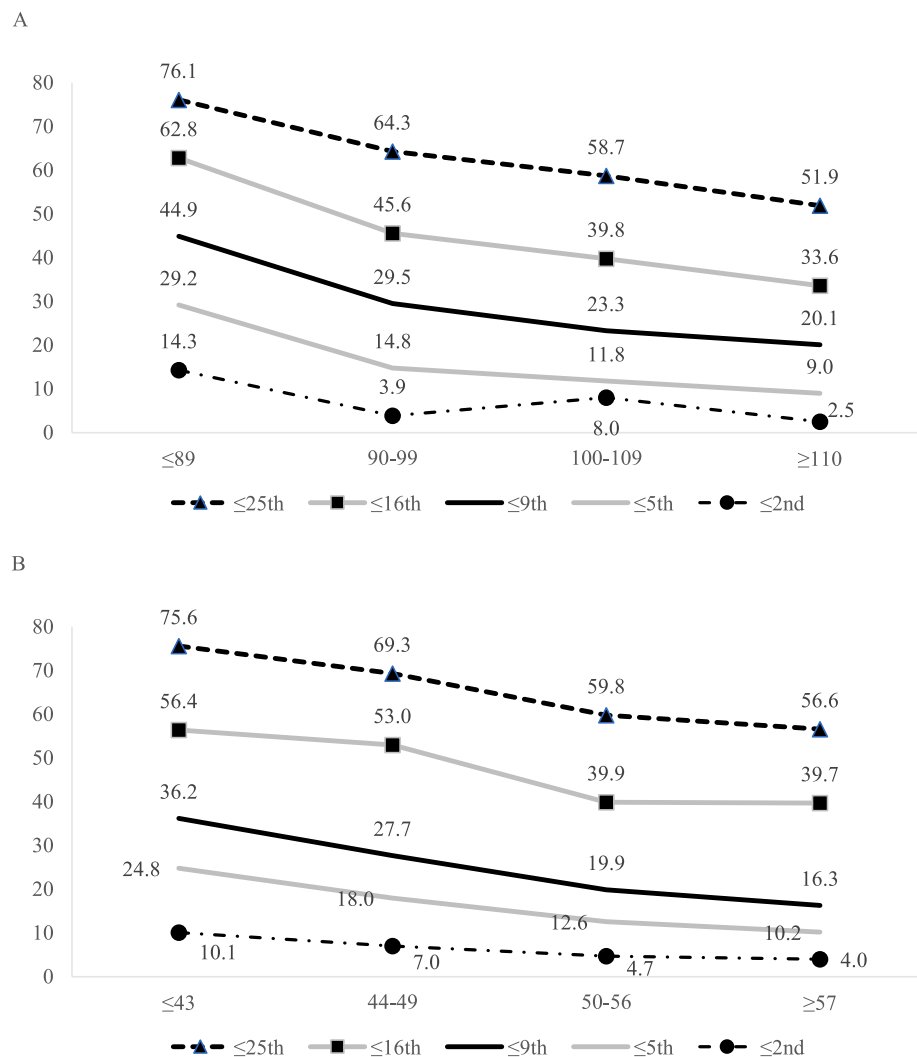


FIGURE 1

Base rates of one or more low fluid subtest scores across various cutoff scores stratified by NIHTB-CB crystallized composite (5 fluid scores were used in base rate calculations). (A) Base rates of one or more low age-adjusted fluid subtest scores across various cutoff scores stratified by NIHTB-CB age-adjusted crystallized composite (5 fluid scores were used in base rate calculations). (B) Base rates of one or more low demographically adjusted fluid subtest scores across various cutoff scores stratified by NIHTB-CB demographically adjusted crystallized composite (5 demographically adjusted fluid scores were used in base rate calculations).

#11). For Algorithm #10, 10.6% of children and adolescents from the NIHTB-CB normative sample had one of the two performance patterns. For Algorithm #11, 12.6% of the NIHTB-CB normative sample had this performance pattern.

It is important to note that these algorithms are meant to be selected *a priori* and then applied to interpret the results. This is critical to avoid inflating the false positive rate. For example, assume a child obtained a crystallized composite score of 99. As shown in Table 2, it is *Uncommon* for children with age-adjusted crystallized composites between 90 and 99 to obtain (a) 4+ scores ≤ 25th percentile (BR = 5.6%), (b) 3+ scores ≤ 16th percentile (BR = 4.3%), and (c) 2+ scores ≤ 9th percentile (BR = 6.9%). However, per Algorithm #7 in Table 3, the likelihood of an examinee obtaining *any* of these three performance patterns is 11.1%. If a clinician or researcher were to apply each of these performance patterns in succession, rather than selecting and applying one criterion *a priori*, the base rate and, therefore the probability of false-positive findings,

would be considerably higher than any of the individual performance patterns applied in isolation.

