HHS Public Access

Author manuscript

Pediatr Neurol. Author manuscript; available in PMC 2024 April 01.

Published in final edited form as:

Pediatr Neurol. 2023 April ; 141: 118–132. doi:10.1016/j.pediatrneurol.2023.01.009.

Consensus-Based Evaluation of Outcome Measures in Pediatric Stroke Care: A Toolkit

Samantha J. Feldman, MSc[a], **Lauren A. Beslow, MD, MSCE**[b], **Ryan J. Felling, MD, PhD**[c], **Laura A. Malone, MD, PhD**[d], **Michaela Waak, MD**[e,f], **Stuart Fraser, MD**[g], **Nihal Bakeer, MD**[h], **Jo Ellen M. Lee, CPNP-PC**[i], **Victoria Sherman, PhD**[j], **Melissa M. Howard, PhD**[k], **Beth Anne Cavanaugh, MD**[l], **Robyn Westmacott, PhD**[a], **Lori C. Jordan, MD, PhD**[m,*]

[a]Department of Psychology, The Hospital for Sick Children, Toronto, Ontario, Canada

[b]Division of Neurology, Children's Hospital of Philadelphia, Departments of Neurology and Pediatrics, Perelman School of Medicine at the University of Pennsylvania, Philadelphia, Pennsylvania

[c]Department of Neurology, Johns Hopkins University School of Medicine, Baltimore, Maryland

[d]Johns Hopkins University School of Medicine and the Kennedy Krieger Institute, Baltimore, Maryland

[e]Pediatric Critical Care Research Group, Child Health Research Centre, The University of Queensland, Queensland, Australia

[f]Pediatric Intensive Care Unit, Queensland Children's Hospital, South Brisbane, Australia

[g]Division of Vascular Neurology, Department of Pediatrics, University of Texas Health Science Center, Houston, Texas

[h]Indiana Hemophilia and Thrombosis Center, Indianapolis, Indiana

[i]Department of Neurology, Nationwide Children's Hospital, Columbus, Ohio

[j]St. Mary's General Hospital, Guelph, Ontario, Canada

[k]Casa Colina Hospital and Centers for Healthcare, Pomona, California

[l]Division of Pediatric Neurology, Department of Pediatrics, University of Tennessee Health Science Center, Le Bonheur Children's Hospital, Memphis, Tennessee

[m]Division of Pediatric Neurology, Department of Pediatrics, Vanderbilt University Medical Center, Nashville, Tennessee

Abstract

Following a pediatric stroke, outcome measures selected for monitoring functional recovery and development vary widely. We sought to develop a toolkit of outcome measures that are currently

*Communications should be addressed to: Dr. Jordan; Division of Pediatric Neurology; Department of Pediatrics; Vanderbilt University Medical Center; 2200 Children's Way, DOT 11212; Nashville, TN 37232. lori.jordan@vumc.org (L.C. Jordan).

Supplementary data

Supplementary data related to this article can be found at https://doi.org/10.1016/j.pediatrneurol.2023.01.009.

available to clinicians, possess strong psychometric properties, and are feasible for use within clinical settings. A multidisciplinary group of clinicians and scientists from the International Pediatric Stroke Organization comprehensively reviewed the quality of measures in multiple domains described in pediatric stroke populations including global performance, motor and cognitive function, language, quality of life, and behavior and adaptive functioning. The quality of each measure was evaluated using guidelines focused on responsiveness and sensitivity, reliability, validity, feasibility, and predictive utility. A total of 48 outcome measures were included and were rated by experts based on the available evidence within the literature supporting the strengths of their psychometric properties and practical use. Only three measures were found to be validated for use in pediatric stroke: the Pediatric Stroke Outcome Measure, the Pediatric Stroke Recurrence and Recovery Questionnaire, and the Pediatric Stroke Quality of Life Measure. However, multiple additional measures were deemed to have good psychometric properties and acceptable utility for assessing pediatric stroke outcomes. Strengths and weaknesses of commonly used measures including feasibility are highlighted to guide evidence-based and practicable outcome measure selection. Improving the coherence of outcome assessment will facilitate comparison of studies and enhance research and clinical care in children with stroke. Further work is urgently needed to close the gap and validate measures across all clinically significant domains in the pediatric stroke population.

Keywords

Pediatric stroke; Outcomes; Measurement; Assessment

Introduction

Pediatric stroke is associated with significant morbidity that can affect a child's developmental trajectory.[1] Typical growth and maturation processes in the developing brain are frequently disrupted, resulting in a spectrum of neurological impairments.[1] Consequent deficits may include hemiparesis, language impairment such as aphasia, cognitive difficulties, and social-emotional problems, some of which only emerge later in childhood when developmental and educational demands increase.[2–4] In addition, variability in age at stroke, stroke etiology, and premorbid risk factors make predictions of poststroke recovery trajectories and outcomes challenging.[5–7] Outcome measures are one effective way to monitor recovery and screen for potential emerging deficits.[8] Clinicians can evaluate initial deficits and track impairments during rehabilitation through the use of global functional outcome measures, which are brief, easy to administer, and broadly applicable. Outcome measures that focus on a particular area, such as motor function, language, cognition, adaptive function, or mood and behavior allow for a more precise understanding of a child's strengths and weaknesses. These domain-specific measures often take longer to administer and might require expertise, but provide valuable insight into specific neurological impairments to inform individualized treatment recommendations.

In practical terms, outcome measures help clinicians make decisions about how to direct resources and care. Given their importance, measurement tools should possess high-quality psychometric properties. International programs such as the COSMIN initiative

(Consensus-based Standards for the Selection of Health Measurement Instruments) have been developed to encourage the use of psychometrically sound health instruments.[9–11] COSMIN advocates for systematic evaluation of health instruments, including clinical outcome measures, through a framework with checklists that incorporate validity, reliability, internal consistency, responsiveness, and interpretability. This approach of identifying high-quality measures with a uniform set of criteria may also aid in the development of preferred or gold-standard instruments to standardize the collection of data across clinical sites and hospitals.

A systematic review in 2012 found wide variation in the use of outcome measures in pediatric stroke research, with 34 studies utilizing 38 different outcome measures.[8] Unfortunately, such variation limits the comparison of studies. Harmonizing outcome measures across clinical sites would allow for easier cross-site comparison of pediatric stroke outcomes and treatment results. To encourage the adoption of a common set of outcome measures, we utilized a multidisciplinary team of health care providers with pediatric stroke expertise to evaluate a wide range of commonly used clinical outcome measures in pediatric stroke care using the COSMIN framework. The purpose of the evaluation was to develop an expert-informed compendium of outcome measures, a “toolkit” for the evaluation of children with stroke. Our goal was to provide clinicians and researchers with valuable information related to psychometric properties and practical features of measures to inform their selection within clinical settings and research studies.

Methods

The expert group for the current study was created and coordinated through the International Pediatric Stroke Study (IPSS) and International Pediatric Stroke Organization (IPSO) network, which consists of a multidisciplinary group of clinicians, scientists, and research staff. The IPSO members who participated in the evaluation process were selected to ensure diverse expertise in clinical backgrounds and research areas related to pediatric stroke. The group of 13 contributors included neurologists, a neurointensivist, neuropsychologists, a nurse practitioner, a physical therapist, and a speech and language pathologist.

To develop the measurement toolkit, the expert group evaluated the measures through the following stages.

Stage 1: identification of domains and generation of the initial list of measures for inclusion in the expert review

Commonly used measures for evaluating stroke-related impairment were identified within the clinical research literature and through existing recommendations for the following outcome domains of interest: (1) global performance; (2) motor function; (3) behavioral assessments and adaptive function; (4) cognitive function; (5) language; (6) quality of life (QoL); and (7) mood. The expert team met and reviewed the list of domains and their associated outcome measures. An initial list of measures was discussed to finalize a list for review before stage 2.

Stage 2: quality criteria and content for review measures and REDCap survey ratings of quality criteria

A survey was utilized to rate the importance of quality criteria for subsequent evaluation of the individual measures within the domains of interest. These quality criteria were adapted from the COSMIN guidelines international consensus on measurement properties.[10,11] Additional information to guide the ratings such as practical features and scoring information was reviewed further and approved by expert group members. The working group completed an anonymous REDCap survey in which each member rated the following quality criteria adapted from the COSMIN guidelines on a scale of 1 to 10 with 1 considered least important and 10 considered most important: responsiveness and sensitivity, reliability, repeatability, validity, feasibility, and predictive utility. The operationalized definitions for each quality criterion were provided to each group member to ensure consistent interpretation (refer to Supplementary Document 1).[10,11] The REDCap survey ratings of the importance of each quality criteria provided by each group member were rank ordered, and then each criterion was averaged across the 12 raters with a derived mean level of importance to the expert working group.