4.2 Clinical implications

These results align with many prior studies demonstrating that the likelihood of a child or adolescent obtaining a low score is not equivalent to the percentile rank for that score. That is, a score at the 9th percentile on one of the NIHTB-CB fluid subtests indicates that only 9% of individuals score at that level or lower when administered just that test; but the likelihood of a child or adolescent obtaining at least one score at or below the 9th percentile when administered all five fluid subtests is 29.1% (i.e., nearly one third of the normative sample using age-adjusted norms). Base rates help reduce overinterpretation of isolated low scores, and false positive classifications of cognitive weakness, by

TABLE 2 Classification ranges for number of low NIHTB-CB fluid subtest scores for children and adolescents.

	Age-adjusted norms												Demographic-adjusted norms				
	Total Sample	Boys	Girls	Age					Crystallized composite				Total Sample	Crystallized composite			
				7–8	9–10	11–12	13–14	15–17	≤89	90–99	100–109	≥110		≤43	44–49	50–56	≥57
Sample size	1,269	597	672	220	260	258	235	296	301	305	339	324	1,233	307	300	301	325
≤25th percentile																	
Broadly normal	0–2	0–2	0–2	0–2	0–2	0–2	0–2	0–2	0–2	0–2	0–2	0–1	0–2	0–2	0–2	0–2	0–1
Below average	3	3	3	–	3	3	3	3	3	3	3	2	3	3	3	3	2
Uncommon	4	4	4	3	4	4	4	4	4	4	–	3	4	4	4	4	3
Very uncommon	5	5	5	4–5	5	5	5	5	5	5	4–5	4–5	5	5	5	5	4–5
≤16th percentile																	
Broadly normal	0–1	0–1	0–1	0–1	0–1	0–1	0–1	0–1	0–2	0–1	0–1	0–1	0–1	0–2	0–1	0–1	0–1
Below average	2	2	2	2	2	2	2	2	3	2	2	–	2	–	2	2	2
Uncommon	3	3	3	–	3	3	3	3	4	3	3	2	3	3	3	3	–
Very uncommon	4–5	4–5	4–5	3–5	4–5	4–5	4–5	4–5	5	4–5	4–5	3–5	4–5	4–5	4–5	4–5	3–5
≤9th percentile																	
Broadly normal	0–1	0–1	0–1	0–1	0–1	0–1	0–1	0–1	0–1	0–1	0	0	0	0–1	0–1	0	0
Below average	–	–	–	–	–	–	2	–	2	–	1	1	1	–	–	1	1
Uncommon	2	2	2–3	2	2	2	–	2	3	2	2	–	2	2	2	–	–
Very uncommon	3–5	3–5	4–5	3–5	3–5	3–5	3–5	3–5	4–5	3–5	3–5	2–5	3–5	3–5	3–5	2–5	2–5
≤5th percentile																	
Broadly normal	0	0	0	0	0	0	0	0	0–1	0	0	0	0	0	0	0	0
Below average	1	1	1	1	1	1	1	1	–	1	1	–	1	1	1	1	1
Uncommon	–	–	–	–	–	2	2	–	2	–	–	1	–	2	2	–	–
Very uncommon	2–5	2–5	2–5	2–5	2–5	3–5	3–5	2–5	3–5	2–5	2–5	2–5	2–5	3–5	3–5	2–5	2–5
≤2nd percentile																	
Broadly normal	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Below average	–	–	–	–	–	–	–	–	1	–	–	–	–	1	–	–	–
Uncommon	1	1	1	1	1	1	1	1	–	1	1	–	1	–	1	1	1
Very uncommon	2–5	2–5	2–5	2–5	2–5	2–5	2–5	2–5	2–5	2–5	2–5	1–5	2–5	2–5	2–5	2–5	2–5

NIHTB-CB, National Institutes of Health Toolbox for the Assessment of Neurological and Behavioral Function Cognition Battery. Age norms are provided as age-corrected standard scores (SS; M = 100, SD = 15) and fully demographic corrected norms are provided as T scores (M = 50, SD = 10) adjusted for age, sex, maternal education level, and race/ethnicity. The NIHTB-CB includes five tests of fluid abilities: Flanker Inhibitory Control and Attention Test, Picture Sequence Memory Test, List Sorting Working Memory Test, Dimensional Change Card Sort Test, and Pattern Comparison Processing Speed Test. Two tests measure crystallized abilities, from which the crystallized composite is calculated: Picture Vocabulary Test and Oral Reading Recognition Test. The classification ranges in the first column are as follows: Broadly normal: this number of low scores were obtained by the top 75% of the normative sample; Below average: this number of low scores were obtained by less than 25% of the normative sample; Uncommon: this number of low scores were obtained by less than 10% of the normative sample, and Very uncommon: this number of low scores were obtained by less than 3% of the normative sample.