Stage 3: literature search and drafting of data tables

The literature search was performed within PubMed using search terms “stroke” OR “pediatric stroke” AND (“outcomes” OR “measures” OR “psychometrics” OR “rehabilitation” OR “neurological rehabilitation” OR “therapy” OR “recovery”). An additional search was undertaken in PubMed and on Google Scholar for each domain (e.g., “cognition” or “language”) with the aforementioned search terms, and for each individual measurement scale (e.g., “behavior rating inventory of executive function” or “clinical evaluation of language fundamentals”) that was included within the designated domain. Tables for each outcome domain were initially developed by a single working group member based on literature review. Scales were identified from the literature search according to quality criteria that included psychometric properties (i.e., responsiveness and sensitivity, reliability, repeatability, validity, feasibility, and predictive utility), time of administration, age range, scoring range, and practical features of each measure. The identified scales were then assigned to the appropriate domains (global performance, motor function, language, etc.) within the evaluation framework.

Stage 4: expert review of data tables

Three expert working group members were assigned to each domain of interest to verify the preliminary information in the table for accuracy and to ensure completeness (refer to Supplementary Tables 1–6). These assignments were determined according to the specialized knowledge and expertise of the members.

Stage 5: rating of individual measures using the modified COSMIN checklist

A modified version of the COSMIN checklist was utilized to assess the quality of each outcome measure (Table 1). With measure names removed to reduce bias, blinded coauthors reviewed all measures within each domain and rated the measures assigning a score for each measure from 1 to 5.

Stage 6: rating and recommendation review of measures for the toolkit

The rated measures were reviewed by additional members of the expert group who did not participate in stage 5, and these members provided comments and feedback.

Results

Stage 1

Outcome measures in the seven domains of interest were identified by the expert working group: (1) Global Performance (eight outcome measures); (2) Motor Function (eleven outcome measures); (3) Behavioral Assessments and Adaptive Function (eight outcome measures); (4) Language (six outcome measures); (5) Quality of Life (four measures); and (6) Cognitive Function (ten measures) and Mood (two measures).

Stage 2

Quality criteria that were selected included responsiveness and sensitivity, reliability, repeatability, validity, feasibility, and predictive utility. Based on working group input, the following content areas were added to the previously described modified COSMIN quality criteria: instrument description, scoring range, and practical features of the measurement tool. The REDCap survey results demonstrated that validity, sensitivity, and responsivity received the highest mean scores indicating their greater importance to the raters. Reliability and feasibility were also reported to be important, with slightly lower mean ratings than validity and sensitivity. Repeatability and describing a measure's practice effect impact were the least important qualities to the majority of expert raters. However, across all measures, there was significant variability in importance ratings of a given criteria. For example, repeatability received the majority of the lowest ratings; however, two experts rated it highly (nine out of ten rating of importance).

Stage 3

7608 articles were screened via title review followed by abstract review when relevant on PubMed and Google Scholar during the literature search. Information from 304 articles and six technical manuals were referenced and used to populate the domain tables.

Stage 4

Domain tables were verified by group members (see Supplemental Tables 1–6).

Stage 5

Tables of domain-specific measures are summarized below and are organized from highest to lowest rated.

Global performance

The Pediatric Stroke Outcome Measure (PSOM) received the highest rating (mean = 4.66, range 4–5) among the eight scales (Table 2). The PSOM was recognized as being the only validated global composite performance measure developed specifically to assess outcomes in pediatric stroke. The PSOM was rated to have good to excellent validity and reliability

and was noted to be easy to administer at the bedside. The Pediatric Stroke Recurrence and Recovery Questionnaire (psRRQ) and the Hammersmith Infant Neurological Examination (HINE) had the next highest ratings, respectively. Importantly, the psRRQ is a derivative of the PSOM, which can be conducted remotely, an important consideration in easing the burden of attending clinic and assessing outcome for children and families who live remote to the stroke center. The HINE was rated highly due to sensitivity and accuracy in identifying mild delays in infants with cerebral palsy relative to typically developing infants and based on evidence supporting its predictive validity, as it is highly correlated with gross motor function at two years and full-scale IQ scores.[18] Other measures within the domain received lower scores because they were not specific, were generally more rough estimates of function, were not validated in pediatric samples, or required multiple informants to accurately identify a child's functional capacity across a number of different domains making their use somewhat impractical except in team-based care settings. The shortcomings of many of the measures within the global performance domain are the limited range and lack of specificity of items that determine a child's overall functioning; thus, substantial changes in functional ability would be necessary to shift scores to a different performance category (from mild to moderate or moderate to severe).[22–24,28,29,31] The limited range and lack of specificity are desirable in that only clinically meaningful differences in scores are likely to be detected; however, this could limit detection of subtle changes that could prove important at an individual level.

Motor function

The Gross Motor Functional Measure (GMFM) was the highest rated outcome measure for motor function (refer to Table 3 and Supplemental Table 2) based on excellent validity and reliability in children with cerebral palsy.[32–38] The GMFM is designed to measure gross motor function over the course of typical development and therefore is sensitive to motor impairments and change over time. The GMFM also rates capacity to complete a movement rather than quality of performance.[36] For children with hemiparesis, this is a meaningful outcome because less fluid, but functional movements may allow children to complete activities of daily living. One significant drawback is that the GMFM is an hour-long assessment, which is not usually practical during a routine clinic appointment. Some of the alternative measures have not been validated in children or in relevant clinical populations (e.g., pediatric stroke or cerebral palsy). These measures also may be impractical or lack adequate sensitivity, reliability, or validity. Future work should aim to validate motor function measures such as the GMFM in pediatric stroke samples. The Dysphagia Disorder Survey is the only validated measure reported for use in pediatric stroke that assesses swallowing and feeding function; however, the Dysphagia Disorder Survey requires substantial training to administer and interpret, and therefore received a lower score (mean rating = 3).[74]

Adaptive functioning and behavior

The Adaptive Behavior Assessment System (ABAS) is one of the most commonly used measures to evaluate adaptive functioning (Table 4) in children and adults and is normed on a large sample representative of the US population. The ABAS received a mean rating of 4.66 and is considered a gold standard in terms of assessment of adaptive functioning. The

ABAS possesses strong psychometric properties (i.e., validity, reliability) and is frequently used in pediatric stroke research.[79] The Child and Adolescent Scale of Participation (CASP) also received high scores (mean = 4.33) and is freely available, unlike the ABAS, which requires a cost per use.[80–84] The Vineland Adaptive Behavior Scale received lower ratings due to variable test-retest reliability and its long administration time.[104]

The Child Behavior Checklist (CBCL), Vanderbilt Assessment Scale (VAS), and the Behavioral Assessment System for Children (BASC) evaluate children's internalizing and externalizing behaviors and assist in diagnosing behavioral and emotional problems in children. All three measures indicate items that correspond with the DSM-V diagnostic criteria for disorders diagnosed in children (e.g., attention-deficit/hyperactivity disorder, generalized anxiety disorder, major depressive disorder, oppositional defiance disorder).[107–109] The CBCL and BASC possess similarly strong psychometric properties, and both received ratings with a mean of 4 and were rated higher than the VAS, which was found to have low-inter-rater reliability and no evidence of discriminant validity (mean rating = 3.66).[107,109] The CBCL possesses good cross-cultural validity; it is available in multiple languages and has been validated in countries outside of North America.[100–102]

Quality of life (QoL)

Four measures of QoL were evaluated, and all were rated positively (mean rating≥4). The Pediatric Stroke Quality of Life Measure (PSQLM) received the highest rating of 4.66. In addition to the PSOM, the PSQLM is one of the few scales developed specifically for children with stroke. The PSQLM has excellent sensitivity and validity, and the items were informed by the experiences of children with stroke and their families.[111] Generic QoL measurement tools such as the PedsQL tend to lack the elements of QoL related to cognition, language, and memory issues, which are of critical importance in pediatric stroke populations.

Mood

The Revised Children's Anxiety and Depression Scale (RCADS) had a higher rating (mean rating 4.33) than the Child Depression Inventory (CDI) (mean rating = ADD). The RCADs has been validated in a large number of children representative of the US population, spans a wide range of ages, and is less costly than the CDI.[119–121]

Cognition

Cognitive performance is often characterized as consisting of five subdomains: memory and learning, language, attention, executive functions, and perceptual and motor functions. Motor function and language were addressed separately, so this section focused primarily on batteries assessing overall intellectual ability, memory and learning, and executive function (Table 5). Two batteries that assess core domains of cognition were evaluated. The Wechsler Intellectual Ability Tests (WPPSI-IV, WISC-V, WAIS-IV) received a mean rating of 4.66. The Wechsler tests are the gold-standard measures to assess intellectual ability across the life span.[160] The child version has been normed on over 2000 children, possesses excellent psychometric validity, and is continually updated and improved upon.[155] The Weschler tests require significant training, have high associated costs, and are known to have practice

effects. The NIH toolbox also received high ratings, with a mean score of 4. The NIH toolbox has a shorter administration time, is less expensive, and has good psychometric validity. The toolbox is far less widely used than the Wechsler tests as it is a newer battery and has only recently been validated for use in children with TBI.

Within the memory subdomain, the California Verbal Learning Test (CVLT-C) was most highly rated with a mean score of 4.66. The CVLT-C is a standardized test with a short administration time, excellent reliability and validity in a pediatric TBI sample, and has been used in a research context in pediatric stroke.[137] Within the subdomain of executive function, the Behavior Rating Inventory of Executive Function received a score of 4.33 due to short administration time, strong psychometric properties, availability in several languages, and ability for parents to complete in clinic.[144] Only one objective attention measure, the Test of Everyday Attention-Child, was included in our evaluation, which received a mean rating of 3.66. The Test of Everyday Attention-Child has good validity; however, it possesses weak test-retest reliability, and some clinicians tend to prefer previous versions of the measure. However, the BASC, CBCL, Conners, and the VAS (covered under the Adaptive Function and Behavior domain) all will identify concerns about attention.

Speech and language

Very few language-specific measures were identified as being commonly used in pediatric stroke. The highest rated measure, the Focus on the Outcome of Communication Under Six has strong psychometric properties and received a mean score of 4. However, the age range is limited and therefore does not have wide applicability across the pediatric age span with low utility as a longitudinal measure. The Clinical Evaluation of Language Fundamentals (CELF) covers a broader range of ages (mean rating = 3.66); however, it is quite long to administer, requires specific expertise to interpret, and requires some degree of motor function.

Discussion

This study identified and evaluated 48 commonly used outcome measures across seven domains of function through a review of the literature and expert ratings by a multidisciplinary group of clinicians who care for children with strokes. Although there are a range of different outcome measures utilized in clinical care and clinical trials in pediatric stroke, existing literature provides little guidance regarding outcome measure quality and utility. Our comprehensive assessment of commonly used instruments addresses an important gap in knowledge regarding outcome measures by providing systematic multi-rater scoring of instruments for their utility across multiple domains of function. The selected outcome measures were evaluated based on their psychometric properties pertaining to the relevant clinical groups (i.e., pediatric stroke, cerebral palsy, TBI). Instrument strengths and weaknesses were summarized from the current literature by expert users. These evaluations can guide outcome measure selection for clinical trials or observational studies. Over the long-term, the use of a shared set of high-quality outcome measures could facilitate comparison between research studies, improve understanding of the recovery

phases following pediatric stroke, and advance pediatric stroke recovery based on this knowledge.

Consistent with a prior systematic review completed a decade ago, most outcome measures have been validated in related populations of children with cerebral palsy or TBI.[8] Only three measures were specifically validated for use in pediatric stroke populations: the PSOM; the psRRQ, which is a derivative of the PSOM; and the PSQLM. The PSOM received the highest rating of the global performance measures, with strengths including construct validity, inter-rater reliability, and ease of use either prospectively or retrospectively.[13] The PSOM has been used in multiple outcome studies and has been previously strongly recommended for prospective clinical trials in pediatric stroke.[13] The psRRQ as a remote administration option expands its use.[16] One inherent limitation of global performance measures is that the corollary of their strengths as general screening measures is their limitation in being able to identify more subtle or focal deficits with sufficient sensitivity. Some global performance measures may misclassify a patient with minor neurological impairment into a more severe category and in turn predict an unnecessarily poorer outcome.[24] For example, on the KOSCHI, a child must meet all criteria outlined within a given category; otherwise, the child would be classified in a lower category suggesting an increased level of disability than might otherwise be warranted. With the PSOM, a child with mild functional impairment in four domains can receive the same score as a child with severe or profound impairment in one domain, which can lead to different levels of functional impairment within the same score. A variation of the PSOM, the Severity Classification Scheme (PSOM-SCS) has recently been developed to capture overall functional impact across domains better[12] and has been used in pediatric stroke outcome studies as well.[7]

A number of the global outcomes and motor scales have been adapted from adult scales such as the Glasgow Outcome Scale (GOS) and the Assisting Hand Assessment (AHA), which provides the advantage of supporting comparison with adult populations as well as facilitating the evaluation of teens into adulthood.[28] However, definitions may be hard to interpret, given the developmental stage of a child (e.g., “age appropriately independent for daily living”). Also, young children with stroke may have deficits that become more apparent over time as language and motor skills become more complex.[3]

Many of the motor performance and cognition measures evaluated in the current study are designed to assess different subdomains of function. Broad batteries of motor function or cognitive measures that screen many subdomains of function such as the Gross Motor Function Measure or the Wechsler Intelligence Scales are valuable and received high ratings from the experts as they are well validated and standardized. However, these measures are often impractical and time-consuming to administer within a routine neurology clinic visit or a clinical trial visit. A global performance measure may be useful to identify areas of impairment, whereas well-validated subdomain specific measures such as the California Verbal Learning Test, which is used to assess verbal memory, or the Community Balance and Mobility scale may be better suited to provide greater specificity and detail of specific impairments after areas of low performance are identified on a broad screening battery (e.g., WISC-V).

Often deficits in social-emotional functioning and behavior have a greater impact on health and well-being of children after stroke than deficits in physical functioning and mobility.[112] Therefore, an index of well-being such as a quality-of-life measure or a behavioral measure to track patient outcomes over time is recommended. Through the current evaluation of commonly used measures, it was noted that several of the adaptive functioning and behavior measures were highly correlated with one another (e.g., the BASC and CBCL) and are described to evaluate similar constructs.[97] During selection of outcome measures in the domains of behavior, QoL, and adaptive function for clinical practice or a clinical trial, care must be taken to avoid administering multiple highly correlated measures to reduce redundancy as well as patient burden.

There are some notable limitations to this study. Although a comprehensive evaluation of many of the outcome measures is provided, not every measure available to evaluate outcomes in pediatric stroke has been captured. The list of measures included in the current study was informed through a thorough literature search and modified by experts; however, the measures chosen were those most frequently used within the North American context. Measures used that are in accordance with the European International Classification of Disease system were outside the scope of the current study. Nevertheless, many of the measures selected for use are validated cross-culturally and are available in numerous languages. Our assessment suggests the need for new practical outcome measures to assess specific outcome subdomains and the necessity of further validation of commonly used measures within a pediatric stroke population.

Although making definitive outcome measure recommendations for all situations and studies remains difficult, pediatric stroke centers should include global measures at each follow-up time point as well as more domain-specific measures as appropriate. The rankings established in this article provide a reference for selecting outcome measures depending on the clinical or research question, assessment capabilities, and age of the child. Typical timing of assessments varies (also see Felling et al., (2023) which provides a roadmap for the timing of pediatric stroke outcome assessment). However, at hospital discharge, global assessments such as the PSOM are appropriate. More detailed assessments may occur in inpatient rehabilitation, three months poststroke, 12 months poststroke, and as clinical needs dictate, especially at critical time points of transition. Some measures have practice effects such that they cannot be used more often than every 12 months.

This work is intended to provide a toolkit for clinicians and clinical researchers to tailor outcome measure choices for children with stroke in clinical care, observational research studies, or clinical trials. The compendium of assessments and evaluation of their quality and utility should support more consistency across centers, which should facilitate research and care pathways. In the future, the development of additional pediatric stroke-specific outcome measures or validation of existing measures in pediatric stroke populations would be helpful.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Funding:

S.J.F's contribution to this work was generously supported by the Canadian Institute for Health Research (CIHR), Funding Reference Number: 181532. NIH grant 1K24-HL147017 focused on mentoring in patient-oriented research funded Dr. Jordan's time.