TABLE 3 Flexible algorithms for identifying cognitive weakness or impairment on the NIHTB-CB fluid subtests for children and adolescents ages 7–17.

Comparison group	12 algorithms		Age-adjusted norms		Demographically-adjusted norms	
			Cutoff base rate	Overall base rate	Cutoff base rate	Overall base rate
Full sample (ages 7–17)	1)	4+ scores ≤ 25th, or	4.4%	9.7%	4.3%	7.9%
		3+ scores ≤ 16th, or	5.0%	–	5.4%	–
		2+ scores ≤ 9th	6.8%	–	4.4%	–
	2)	3+ scores ≤ 25th	13.3%	13.3%	14.9%	14.9%
	3)	2+ scores ≤ 16th	16.8%	16.8%	17.1%	17.1%
Crystallized composite SS ≤ 89/T ≤ 43 (low average)	4)	5 scores ≤ 25th, or	2.0%	7.3%	1.3%	6.2%
		4+ scores ≤ 16th, or	3.0%	–	2.3%	–
		3+ scores ≤ 9th, or	3.7%	–	2.0%	–
		2+ scores ≤ 5th	5.0%	–	4.2%	–
	5)	3+ scores ≤ 16th	10.0%	10.0%	8.5%	8.5%
	6)	2+ scores ≤ 9th	13.3%	13.3%	9.1%	9.1%
Crystallized composite SS = 90–99/T = 44–49 (average)	7)	4+ scores ≤ 25th, or	5.6%	11.1%	5.7%	9.3%
		3+ scores ≤ 16th, or	4.3%	–	7.3%	–
		2+ scores ≤ 9th	6.9%	–	4.3%	–
	8)	3+ scores ≤ 25th	15.7%	15.7%	18.0%	18.0%
	9)	2+ scores ≤ 16th	17.7%	17.7%	19.0%	19.0%
Crystallized composite SS = 100–109/T = 50–56 (average)	10)	3+ scores ≤ 25th, or	10.6%	11.5%	10.3%	10.6%
		2+ scores ≤ 9th	5.0%	–	2.7%	–
	11)	2+ scores ≤ 16th	14.2%	14.2%	12.6%	12.6%
Crystallized composite SS ≥ 110/T ≥ 57 (above average)	12)	3+ scores ≤ 25th, or	7.1%	11.1%	9.2%	14.2%
		2+ score ≤ 16th	9.3%	–	11.7%	–

NIHTB-CB, National Institutes of Health Toolbox for the Assessment of Neurological and Behavioral Function Cognition Battery. Age norms are provided as age-corrected standard scores (SS; M = 100, SD = 15) and fully demographic corrected norms are provided as T scores (M = 50, SD = 10) adjusted for age, sex, maternal education level, and race/ethnicity. The NIHTB-CB includes five tests of fluid abilities: Flanker Inhibitory Control and Attention Test, Picture Sequence Memory Test, List Sorting Working Memory Test, Dimensional Change Card Sort Test, and Pattern Comparison Processing Speed Test. Two tests measure crystallized abilities, from which the crystallized composite is calculated: Picture Vocabulary Test and Oral Reading Recognition Test.

considering that low scores are common among youth in the general population, in the absence of any illness, injury, or disorder.

The clinical importance of multivariate base rates is highlighted when considering clinical diagnoses that consider cognitive test results in their criteria, such as the DSM-5 Mild Neurocognitive Disorder, which designates neuropsychological test performances 1–2 SDs below the normative mean (i.e., between the 3rd and 16th percentiles) as indicative of a modest impairment in cognitive functioning (American Psychiatric Association, 2013). Applying those criteria to the NIHTB-CB five fluid subtests, 45.1% (using age-adjusted norms) to 47.1% (using demographically adjusted norms) of children and adolescents from the normative sample would meet the psychometric criteria

for mild neurocognitive disorder, which is clearly an unacceptably high false positive rate.

Conversely, base rates help reduce false negatives by considering unusual performance patterns that otherwise might be considered normal. For example, obtaining three or more scores within a conventional *low average* classification range (i.e., 10th–25th percentile) is *Uncommon* for children and adolescents of high average crystallized ability, occurring in fewer than 10% of the normative sample at this ability level. Some clinicians may consider these individual scores as broadly normal, but in aggregate, they might reflect an unusual pattern of performance potentially indicative of a decline from prior levels of cognitive functioning (in the appropriate clinical context, such as after a moderate traumatic brain injury or a critical illness requiring hospitalization or intensive care).