References

1. deVeber GA, Kirton A, Booth FA, et al. Epidemiology and outcomes of arterial ischemic stroke in children: the Canadian Pediatric Ischemic Stroke Registry. Pediatr Neurol. 2017;69:58–70. [PubMed: 28254555]
2. Porcari GS, Jordan LC, Ichord RN, Licht DJ, Smith SE, Beslow LA. Outcome trajectories after primary perinatal hemorrhagic stroke. Pediatr Neurol. 2020;105:41–47. [PubMed: 31952959]
3. Lo W, Gordon AL, Hajek C, et al. Pediatric stroke outcome measure: predictor of multiple impairments in childhood stroke. J Child Neurol. 2014;29:1524–1530. [PubMed: 24163399]
4. Yvon E, Lamotte D, Tiberghien A, et al. Long-term motor, functional, and academic outcome following childhood ischemic and hemorrhagic stroke: a large rehabilitation center-based retrospective study. Dev Neurorehabil. 2018;21:83–90. [PubMed: 27841719]
5. Felling RJ, Sun LR, Maxwell EC, Goldenberg N, Bernard T. Pediatric arterial ischemic stroke: epidemiology, risk factors, and management. Blood Cells Mol Dis. 2017;67:23–33. [PubMed: 28336156]
6. Malone LA, Felling RJ. Pediatric stroke: unique implications of the immature brain on injury and recovery. Pediatr Neurol. 2020;102:3–9. [PubMed: 31371122]
7. Felling RJ, Rafay MF, Bernard TJ, et al. Predicting recovery and outcome after pediatric stroke: results from the international pediatric stroke study. Ann Neurol. 2020;87:840–852. [PubMed: 32215969]
8. Engelmann KA, Jordan LC. Outcome measures used in pediatric stroke studies: a systematic review. Arch Neurol. 2012;69:23–27. [PubMed: 22232344]
9. Mokkink LB, Terwee CB, Patrick DL, et al. The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. J Clin Epidemiol. 2010;63:737–745. [PubMed: 20494804]
10. Mokkink LB, Terwee CB, Patrick DL, et al. The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study. Qual Life Res. 2010;19:539–549. [PubMed: 20169472]
11. Mokkink L, Terwee CB, Knol DL, et al. Protocol of the COSMIN study: COnsensus-based Standards for the selection of health Measurement INstruments. BMC Med Res Methodol. 2006;6:2. [PubMed: 16433905]
12. Slim M, Fox CK, Friefeld S, et al. Validation of the pediatric stroke outcome measure for classifying overall neurological deficit. Pediatr Res. 2020;88:234–242. [PubMed: 32179868]
13. Kitchen L, Westmacott R, Friefeld S, et al. The pediatric stroke outcome measure: a validation and reliability study. Stroke. 2012;43:1602–1608. [PubMed: 22474056]
14. DeVeber GA, MacGregor D, Curtis R, Mayank S. Neurologic outcome in survivors of childhood arterial ischemic stroke and sinovenous thrombosis. J Child Neurol. 2000;15:316–324. [PubMed: 10830198]
15. Chung MG, Lo W. Commentary on "validation of the pediatric stroke outcome measure for classifying overall neurological deficit". Pediatr Res. 2020;88:157–158. [PubMed: 32359224]
16. Lo WD, Ichord RN, Dowling MM, et al. The pediatric stroke recurrence and recovery questionnaire: validation in a prospective cohort. Neurology. 2012;79:864–870. [PubMed: 22895580]
17. Hay K, Nelin M, Carey H, et al. Hammersmith infant neurological examination asymmetry score distinguishes hemiplegic cerebral palsy from typical development. Pediatr Neurol. 2018;87:70–74. [PubMed: 30190180]

18. Uusitalo K, Haataja L, Nyman A, Lehtonen T, Setänen S. Hammersmith infant neurological examination and long-term cognitive outcome in children born very preterm. Dev Med Child Neurol. 2021;63:947–953. [PubMed: 33834473]
19. Ljungblad UW, Paulsen H, Tangeraas T, Evensen KAI. Reference material for hammersmith infant neurologic examination scores based on healthy, term infants age 3–7 months. J Pediatr. 2022;244:79–85. [PubMed: 35093317]
20. Romeo DM, Ricci D, Brogna C, Mercuri E. Use of the hammersmith infant neurological examination in infants with cerebral palsy: a critical review of the literature. Dev Med Child Neurol. 2016;58:240–245. [PubMed: 26306473]
21. Tedla JS, Bajaj A, Joshua AM, Kamath G. Psychometric properties of hammersmith infant neurological examination in 12 months old high-risk infants: a cross sectional study. Indian J Physiother Occup Ther. 2014;8:169.
22. Pollack MM, Holubkov R, Funai T, et al. Relationship between the functional status scale and the pediatric overall performance category and pediatric cerebral performance category scales. JAMA Pediatr. 2014;168:671–676. [PubMed: 24862461]
23. Fiser DH, Long N, Roberson PK, Hefley G, Zolten K, Brodie-Fowler M. Relationship of pediatric overall performance category and pediatric cerebral performance category scores at pediatric intensive care unit discharge with outcome measures collected at hospital discharge and 1- and 6-month follow-up assessments. Crit Care Med. 2000;28:2616–2620. [PubMed: 10921604]
24. Crouchman M, Rossiter L, Colaco T, Forsyth R. A practical outcome scale for paediatric head injury. Arch Dis Child. 2001;84:120–124. [PubMed: 11159284]
25. Geary M, Kirkham F, Drever E, Best K, Anwar DR, Palmer J. OP36 – 2640: King's Outcome Scale for Childhood Head Injury (KOSCHI) – prospective and retrospective comparison of outcome, and level of agreement, within the neuro-rehabilitation cohort at Southampton Children's Hospital. Eur J Paediatr Neurol. 2015;19:S12.
26. Calvert S, Miller HE, Curran A, et al. The King's Outcome Scale for Childhood Head Injury and injury severity and outcome measures in children with traumatic brain injury. Dev Med Child Neurol. 2008;50:426–431. [PubMed: 18422680]
27. Greenspoon D, Rumney P, Hung R, et al. Feasibility study of the King's outcome scale for childhood head injury in children attending a rehabilitation hospital. Arch Phys Med Rehabil. 2014;95:e56.
28. Beers SR, Wisniewski SR, Garcia-Filion P, et al. Validity of a pediatric version of the Glasgow Outcome Scale-Extended. J Neurotrauma. 2012;29:1126–1139. [PubMed: 22220819]
29. Hudak AM, Caesar RR, Frol AB, et al. Functional outcome scales in traumatic brain injury: a comparison of the Glasgow Outcome Scale (Extended) and the functional status examination. J Neurotrauma. 2005;22:1319–1326. [PubMed: 16305320]
30. Banks JL, Marotta CA. Outcomes validity and reliability of the modified Rankin scale: implications for stroke clinical trials. Stroke. 2007;38:1091–1096. [PubMed: 17272767]
31. Maddux AB, Cox-Martin M, Dichiaro M, Bennett TD. The association between the functional status scale and the pediatric functional independence measure in children who survive traumatic brain injury. Pediatr Crit Care Med. 2018;19:1046–1053. [PubMed: 30119094]
32. Morris C, Bartlett D. Gross motor function classification system: impact and utility. Dev Med Child Neurol. 2004;46. [PubMed: 14974647]
33. Russell DJ, Avery LM, Rosenbaum PL, Raina PS, Walter SD, Palisano RJ. Improved scaling of the gross motor function measure for children with cerebral palsy: evidence of reliability and validity. Phys Ther. 2000;80:873–885. [PubMed: 10960935]
34. Gray L, Ng H, Bartlett D. The gross motor function classification system. Pediatr Phys Ther. 2010;22:315–320. [PubMed: 20699783]
35. Towns M, Rosenbaum P, Palisano R, Wright FV. Should the Gross Motor Function Classification System be used for children who do not have cerebral palsy? Dev Med Child Neurol. 2018;60:147–154. [PubMed: 29105760]
36. Palisano RJ, Rosenbaum P, Bartlett D, Livingston MH. Content validity of the expanded and revised gross motor function classification system. Dev Med Child Neurol. 2008;50:744–750. [PubMed: 18834387]

37. Piscitelli D, Ferrarello F, Ugolini A, Verola S, Pellicciari L. Measurement properties of the gross motor function classification system, gross motor function classification system-expanded & revised, manual ability classification system, and communication function classification system in cerebral palsy: a systematic review with meta-analysis. Dev Med Child Neurol. 2021;63:1251–1261. [PubMed: 34028793]
38. Gunel MK, Mutlu A, Tarsuslu T, Livanelioglu A. Relationship among the Manual Ability Classification System (MACS), the Gross Motor Function Classification System (GMFCS), and the functional status (WeeFIM) in children with spastic cerebral palsy. Eur J Pediatr. 2009;168:477–485. [PubMed: 18551314]
39. Vohr BR, Msall ME, Wilson D, Wright LL, McDonald S, Poole WK. Spectrum of gross motor function in extremely low birth weight children with cerebral palsy at 18 months of age. Pediatrics. 2005;116:123–129. [PubMed: 15995042]
40. Harvey EM, Leonard-Green TK, Mohan KM, et al. Interrater and test-retest reliability of the beery visual-motor integration in schoolchildren. Optom Vis Sci. 2017;94:598–605. [PubMed: 28422801]
41. McCrimmon AW, Altomare AA, Matchullis RL, Jitlina K. Test review: the beery developmental test of visual-motor integration. J Psychoeduc Assess. 2012;30:588–592.
42. van Hartingsveldt MJ, Cup EH, Hendriks JC, de Vries L, de Groot IJ, Nijhuis-van der Sanden MW. Predictive validity of kindergarten assessments on handwriting readiness. Res Dev Disabil. 2015;36:114–124.
43. Holmefur MM, Krumlinde-Sundholm L. Psychometric properties of a revised version of the assisting hand assessment (Kids-AHA 5.0). Dev Med Child Neurol. 2016;58:618–624. [PubMed: 26507383]
44. Holmefur M, Krumlinde-Sundholm L, Eliasson A-C. Interrater and intrarater reliability of the assisting hand assessment. Am J Occup Ther. 2007;61:79–84. [PubMed: 17302108]
45. Louwers A, Krumlinde-Sundholm L, Boeschoten K, Beelen A. Reliability of the assisting hand assessment in adolescents. Dev Med Child Neurol. 2017;59:926–932. [PubMed: 28555755]
46. Krumlinde-Sundholm L, Holmefur M, Kottorp A, Eliasson AC. The assisting hand assessment: current evidence of validity, reliability, and responsiveness to change. Dev Med Child Neurol. 2007;49:259–264. [PubMed: 17376135]
47. Wallen M Reflections on the contribution of the assisting hand assessment. Dev Med Child Neurol. 2016;58:537–538. [PubMed: 26566733]
48. Holmefur M, Aarts P, Hoare B, Krumlinde-Sundholm L. Test-retest and alternate forms reliability of the assisting hand assessment. J Rehabil Med. 2009;41:886–891. [PubMed: 19841839]
49. Ek L, Eliasson AC, Sicola E, et al. Hand assessment for infants: normative reference values. Dev Med Child Neurol. 2019;61:1087–1092. [PubMed: 30719697]
50. Krumlinde-Sundholm L, Ek L, Sicola E, et al. Development of the hand assessment for infants: evidence of internal scale validity. Dev Med Child Neurol. 2017;59:1276–1283. [PubMed: 28984352]
51. Ryll UC, Krumlinde-Sundholm L, Verhage CH, et al. Predictive validity of the hand assessment for infants in infants at risk of unilateral cerebral palsy. Dev Med Child Neurol. 2021;63:436–443. [PubMed: 33251586]
52. Deitz JC, Kartin D, Kopp K. Physical & occupational therapy in pediatrics review of the Bruininks-Oseretsky test of motor proficiency, second edition (BOT-2). Phys Occup Ther Pediatr. 2007;27:87–102. [PubMed: 18032151]
53. Wuang YP, Su CY. Reliability and responsiveness of the Bruininks-Oseretsky test of motor proficiency-second edition in children with intellectual disability. Res Dev Disabil. 2009;30:847–855. [PubMed: 19181480]
54. Inness EL, Howe JA, Niechwiej-Szwedo E, Jaglal SB, McIlroy WE, Verrier MC. Measuring balance and mobility after traumatic brain injury: validation of the community balance and mobility scale (CB&M). Physiother Can. 2011;63:199–208. [PubMed: 22379260]
55. Wright F, Ryan J, Brewer K. Reliability of the Community Balance and Mobility Scale (CB&M) in high-functioning school-aged children and adolescents who have an acquired brain injury. Brain Inj. 2010;24:1585–1594. [PubMed: 20973626]

56. Knorr S, Brouwer B, Garland SJ. Validity of the community balance and mobility scale in community-dwelling persons after stroke. Arch Phys Med Rehabil. 2010;91:890–896. [PubMed: 20510980]
57. Wright MJ, Bos C. Performance of children on the community balance and mobility scale. Phys Occup Ther Pediatr. 2012;32:416–429. [PubMed: 22871209]
58. Franjoine MR, Gunther JS, Taylor MJ. Pediatric balance scale: a modified version of the berg balance scale for the school-age child with mild to moderate motor impairment. Pediatr Phys Ther. 2003;15:114–128. [PubMed: 17057441]
59. Gladstone DJ, Danells CJ, Black SE. The Fugl-Meyer assessment of motor recovery after stroke: a critical review of its measurement properties. Neurorehabil Neural Repair. 2002;16:232–240. [PubMed: 12234086]
60. Lee YY, Hsieh YW, Wu CY, Lin KC, Chen CK. Proximal Fugl-Meyer assessment scores predict clinically important upper limb Improvement after 3 stroke rehabilitative interventions. Arch Phys Med Rehabil. 2015;96:2137–2144. [PubMed: 26260019]
61. Fasoli S, Fragala-Pinkham M, Haley S. Fugl-meyer assessment: reliability for children with hemiplegia. Arch Phys Med Rehabil. 2009;90:e4–e5.
62. Rabadi MH, Rabadi FM. Comparison of the action research arm test and the Fugl-Meyer assessment as measures of upper-extremity motor weakness after stroke. Arch Phys Med Rehabil. 2006;87:962–966. [PubMed: 16813784]
63. Fugl-Meyer AR, Jääskö L, Leyman I, Olsson S, Steglind S. The post-stroke hemiplegic patient. A method for evaluation of physical performance. Scand J Rehabil Med. 1975;7:13–31. [PubMed: 1135616]
64. Geiger R, Strasak A, Treml B, et al. Six-Minute walk test in children and adolescents. J Pediatr. 2007;150:395–399.e2. [PubMed: 17382117]
65. Fitzgerald D, Hickey C, Delahunt E, Walsh M, O'Brien T. Six-minute walk test in children with spastic cerebral palsy and children developing typically. Pediatr Phys Ther. 2016;28:192–199. [PubMed: 26808959]
66. Wevers LE, Kwakkel G, van de Port IG. Is outdoor use of the six-minute walk test with a global positioning system in stroke patients' own neighbourhoods reproducible and valid? J Rehabil Med. 2011;43:1027–1031. [PubMed: 22031349]
67. McDonnell M Action research arm test. Aust J Physiother. 2008;54:220. [PubMed: 18833688]
68. van der Lee JH, Roorda LD, Beckerman H, Lankhorst GJ, Bouter LM. Improving the action research arm test: a unidimensional hierarchical scale. Clin Rehabil. 2002;16:646–653. [PubMed: 12392340]
69. Yozbatiran N, Der-Yeghiaian L, Cramer SC. A standardized approach to performing the action research arm test. Neurorehabil Neural Repair. 2008;22:78–90. [PubMed: 17704352]
70. van der Lee JH, De Groot V, Beckerman H, Wagenaar RC, Lankhorst GJ, Bouter LM. The intra- and interrater reliability of the action research arm test: a practical test of upper extremity function in patients with stroke. Arch Phys Med Rehabil. 2001;82:14–19. [PubMed: 11239280]
71. DeMatteo C, Law M, Russell D, Pollock N, Rosenbaum P, Walter S. The reliability and validity of the quality of upper extremity skills test. Phys Occup Ther Pediatr. 1993;13:1–18.
72. Thorley M, Lannin N, Cusick A, Novak I, Boyd R. Construct validity of the quality of upper extremity skills test for children with cerebral palsy. Dev Med Child Neurol. 2012;54:1037–1043. [PubMed: 22845645]
73. Thorley M, Lannin N, Cusick A, Novak I, Boyd R. Reliability of the quality of upper extremity skills test for children with cerebral palsy aged 2 to 12 years. Phys Occup Ther Pediatr. 2012;32:4–21. [PubMed: 21838618]
74. Sheppard JJ, Hochman R, Baer C. The dysphagia disorder survey: validation of an assessment for swallowing and feeding function in developmental disability. Res Dev Disabil. 2014;35:929–942. [PubMed: 24637033]
75. Calis EA, Veugelers R, Sheppard JJ, Tibboel D, Evenhuis HM, Penning C. Dysphagia in children with severe generalized cerebral palsy and intellectual disability. Dev Med Child Neurol. 2008;50:625–630. [PubMed: 18754902]

76. Dodrill P, Gosa MM. Pediatric dysphagia: physiology, assessment, and management. Ann Nutr Metab. 2015;66:24–31. [PubMed: 26226994]
77. Community-University Partnership for the Study of Children. Early childhood measurement and evaluation tool review adaptive behaviour assessment system system-second edition (ABAS-II). Available at: https://www.ualberta.ca/community-university-partnership/media-library/community-university-partnership/resources/tools—assessment/abas-ii-jan-2012.pdf; 2011. Accessed January 14, 2022.
78. Rust JO, Wallace MA. Test review: Adaptive Behaviour Assessment System – Second Edition (ABAS-II). J of Psych Assess. 2003;22:367–373.
79. Harrison P, Oakland T. Adaptive Behavior Assessment System. 3rd ed. Torrance: WPS; 2015. Manual.
80. Bedell G Further validation of the child and adolescent scale of participation (CASP). Dev Neurorehabil. 2009;12:342–351. [PubMed: 20477563]
81. Golos A, Bedell G. Responsiveness and discriminant validity of the Child and Adolescent Scale of Participation across three years for children and youth with traumatic brain injury. Dev Neurorehabil. 2018;21:431–438. [PubMed: 28692352]
82. Lambregts SAM, Smetsers JEM, Verhoeven IMAJ, et al. Cognitive function and participation in children and youth with mild traumatic brain injury two years after injury. Brain Inj. 2018;32:230–241. [PubMed: 29190153]
83. Golos A, Bedell G. Psychometric properties of the child and adolescent scale of participation (CASP) across a 3-year period for children and youth with traumatic brain injury. NeuroRehabilitation. 2016;38:311–319. [PubMed: 27061159]
84. De Bock F, Bosle C, Graef C, Oepen J, Philippi H, Urschitz MS. Measuring social participation in children with chronic health conditions: validation and reference values of the child and adolescent scale of participation (CASP) in the German context. BMC Pediatr. 2019;19:125. [PubMed: 31018847]
85. Ottenbacher KJ, Msall ME, Lyon N, et al. The WeeFIM instrument: its utility in detecting change in children with developmental disabilities. Arch Phys Med Rehabil. 2000;81:1317–1326. [PubMed: 11030496]
86. Wong V, Chung B, Hui S, et al. Cerebral palsy: correlation of risk factors and functional performance using the functional independence measure for children (WeeFIM). J Child Neurol. 2004;19:887–893. [PubMed: 15658794]
87. Msall ME, DiGaudio K, Rogers BT, et al. The functional independence measure for children (WeeFIM). Clin Pediatr (Phila). 1994;33:421–430. [PubMed: 7525140]
88. Grilli L, Feldman DE, Majnemer A, Couture M, Azoulay L, Swaine B. Associations between a functional independence measure (WeeFIM) and the pediatric quality of life inventory (PedsQL4.0) in young children with physical disabilities. Qual Life Res. 2006;15:1023–1031. [PubMed: 16900282]
89. Singh A, Yeh CJ, Blanchard SB. Ages and stages questionnaire: a global screening scale. Bol Med Hosp Infant Mex. 2017;74:5–12. [PubMed: 29364814]
90. Charkaluk ML, Rousseau J, Calderon J, et al. Ages and stages questionnaire at 3 years for predicting IQ at 5–6 years. Pediatrics. 2017;139:e20162798. [PubMed: 28360034]
91. Rothstein A, Miskovic A, Nitsch K. Brief review of psychometric properties and clinical utility of the ages and stages questionnaires, third edition for evaluating pediatric development. Arch Phys Med Rehabil. 2017;98:809–810.
92. Woodward BJ, Papile LA, Lowe JR, et al. Use of the ages and stages questionnaire and bayley scales of infant development-II in neurodevelopmental follow-up of extremely low birth weight infants. J Perinatol. 2011;31: 641–646. [PubMed: 21311498]
93. Reynolds CR, Kamphaus RW. The Clinician's Guide to the Behavior Assessment System for Children (BASC). New York: Guilford Press; 2002.
94. Kamphaus RW, Reynolds C, Hatcher N, Kim S. Treatment Planning and Evaluation with the Behavior Assessment System for Children (BASC). In: Maruish ME, ed. The use of psychological testing for treatment planning and outcomes assessment: Instruments for children and adolescents. New Jersey: Lawrence Erlbaum Associates Publishers; 2004:331–354.