It is important to carefully consider factors that might be associated with obtaining low scores, as opposed to relying fully on multivariate base rates to guide an interpretation of low individual scores or low patterns of performances. Examinees might obtain low scores on the NIHTB-CB due to such factors as a longstanding developmental cognitive weakness, an acquired deficit due to a brain injury or medical condition, inconsistent or suboptimal effort, misunderstanding test instructions, sacrificing speed for accuracy (e.g., on the Pattern Comparison test), fatigue during testing, or anxiety associated with testing.

In contrast, it is possible that a single low test score might represent an area of clinical focus, particularly if there is evidence of associated functional difficulties in daily life. Thus, though a single isolated low score might not provide strong evidence of cognitive impairment *per se*, it might still reflect an area of weakness amenable to some form of accommodation, intervention, or remediation efforts. Thus, clinicians are encouraged to consider base rate information in the context of other sources of data to include information about whether and how clinical conclusions drawn from base rate information correlate with everyday functioning.

The base rates and algorithms reported in this study are designed to supplement clinical interpretation and require an appreciation of additional sources of clinical information to inform a diagnosis of cognitive impairment. Low test scores alone do not define cognitive impairment, and should be interpreted in addition to observations, developmental history, medical history, neuroimaging (if available), and presenting clinical concerns. Neither the base rates, nor the algorithms, definitively establish the presence or absence of a clinical condition, loss of functioning due to an injury or illness, or clinically significant problems that require intervention. Like other normative and psychometric information, these base rates are provided to strengthen the scientific basis of test score interpretation in the context of other information and details regarding a specific case (Brooks and Iverson, 2012).

4.3 Limitations

The reported base rates apply to the interpretation of all five fluid tests among participants administered all seven NIHTB-CB subtests (i.e., five fluid and two crystalized tests). If fewer subtests are administered in practice, these base rates and algorithms will not be applicable. These base rates also represent point estimates of cumulative percentages based on the normative sample. Clinicians and researchers applying these percentages to the interpretation of cognitive test performances should be mindful of uncertainty surrounding these estimates, as is typical for any individual cognitive test score interpretation. This uncertainty would be especially true for stratifications with smaller sample sizes, as opposed to estimates for the full normative sample.

There are longstanding conceptual and psychometric limitations and concerns relating to cognitive profile analyses, such as scatter, strengths and weaknesses, and discrepancy scores in school psychology (McGill et al., 2018) that also are relevant and important for clinical neuropsychology. Principles from evidence-based medicine can be used to mitigate some of these conceptual and psychometric limitations, such as tying assessment

data directly to clinical conclusions or decision-making using Bayesian methods (Youngstrom, 2013)—an approach that can be applied when considering multivariate base rates and clinical algorithms, such as those provided in the present study.

Another limitation pertains to the crystalized composite. The crystalized composite may serve as an estimate of “premorbid” (i.e., longstanding) functioning when premorbid testing data is unavailable. Performance on the crystalized composite is relatively resistant to the effects of brain injury (Tulsky et al., 2017), but it is not impervious to the effects of neurological injury, considering verbal intelligence can be reduced in acute and chronic pediatric brain injury (Babikian and Asarnow, 2009). Therefore, when evaluating children with acquired injuries to the brain it is important to appreciate that their obtained crystalized composite scores might not closely correspond with their premorbid functioning. This is especially important to consider for children whose obtained crystalized composite falls on the border between classification ranges. For example, imagine an adolescent who sustained a moderate traumatic brain injury 1 month prior to completing the NIHTB-CB. If that youth obtained a crystalized composite score of 109, the clinician might choose to assume that the adolescent’s premorbid (i.e., pre-injury) score was more likely to be 110 or greater, and thus apply base rate comparisons, or algorithms, based on having estimated longstanding intellectual abilities in the above average, not the average, classification range. Of note, however, stratifying base rates using a measure of crystalized intelligence (as defined by the NIHTB-CB with measures of receptive vocabulary and single word reading) may underestimate the abilities of English language learners (Ortiz, 2019) and youth with certain neurodevelopmental disorders, such as reading disorder (i.e., dyslexia) and language disorder (de Bree et al., 2022; Gray et al., 2019).

5 Conclusion and future directions

This study provides multivariate base rates of low scores for children and adolescents on the NIHTB-CB. It also provides flexible diagnostic algorithms that may be useful for clinicians and researchers seeking criteria to operationalize cognitive impairment. These algorithms can be helpful when evaluating patients and participants of varying levels of intellectual ability, allowing for different operational definitions of cognitive impairment based on crystalized ability. They may also have value when examining individuals of diverse cultural backgrounds, with algorithms based on demographically adjusted norms that consider age, gender, race/ethnicity, and maternal education as a proxy of socioeconomic status. The goal of such demographically adjusted algorithms would be to improve the accuracy of neuropsychological assessment and test score interpretation of minoritized groups. The algorithms reported in this study would benefit from validation within clinical samples to determine whether an individual classified as having cognitive impairment is more likely to have a clinical condition, functional impairment, or evidence of structural or functional changes on neuroimaging.