95. Garcia-Barrera MA, Duggan EC, Karr JE, Reynolds CR. Examining executive functioning using the behavior assessment system for children (BASC). In: Handbook of executive functioning 283–299. New York: Springer; 2014.

96. Floyd RG, Kirby EA. Psychometric properties of measures of behavioral inhibition with preschool-age children: implications for assessment of children at risk for ADHD. J Atten Disord. 2001;5:79–91.

97. Gladman M, Lancaster S. A review of the behaviour assessment system for children. Sch Psych Int. 2003;24:276–291.

98. Achenbach TM. The Achenbach System of Empirically Based Assessment (ASEBA): Development, Findings, Theory, and Applications. Vermont: University of Vermont Research Center for Children, Youth, & Families; 2009.

99. Bordin IA, Rocha MM, Paula CS, et al. Child Behavior Checklist (CBCL): Youth Self-Report (YSR) and Teacher's Report Form (TRF): an overview of the development of the original and Brazilian versions. Cad Saúde Pública. 2013;29:13–28. [PubMed: 23370021]

100. de Groot A, Koot HM, Verhulst FC. Cross-cultural generalizability of the child behavior checklist cross-informant syndromes. Psychol Assess. 1994;6: 225–230.

101. Liu J, Leung P, Sun R, Li H-T, Liu J-M. Cross-cultural application of achenbach system of empirically based assessment: instrument translation in Chinese, challenges, and future directions. World J Clin Pediatr. 2012;8:5–10.

102. Wild D, Furtado T, Angalakuditi M. The translation and cultural adaptation of the Child Behavior Checklist for use in Israel (Hebrew), Korea, the US (Spanish), India (Malayalam and Kannada), and Spain. Psychol Res Behav Manag. 2012;5:51. [PubMed: 22715318]

103. Farmer C, Adedipe D, Bal V, Chlebowski C, Thurm A. Reliability of the Vineland Adaptive Behavior Scales, Third Edition. PsyArXiv Preprints. 2019. 10.31234/osf.io/pn463.

104. Pepperdine CR, McCrimmon AW. Test review: Vineland adaptive behavior scales, Third edition (Vineland-3) by Sparrow, S. S., Cicchetti, D. V., & Saulnier, C. A. Can J Sch Psychol. 2018;33:157–163.

105. Bard DE, Wolraich ML, Neas B, Doffing M, Beck L. The psychometric properties of the Vanderbilt attention-deficit hyperactivity disorder diagnostic parent rating scale in a community population. J Dev Behav Pediatr. 2013;34:72–82. [PubMed: 23363972]

106. Becker SP, Langberg JM, Vaughn AJ, Epstein JN. Clinical utility of the Vanderbilt ADHD diagnostic parent rating scale comorbidity screening scales. J Dev Behav Pediatr. 2012;33:221–228. [PubMed: 22343479]

107. Collett BR, Ohan JL, Myers KM. Ten-year review of rating scales. V: scales assessing attention-deficit/hyperactivity disorder. J Am Acad Child Adolesc Psychiatry. 2003;42:1015–1037. [PubMed: 12960702]

108. Pelham WE Jr, Fabiano GA, Massetti GM. Evidence-based assessment of attention deficit hyperactivity disorder in children and adolescents. J Clin Child Adolesc Psychol. 2005;34:449–476. [PubMed: 16026214]

109. Kelsay K, Dardar S Screening for Pediatric ADHD in the Primary Care Clinic. In: Maruish ME editor. Handbook of Pediatric Psychological Screening and Assessment in Primary Care. Oxford Routledge. p. 381–394.

110. Wolraich ML, Bard DE, Neas B, Doffing M, Beck L. The psychometric properties of the Vanderbilt attention-deficit hyperactivity disorder diagnostic teacher rating scale in a community population. J Dev Behav Pediatr. 2013;34:83–93. [PubMed: 23363973]

111. Fiume A, Deveber G, Jang SH, Fuller C, Viner S, Friefeld S. Development and validation of the pediatric stroke quality of life measure. Dev Med Child Neurol. 2018;60:587–595. [PubMed: 29451699]

112. Rohner A, Gutbrod K, Kohler B, et al. Health-related quality of life in young adults following pediatric arterial ischemic stroke. Stroke. 2020;51:952–957. [PubMed: 31865895]

113. Lai JS, Nowinski C, Victorson D, et al. Quality-of-life measures in children with neurological conditions: pediatric neuro-QOL. Neurorehabil Neural Repair. 2012;26:36–47. [PubMed: 21788436]

114. Lai JS, Nowinski CJ, Zelko F, et al. Validation of the Neuro-QoL measurement system in children with epilepsy. Epilepsy Behav. 2015;46:209–214. [PubMed: 25862469]
115. Cella D, Lai JS, Nowinski CJ, et al. Neuro-QOL: brief measures of health-related quality of life for clinical research in neurology. Neurology. 2012;78:1860–1867. [PubMed: 22573626]
116. Bertisch H, Rivara FP, Kisala PA, et al. Psychometric evaluation of the pediatric and parent-proxy patient-reported outcomes measurement information system and the neurology and traumatic brain injury quality of life measurement item banks in pediatric traumatic brain injury. Qual Life Res. 2017;26:1887–1899. [PubMed: 28271316]
117. McCarthy ML, MacKenzie EJ, Durbin DR, et al. The pediatric quality of life inventory: an evaluation of its reliability and validity for children with traumatic brain injury. Arch Phys Med Rehabil. 2005;86:1901–1909. [PubMed: 16213229]
118. Desai AD, Zhou C, Stanford S, Haaland W, Varni JW, Mangione-Smith RM. Validity and responsiveness of the pediatric quality of life inventory (PedsQL) 4.0 generic core scales in the pediatric inpatient setting. JAMA Pediatr. 2014;168:1114–1121. [PubMed: 25347549]
119. Chorpita BF, Moffitt CE, Gray J. Psychometric properties of the revised child anxiety and depression scale in a clinical sample. Behav Res Ther. 2005;43:309–322. [PubMed: 15680928]
120. Chorpita BF, Yim L, Mofatt C, Umemoto LA, Francis SE. Assessment of symptoms of DSM-IV anxiety and depression in children: a revised child anxiety and depression scale. Behav Res Ther. 2000;38:835–855. [PubMed: 10937431]
121. Ebesutani C, Korathu-Larson P, Nakamura BJ, Higa-McMillan C, Chorpita B. The revised child anxiety and depression scale 25eparent version: scale development and validation in a school-based and clinical sample. Assessment. 2017;24:712–728. [PubMed: 26834091]
122. Bae Y Test review: children's depression inventory 2 (CDI 2). J Psychoeduc Assess. 2012;30:304–308.
123. Figueras Masip A, Amador-Campos JA, Gómez-Benito J, del Barrio Gándara V. Psychometric properties of the children's depression inventory in community and clinical samples. Span J Psychol. 2010;13:990–999. [PubMed: 20977046]
124. Stumper A, Olino TM, Abramson LY, Alloy LB. A factor analysis and test of longitudinal measurement invariance of the children's depression inventory (CDI) across adolescence. J Psychopathol Behav Assess. 2019;41:692–698. [PubMed: 33132495]
125. de la Vega R, Racine M, Sánchez-Rodríguez E, et al. Psychometric properties of the short form of the children's depression inventory (CDI-S) in young people with physical disabilities. J Psychosom Res. 2016;90:57–61. [PubMed: 27772560]
126. Block GW, Nanson JL, Lowry NJ. Attention, memory, and language after pediatric ischemic stroke. Child Neuropsychol. 1999;5:81–91.
127. Fine EM, Delis DC. California Verbal Learning Test – Children's Version in Encyclopedia of Clinical Neuropsychology. New York: Springer; 2011:476–479.
128. Yeates KO, Blumenstein E, Patterson CM, Delis DC. Verbal learning and memory following pediatric closed-head injury. J Int Neuropsychol Soc. 1995;1:78–87. [PubMed: 9375212]
129. O'Jile JR, Schrimsher GW, O'Bryant SE. The California verbal learning test-children's version: relation to factor indices of the Wechsler intelligence scale for children-third edition. J Clin Exp Neuropsychol. 2005;27:815–822. [PubMed: 16183615]
130. Woods S, Delis D, Scott J, Kramer J, Holdnack J. The California verbal learning test – second edition: test-retest reliability, practice effects, and reliable change indices for the standard and alternate forms. Arch Clin Neuropsychol. 2006;21:413–420. [PubMed: 16843636]
131. DeJong J, Donders J. A confirmatory factor analysis of the California Verbal Learning Test–Second Edition (CVLT-II) in a traumatic brain injury sample. Assessment. 2009;16:328–336. [PubMed: 19546480]
132. Mottram L, Donders J. Cluster subtypes on the California verbal learning testechildren's version after pediatric traumatic brain injury. Dev Neuropsychol. 2006;30:865–883. [PubMed: 17083297]
133. Donders J, Minnema MT. Performance discrepancies on the California Verbal Learning TesteChildren's Version (CVLTeC) in children with traumatic brain injury. J Int Neuropsychol Soc. 2004;10:482–488. [PubMed: 15327727]

134. Goodman AM, Delis DC, Mattson SN. Normative data for 4-year-old children on the California verbal learning test-children's version. Clin Neuropsychol. 1999;13:274–282. [PubMed: 10726599]
135. Wiegner S, Donders J. Performance on the California verbal learning test after traumatic brain injury. J Clin Exp Neuropsychol. 1999;21:159–170. [PubMed: 10425514]
136. Dockrell JE, Marshall CR. Measurement issues: assessing language skills in young children. Child Adolesc Ment Health. 2015;20:116–125. [PubMed: 32680388]
137. Lansing AE, Max JE, Delis DC, et al. Verbal learning and memory after childhood stroke. J Int Neuropsychol Soc. 2004;10:742–752. [PubMed: 15327721]
138. Vaughan-Jensen J, Adame C, McLean L, Gámez B. Test review: D. Wechsler individual achievement test. J Psychoeduc Assess. 2011;29:286–291.
139. McCrimmon AW, Climie EA. Test review: D. Wechsler individual achievement testdThird edition. San Antonio, TX: NCS Pearson, 2009. Can J Sch Psychol. 2011;26:148–156.
140. Burns TG. Wechsler individual achievement test-III: what is the 'gold standard' for measuring academic achievement? Appl Neuropsychol. 2010;17:234–236. [PubMed: 20799115]
141. Caemmerer JM, Maddocks DLS, Keith TZ, Reynolds MR. Effects of cognitive abilities on child and youth academic achievement: evidence from the WISCV and WIAT-III. Intelligence. 2018;68:6–20.
142. Breaux K, Lichtenberger E. Essentials of KTEA-3 and WIAT-III Assessment. New York: John Wiley & Sons; 2016.
143. Donders J, Wildeboer MA. Validity of the WCST-64 after traumatic brain injury in children. Clin Neuropsychol. 2004;18:521–527. [PubMed: 15841955]
144. Donders J, DenBraber D, Vos L. Construct and criterion validity of the Behaviour Rating Inventory of Executive Function (BRIEF) in children referred for neuropsychological assessment after paediatric traumatic brain injury. J Neuropsychol. 2010;4:197–209. [PubMed: 19930791]
145. Cederfeldt M, Widell Y, Andersson EE, Dahlin-Ivanoff S, Gosman-Hedström G. Concurrent validity of the executive function performance test in people with mild stroke. Br J Occup Ther. 2011;74:443–449.
146. Castellanos I, Kronenberger WG, Pisoni DB. Questionnaire-based assessment of executive functioning: psychometrics. Appl Neuropsychol Child. 2018;7:93–109. [PubMed: 27841670]
147. Anderson V, Le Brocque R, Iselin G, et al. Adaptive ability, behavior and quality of life pre and posttraumatic brain injury in childhood. Disabil Rehabil. 2012;34:1639–1647. [PubMed: 22416951]
148. Anderson V, Godfrey C, Rosenfeld Jv, Catroppa C. Predictors of cognitive function and recovery 10 years after traumatic brain injury in young children. Pediatrics. 2012;129:e254–e261. [PubMed: 22271691]
149. Anderson V, Anderson P, Jacobs R, Spencer Smith M. Development and assessment of executive function: From preschool to adolescence. In: Anderson V, Jacobs R, Anderson PJ, eds. Executive Functions and the Frontal Lobes. Oxford: Psychology Press (Imprint) for Routledge; 2008:123–154.
150. Anderson VA, Anderson P, Northam E, Jacobs R, Mikiewicz O. Relationships between cognitive and behavioral measures of executive function in children with brain disease. Child Neuropsychol. 2002;8:231–240. [PubMed: 12759820]
151. Anderson P. Assessment and development of executive function (EF) during childhood. Child Neuropsychol. 2002;8:71–82. [PubMed: 12638061]
152. Aarnoudse-Moens CSH, Smidts DP, Oosterlaan J, Duivenvoorden HJ, Weisglas-Kuperus N. Executive function in very preterm children at early school age. J Abnorm Child Psychol. 2009;37:981–993. [PubMed: 19488851]
153. di Lorenzo M, Desrocher M, Westmacott R. The clinical utility of the behavior rating inventory of executive function in preschool children with a history of perinatal stroke. Appl Neuropsychol Child. 2022;11:429–437. [PubMed: 33535801]
154. Wright AJ. Equivalence of remote, digital administration and traditional, in-person administration of the Wechsler Intelligence Scale for Children, fifth edition (WISC-V). Psychol Assess. 2020;32:809–817. [PubMed: 32718161]