Data availability statement

The datasets analyzed for this study are publicly available (Gershon, 2016).

Ethics statement

The studies involving humans were approved by the Institutional Review Board of Mass General Brigham (Protocol #: 2020P000504). The studies were conducted in accordance with the local legislation and institutional requirements. The ethics committee/institutional review board waived the requirement of written informed consent for participation from the participants or the participants' legal guardians/next of kin because written informed consent for participation was not required from the participants or the participants' legal guardians/next of kin in accordance with the national legislation and institutional requirements.

Author contributions

NC: Conceptualization, Investigation, Methodology, Validation, Visualization, Writing – original draft, Writing – review & editing. GI: Conceptualization, Resources, Supervision, Visualization, Writing – original draft, Writing – review & editing. JK: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Validation, Visualization, Writing – original draft, Writing – review & editing.

Funding

The author(s) declare that financial support was received for the research and/or publication of this article. Grant L. Iverson

References

- Akshoomoff, N., Beaumont, J. L., Bauer, P. J., Dikmen, S. S., Gershon, R. C., Mungas, D., et al. (2013). VIII. NIH toolbox cognition battery (CB): composite scores of crystallized, fluid, and overall cognition. *Monogr. Soc. Res. Child Dev.* 78, 119–132. doi: 10.1111/mono.12038
- Akshoomoff, N., Brown, T. T., Bakeman, R., and Hagler, D. J. (2018). Developmental differentiation of executive functions on the NIH toolbox cognition battery. *Neuropsychology* 32, 777–783. doi: 10.1037/neu0000476
- American Psychiatric Association (2013). Diagnostic and statistical manual of mental disorders. 5th Edn. Arlington, VA: American Psychiatric Publishing.
- Babikian, T., and Asarnow, R. (2009). Neurocognitive outcomes and recovery after pediatric TBI: meta-analytic review of the literature. *Neuropsychology* 23, 283–296. doi: 10.1037/a0015268
- Bauer, P. J., Dikmen, S. S., Heaton, R. K., Mungas, D., Slotkin, J., and Beaumont, J. L. (2013). III. NIH toolbox cognition battery (CB): measuring episodic memory. *Monogr. Soc. Res. Child Dev.* 78, 34–48. doi: 10.1111/mono.12033
- Brooks, B. L. (2010). Seeing the forest for the trees: prevalence of low scores on the Wechsler intelligence scale for children, fourth edition (WISC-IV). *Psychol. Assess.* 22, 650–656. doi: 10.1037/a0019781
- Brooks, B. L. (2011). A study of low scores in Canadian children and adolescents on the Wechsler intelligence scale for children, fourth edition (WISC-IV). *Child Neuropsychol.* 17, 281–289. doi: 10.1080/09297049.2010.537255
- Brooks, B. L., and Iverson, G. L. (2010). Comparing actual to estimated base rates of "abnormal" scores on neuropsychological test batteries: implications for interpretation. *Arch. Clin. Neuropsychol.* 25, 14–21. doi: 10.1093/arclin/acp100
- Brooks, B. L., and Iverson, G. L. (2012). "Improving accuracy when identifying cognitive impairment in pediatric neuropsychological assessments" in *Pediatric forensic neuropsychology*. eds. E. M. Sherman and B. L. Brooks (New York, NY: Oxford University Press), 66–88.
- Brooks, B. L., Iverson, G. L., and Holdnack, J. A. (2013a). "Understanding and using multivariate base rates with the WAIS-IV/WMS-IV" in *WAIS-IV, WMS-IV, and ACS: Advanced clinical interpretation*. eds. J. A. Holdnack, L. W. Drozdick, L. G. Weiss and G. L. Iverson (Waltham, MA: Academic Press), 75–102.
- Brooks, B. L., Iverson, G. L., Koushik, N. S., Mazur-Mosiewicz, A., Horton, A. M. Jr., and Reynolds, C. R. (2013b). Prevalence of low scores in children and adolescents on the test of verbal conceptualization and fluency. *Appl. Neuropsychol. Child* 2, 70–77. doi: 10.1080/21622965.2012.742792
- Brooks, B. L., Iverson, G. L., Sherman, E. M., and Holdnack, J. A. (2009). Healthy children and adolescents obtain some low scores across a battery of memory tests. *J. Int. Neuropsychol. Soc.* 15, 613–617. doi: 10.1017/S1355617709090651
- Brooks, B. L., Iverson, G. L., Sherman, E. M., and Roberge, M. C. (2010a). Identifying cognitive problems in children and adolescents with depression using computerized neuropsychological testing. *Appl. Neuropsychol.* 17, 37–43. doi: 10.1080/09084280903526083
- Brooks, B. L., Sherman, E. M., and Iverson, G. L. (2010b). Healthy children get low scores too: prevalence of low scores on the NEPSY-II in preschoolers, children, and adolescents. *Arch. Clin. Neuropsychol.* 25, 182–190. doi: 10.1093/arclin/acq005
- Carlozzi, N. E., Tulskey, D. S., Kail, R. V., and Beaumont, J. L. (2013). VI. NIH toolbox cognition battery (CB): measuring processing speed. *Monogr. Soc. Res. Child Dev.* 78, 88–102. doi: 10.1111/mono.12036

acknowledges unrestricted philanthropic support from the Schoen Adams Research Institute at Spaulding Rehabilitation. The funders of the Research Institute were not involved in the study design, interpretation of data, the writing of this article, or the decision to submit it for publication.

Conflict of interest

GI has received research support from test publishing companies in the past, including PAR, Inc., ImPACT Applications, Inc., and CNS Vital Signs. He receives royalties for one neuropsychological test (Wisconsin Card Sorting Test-64 Card Version). He serves as a scientific advisor for NanoDx®, Sway Operations, LLC, and Highmark, Inc. He has a clinical and consulting practice in forensic neuropsychology, including expert testimony, involving individuals who have sustained mild TBIs. He acknowledges unrestricted philanthropic support from ImPACT Applications, Inc. and the Mooney-Reed Charitable Foundation.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

- Casaleto, 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. *J. Int. Neuropsychol. Soc.* 21, 378–391. doi: 10.1017/S1355617715000351
- Chadwick, L., Roth, E., Minich, N. M., Taylor, H. G., Bigler, E. D., Cohen, D. M., et al. (2021). Cognitive outcomes in children with mild traumatic brain injury: an examination using the National Institutes of Health toolbox cognition battery. *J. Neurotrauma* 38, 2590–2599. doi: 10.1089/neu.2020.7513
- Cook, N. E., Karr, J. E., Brooks, B. L., Garcia-Barrera, M. A., Holdnack, J. A., and Iverson, G. L. (2019). Multivariate base rates for the assessment of executive functioning among children and adolescents. *Child Neuropsychol.* 25, 836–858. doi: 10.1080/09297049.2018.1543389
- Dakopoulos, A., Condy, E., Smith, E., Harvey, D., Kaat, A. J., Coleman, J., et al. (2024). Developmental associations between cognition and adaptive behavior in intellectual and developmental disability. *Res. Sq.* 8:708. doi: 10.21203/rs.3.rs-3684708/v1
- Davis, A. S., Bernat, D. J., and Reynolds, C. R. (2018). Estimation of premorbid functioning in pediatric neuropsychology: review and recommendations. *J. Pediatr. Neuropsychol.* 4, 49–62. doi: 10.1007/s40817-018-0051-x
- de Bree, E. H., Boerma, T., Hakvoort, B., Blom, E., and van den Boer, M. (2022). Word reading in monolingual and bilingual children with developmental language disorder. *Learn. Individ. Differ.* 98:102185. doi: 10.1016/j.lindif.2022.102185
- Gaudet, C. E., Cook, N. E., Kavanaugh, B. C., Studeny, J., and Holler, K. (2019). Prevalence of low test scores in a pediatric psychiatric inpatient population: applying multivariate base rate analyses. *Appl. Neuropsychol. Child* 8, 163–173. doi: 10.1080/21622965.2017.1417126
- Gershon, R. (2016). NIH Toolbox Norming Study. Harvard Dataverse, V4. Available at: <https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/FF4DI7>
- Gershon, R. C., Slotkin, J., Manly, J. J., Blitz, D. L., Beaumont, J. L., Schnipke, D., et al. (2013). IV. NIH toolbox cognition battery (CB): measuring language (vocabulary comprehension and reading decoding). *Monogr. Soc. Res. Child Dev.* 78, 49–69. doi: 10.1111/mono.12034
- Gladstone, E., Narad, M. E., Hussain, F., Quatman-Yates, C. C., Hugentobler, J., Wade, S. L., et al. (2019). Neurocognitive and quality of life improvements associated with aerobic training for individuals with persistent symptoms after mild traumatic brain injury: secondary outcome analysis of a pilot randomized clinical trial. *Front. Neurol.* 10:1002. doi: 10.3389/fneur.2019.01002
- Gray, S., Fox, A. B., Green, S., Alt, M., Hogan, T. P., Petscher, Y., et al. (2019). Working memory profiles of children with dyslexia, developmental language disorder, or both. *J. Speech Lang. Hear. Res.* 62, 1839–1858. doi: 10.1044/2019_JSLHR-L-18-0148
- Hackman, D. A., Cserbik, D., Chen, J. C., Berhane, K., Minaravesh, B., McConnell, R., et al. (2021). Association of local variation in neighborhood disadvantage in metropolitan areas with youth neurocognition and brain structure. *JAMA Pediatr.* 175:e210426. doi: 10.1001/jamapediatrics.2021.0426
- Holdnack, J. A., Iverson, G. L., Silverberg, N. D., Tulsy, D. S., and Heinemann, A. W. (2017a). NIH toolbox cognition tests following traumatic brain injury: frequency of low scores. *Rehabil. Psychol.* 62, 474–484. doi: 10.1037/rep0000145
- Holdnack, J. A., Tulsy, D. S., Brooks, B. L., Slotkin, J., Gershon, R., Heinemann, A. W., et al. (2017b). Interpreting patterns of low scores on the NIH toolbox cognition battery. *Arch. Clin. Neuropsychol.* 32, 574–584. doi: 10.1093/arclin/acx032
- Isaiah, A., Ernst, T. M., Liang, H., Ryan, M., Cunningham, E., Rodriguez, P. J., et al. (2023). Associations between socioeconomic gradients and racial disparities in preadolescent brain outcomes. *Pediatr. Res.* 94, 356–364. doi: 10.1038/s41390-022-02399-9
- Iverson, G. L., and Karr, J. E. (2021). Improving the methodology for identifying mild cognitive impairment in intellectually high-functioning adults using the NIH toolbox cognition battery. *Front. Psychol.* 12:724888. doi: 10.3389/fpsyg.2021.724888
- Jones, D. R., Dallman, A., Harrop, C., Whitten, A., Pritchett, J., Lecavalier, L., et al. (2022). Evaluating the feasibility of the NIH toolbox cognition battery for autistic children and adolescents. *J. Autism Dev. Disord.* 52, 689–699. doi: 10.1007/s10803-021-04965-2
- Joseph, R. M., Hooper, S. R., Heeren, T., Santos, H. P. Jr., Frazier, J. A., Venuti, L., et al. (2022). Maternal social risk, gestational age at delivery, and cognitive outcomes among adolescents born extremely preterm. *Paediatr. Perinat. Epidemiol.* 36, 654–664. doi: 10.1111/ppe.12893
- Karr, J. E., Garcia-Barrera, M. A., Holdnack, J. A., and Iverson, G. L. (2017). Using multivariate base rates to interpret low scores on an abbreviated battery of the delis-Kaplan executive function system. *Arch. Clin. Neuropsychol.* 32, 297–305. doi: 10.1093/arclin/acw105
- Karr, J. E., Garcia-Barrera, M. A., Holdnack, J. A., and Iverson, G. L. (2018). Advanced clinical interpretation of the delis-Kaplan executive function system: multivariate base rates of low scores. *Clin. Neuropsychol.* 32, 42–53. doi: 10.1080/13854046.2017.1334828
- Karr, J. E., Mindt, M. R., and Iverson, G. L. (2022a). A multivariate interpretation of the Spanish-language NIH toolbox cognition battery: the Normal frequency of low scores. *Arch. Clin. Neuropsychol.* 37, 338–351. doi: 10.1093/arclin/acab064
- Karr, J. E., Rivera Mindt, M., and Iverson, G. L. (2022b). Algorithms for operationalizing mild cognitive impairment using the Spanish-language NIH toolbox cognition battery. *Arch. Clin. Neuropsychol.* 37, 1608–1618. doi: 10.1093/arclin/acac042
- Kavanaugh, B. C., Cancilliere, M. K., Fryc, A., Tirrell, E., Oliveira, J., Oberman, L. M., et al. (2019). Measurement of executive functioning with the National Institute of Health Toolbox and the association to anxiety/depressive symptomatology in childhood/adolescence. *Child Neuropsychol.* 26, 754–769. doi: 10.1080/09297049.2019.1708295
- McGill, R. J., Dombrowski, S. C., and Canivez, G. L. (2018). Cognitive profile analysis in school psychology: history, issues, and continued concerns. *J. Sch. Psychol.* 71, 108–121. doi: 10.1016/j.jsp.2018.10.007
- Molinario, M., Adams, H. R., Mwanza-Kabaghe, S., Mbewe, E. G., Kabundula, P. P., Mwemba, M., et al. (2021). Evaluating the relationship between depression and cognitive function among children and adolescents with HIV in Zambia. *AIDS Behav.* 25, 2669–2679. doi: 10.1007/s10461-021-03193-0
- Ortiz, S. O. (2019). On the measurement of cognitive abilities in English learners. *Calif. School Psychol.* 23, 68–86. doi: 10.1007/s40688-018-0208-8
- Schmithorst, V. J., Badaly, D., Beers, S. R., Lee, V. K., Weinberg, J., Lo, C. W., et al. (2022). Relationships between regional cerebral blood flow and neurocognitive outcomes in children and adolescents with congenital heart disease. *Semin. Thorac. Cardiovasc. Surg.* 34, 1285–1295. doi: 10.1053/j.semtcvs.2021.10.014
- Schoenberg, M. R., Lange, R. T., Brickell, T. A., and Saklofske, D. H. (2007). Estimating premorbid general cognitive functioning for children and adolescents using the American Wechsler intelligence scale for children fourth edition: demographic and current performance approaches. *J. Child Neurol.* 22, 379–388. doi: 10.1177/0883073807301925
- Shapiro, A. L. B., Dabelea, D., Stafford, J. M., D'Agostino, R., Pihoker, C., and Liese, A. D. (2021). Cognitive function in adolescents and young adults with youth-onset type 1 versus type 2 diabetes: the SEARCH for diabetes in youth study. *Diabetes Care* 44, 1273–1280. doi: 10.2337/dc20-2308
- Shields, R. H., Kaat, A., Sansone, S. M., Michalak, C., Coleman, J., Thompson, T., et al. (2023). Sensitivity of the NIH toolbox to detect cognitive change in individuals with intellectual and developmental disability. *Neurology* 100, e778–e789. doi: 10.1212/WNL.000000000000201528
- Siciliano, R. E., Murphy, L. K., Prussien, K. V., Henry, L. M., Watson, K. H., Patel, N. J., et al. (2021). Cognitive and attentional function in children with hypoplastic left heart syndrome: a pilot study. *J. Clin. Psychol. Med. Settings* 28, 619–626. doi: 10.1007/s10880-020-09753-1
- Solomon, M., Gordon, A., Iosif, A. M., Gedder, R., Krug, M. K., Mundy, P., et al. (2021). Using the NIH toolbox to assess cognition in adolescents and young adults with autism Spectrum disorders. *Autism Res.* 14, 500–511. doi: 10.1002/aur.2399
- Taylor, R. L., Cooper, S. R., Jackson, J. J., and Barch, D. M. (2020). Assessment of neighborhood poverty, cognitive function, and prefrontal and hippocampal volumes in children. *JAMA Netw. Open* 3:e2023774. doi: 10.1001/jamanetworkopen.2020.23774
- Thompson, M. D., Martin, R. C., Grayson, L. P., Ampah, S. B., Cutter, G., Szaflarski, J. P., et al. (2020). Cognitive function and adaptive skills after a one-year trial of cannabidiol (CBD) in a pediatric sample with treatment-resistant epilepsy. *Epilepsy Behav.* 111:107299. doi: 10.1016/j.yebeh.2020.107299
- Tulsy, D. S., Carlozzi, N. E., Chevalier, N., Espy, K. A., Beaumont, J. L., and Mungas, D. (2013). V. NIH toolbox cognition battery (CB): measuring working memory. *Monogr. Soc. Res. Child Dev.* 78, 70–87. doi: 10.1111/mono.12035
- Tulsy, D. S., Carlozzi, N. E., Holdnack, J., Heaton, R. K., Wong, A., Goldsmith, A., et al. (2017). Using the NIH toolbox cognition battery (NIHTB-CB) in individuals with traumatic brain injury. *Rehabil. Psychol.* 62, 413–424. doi: 10.1037/rep0000174
- Wallace, J., Ceschin, R., Lee, V. K., Beluk, N. H., Burns, C., Beers, S., et al. (2024). Psychometric properties of the NIH toolbox cognition and emotion batteries among children and adolescents with congenital heart defects. *Child Neuropsychol.* 30, 967–986. doi: 10.1080/09297049.2024.2302690
- Watson, W., Pedowitz, A., Nowak, S., Neumayer, C., Kaplan, E., and Shah, S. (2020). Feasibility of National Institutes of Health toolbox cognition battery in pediatric brain injury rehabilitation settings. *Rehabil. Psychol.* 65, 22–30. doi: 10.1037/rep0000309
- Weintraub, S., Dikmen, S. S., Heaton, R. K., Tulsy, D. S., Zelazo, P. D., Bauer, P. J., et al. (2013). Cognition assessment using the NIH toolbox. *Neurology* 80, S54–S64. doi: 10.1212/WNL.0b013e3182872ded
- Youngstrom, E. A. (2013). Future directions in psychological assessment: combining evidence-based medicine innovations with psychology's historical strengths to enhance utility. *J. Clin. Child Adolesc. Psychol.* 42, 139–159. doi: 10.1080/15374416.2012.736358
- Zelazo, P. D., Anderson, J. E., Richler, J., Wallner-Allen, K., Beaumont, J. L., and Weintraub, S. (2013). II. NIH toolbox cognition battery (CB): measuring executive function and attention. *Monogr. Soc. Res. Child Dev.* 78, 16–33. doi: 10.1111/mono.12032