155. Na SD, Burns TG. Wechsler intelligence scale for children-V: test review. Appl Neuropsychol Child. 2016;5:156–160. [PubMed: 25923224]
156. McGill RJ, Ward TJ, Canivez GL. Use of translated and adapted versions of the WISC-V: Caveat emptor. Sch Psychol Int. 2020;41:276–294.
157. Greathouse D, Shaughnessy MF. Test review: an interview with Amy Gabel: about the WISC-V. J Psychoeduc Assess. 2016;34:800–810.
158. Kaufman A, Raiford S, Coalson D. Intelligent Testing with the WISC-V. New York: John Wiley & Sons; 2016.
159. Hartman DE. Wechsler adult intelligence scale IV (WAIS IV): return of the gold standard. Appl Neuropsychol. 2009;16:85–87. [PubMed: 19205953]
160. Sattler JM. Assessment of Children: Cognitive Foundations and Application. 6th ed. San Diego: Sattler; 2020.
161. Zhu J, Weiss L. The Wechsler scales. In: Contemporary Intellectual Assessment: Theories, Tests, and Issues. The Guilford Press; 2005:297–324.
162. Watkins MW, Canivez GL. Assessing the Psychometric Utility of IQ Scores: a tutorial using the wechsler intelligence scale for childrenefifth edition. School Psych Rev. 2021;51:1–15.
163. Conners K. Conners - Technical Manual. 3rd ed. Bloomington: NCS Pearson; 2008.
164. Conners CK, Sitarenios G, Parker JD, Epstein JN. The revised Conners' Parent Rating Scale (CPRS-R): factor structure, reliability, and criterion validity. J Abnorm Child Psychol. 1998;26:257–268. [PubMed: 9700518]
165. Weintraub S, Bauer PJ, Zelazo PD, et al. I. NIH Toolbox Cognition Battery (CB): introduction and pediatric data. Monogr Soc Res in Child Dev. 2013;78:1–15.
166. Shields RH, Kaat AJ, McKenzie FJ, et al. Validation of the NIH toolbox cognitive battery in intellectual disability. Neurology. 2020;94:e1229–e1240. [PubMed: 32094241]
167. Carlozzi NE, Tulsky DS, Wolf TJ, et al. Construct validity of the NIH toolbox cognition battery in individuals with stroke. Rehabil Psychol. 2017;62:443–454. [PubMed: 29265865]
168. Tulsky DS, Heinemann AW. The clinical utility and construct validity of the NIH toolbox cognition battery (NIHTB-CB) in individuals with disabilities. Rehabil Psychol. 2017;62:409–412. [PubMed: 29265861]
169. Rebchuk AD, Deptuck HM, Kuzmuk LE, Silverberg ND, Field TS. Abstract WP559: the NIH toolbox cognition battery outperforms the MoCA in detecting cognitive impairment following mild stroke in young patients. Stroke. 2019;50:559.
170. Rebchuck AD, Alimohammadi A, Yuan M, et al. Assessment of prorated scoring of an abbreviated protocol for the National Institutes of Health toolbox cognition battery. J Int Neuropsychol Soc. 2020;26:1045–1050. [PubMed: 33081872]
171. Pereira A, Lopes S, Magalhaes P, Sampaio A, Chaleta E, Rosário P. How executive functions are evaluated in children and adolescents with cerebral palsy? A systematic review. Front Psychol. 2018;9:2. [PubMed: 29403414]
172. Baron IS. Delis-kaplan executive function system. Child Neuropsychol. 2004;10:147–152.
173. Swanson J The Delis-Kaplan executive function system. Can J Sch Psychol. 2005;20:117–128.
174. Berg C, Edwards DF, King A. Executive function performance on the children's kitchen task assessment with children with sickle cell disease and matched controls. Child Neuropsychol. 2012;18:432–448. [PubMed: 21961955]
175. Shunk AW, Davis AS, Dean RS. Test review: Dean C. Delis, Edith Kaplan & Kramer JH, Delis Kaplan executive function system (D-KEFS). Appl Neuropsychol. 2006;13:275–327.
176. Homack S, Lee D, Riccio CA. Test review: Delis-Kaplan executive function system. J Clin Exp Neuropsychol. 2005;27:599–609. [PubMed: 16019636]
177. Krivitzky L, Bosenbark DD, Ichord R, Jastrzab L, Billinghurst L. Brief report: relationship between performance testing and parent report of attention and executive functioning profiles in children following perinatal arterial ischemic stroke. Child Neuropsychol. 2019;25:1116–1124. [PubMed: 30909791]

178. Araujo GC, Antonini TN, Anderson V, et al. Profiles of executive function across children with distinct brain disorders: traumatic brain injury, stroke, and brain tumor. J Int Neuropsychol Soc. 2017;23:529–538. [PubMed: 28502261]

179. Betts J, Mckay J, Maruff P, Anderson V. The development of sustained attention in children: the effect of age and task load. Child Neuropsychol. 2006;12:205–221. [PubMed: 16837396]

180. Heaton SC, Reader SK, Preston AS, et al. The test of everyday attention for children (TEA-Ch): patterns of performance in children with ADHD and clinical controls. Child Neuropsychol. 2001;7:251–264. [PubMed: 16210214]

181. Henry LA, Bettenay C. The assessment of executive functioning in children. Child Adolesc Ment Health. 2010;15:110–119. [PubMed: 32847241]

182. Kent P The evolution of the Wechsler memory scale: a selective review. Appl Neuropsychol Adult. 2013;20:277–291. [PubMed: 23445503]

183. Carlozzi NE, Grech J, Tulsky DS. Memory functioning in individuals with traumatic brain injury: an examination of the Wechsler Memory ScaleeFourth Edition (WMSeIV). J Clin Exp Neuropsychol. 2013;35:906–914. [PubMed: 24033318]

184. Kent P The Wechsler Memory Scale: A Guide for Clinicians and Researchers. Oxford: Taylor & Francis; 2020.

185. Brooks BL, Iverson GL, Sherman EMS, Holdnack JA. Healthy children and adolescents obtain some low scores across a battery of memory tests. J Int Neuropsychol Soc. 2009;15:613–617. [PubMed: 19573280]

186. Virani S, Rasmussen C, Zivanovic N, et al. Learning and memory profiles in youth with perinatal stroke: a study of the Child and Adolescent Memory Profile (ChAMP). Child Neuropsychol. 2022;28:99–106. [PubMed: 34375160]

187. Piehl JJ, Wolff M, Hahm J. Test review of the child and adolescent memory profile (ChAMP). J Pediatr Neuropsychol. 2017;3:218–222.

188. Brooks BL, Holdnack JA, Iverson GL. Reliable change on memory tests is common in healthy children and adolescents. Arch Clin Neuropsychol. 2017;32:1001–1009. [PubMed: 28383636]

189. Sherman E, Brooks B. Child and Adolescent Memory Profile Professional Manual. Lutz: Psychological Assessment Resources; 2015.

190. Wilson K, Lesica S, Donders J. Clinical utility of the child and adolescent memory profile (ChAMP) after pediatric traumatic brain injury. Assessment. 2022;29:309–316. [PubMed: 33256457]

191. Washington K, Thomas-Stonell N, Oddson B, et al. Construct validity of the FOCUS© (Focus on the Outcomes of Communication under Six): a communicative participation outcome measure for preschool children. Child Care Health Dev. 2013;39:481–489. [PubMed: 23763249]

192. Thomas-Stonell N, Oddson B, Robertson B, Rosenbaum P. Validation of the focus on the outcomes of communication under six outcome measure. Dev Med Child Neurol. 2013;55:546–552. [PubMed: 23461266]

193. Coret MC, McCrimmon AW. Test review: Wiig EH, Semel E, Secord WA. (2013). Clinical evaluation of Language Fundamentals–Fifth edition (CELF-5). J Psychoeduc Assess. 2015;33:495–500.

194. Denman D, Speyer R, Munro N, Pearce WM, Chen YW, Cordier R. Psychometric properties of language assessments for children aged 4–12 years: a systematic review. Front Psychol. 2017;8:1515. [PubMed: 28936189]

195. Wiig H, Semel E, Secord W. Clinical Evaluation of Language Fundamentals: Technical Manual. 5th ed. Bloomington: NCS Pearson; 2013.

196. Peterson RK, McDonald KP, Vincent M, Williams TS, Dlamini N, Westmacott R. Characterizing language outcomes following childhood basal ganglia stroke. Appl Neuropsychol Child. 2021;10:14–25. [PubMed: 31006275]

197. Williams K Expressive Vocabulary Test. 3rd ed. Bloomington: NCS Pearson; 2018.

198. Dunn D Peabody Picture Vocabulary Test. 5th ed. Bloomington: NCS Pearson; 2019.

199. Whurr R, Evans S. Children's acquired aphasia screening test. Int J Lang Commun Disord. 1998;33:343–344. [PubMed: 10343717]

TABLE 1

Modified COSMIN Checklist for Scoring of Outcome Measures

Score	Meaning	Description
5	Highly Recommend	• Excellent reliability (test-retest, inter-rater, intrarater, and/or responsiveness) • Excellent validity (content, construct, and criterion validity) • Easy to interpret scores (can be used by individuals of differing expertise or no training required) • Evidence of cross-cultural validity • Excellent clinical utility across a majority of the following: short, does not require a lot of supplies, applicable for many intellectual/physical ability levels, can readminister within short period of time, wide age range of assessment allows for follow-up within patient tracking over time) • Data to support psychometric utility is strong (validated in a pediatric stroke population or relevant/related populations) • Sensitive to change over time (minimal clinically important difference) has been established, low likelihood of ceiling or floor effects)
4	Recommend	• Good reliability • Good validity • Easy to interpret scores • Minimal evidence of cross-cultural validity • Strong clinical utility (meets half of the criteria) • Data to support psychometric utility is strong • Sensitive to change over time
3	Recommend, nonpreferred	• Minimal evidence of reliability • Minimal evidence of validity • Some barriers to score interpretation • Minimal or reported evidence of cross-cultural validity • Meets some clinical utility criteria • Data to support psychometric utility is strong but has not been validated in a pediatric stroke or relevant/related population or is an adult measure that has been adapted for use in pediatric populations • Sensitive to change over time—no published data
2	Unable to recommend at this time	• Limited information with regard to psychometrics but has clinical utility (might be a measure that is often used clinically but has little research available)
1	Do not recommend	Poor psychometrics and poor clinical utility or measure is rarely used so there is little to no data on the measure

Abbreviation:

COSMIN = Consensus-based Standards for the Selection of Health Measurement Instruments

Modified COSMIN Guidelines were used for rating the individual measures. A scale of1to 5 was utilized; if majority of the criteria (>5 items) were met within a given score category, the measure would receive that score. The reliability and validity criteria both had to be satisfied to meet a given score category. Definitions of each scoring criteria were provided (refer to Supplemental Document 1).

TABLE 2

Global Performance Measures

Outcome Measure	Ratings (Means and Range)	Pros	Cons
PSOM[3,12–16]	Mean: 4.66 Range: 4–5	• Wide age range • Rapid administration • Validated in pediatric stroke • Phone version available	• Cognitive and behavioral items limited in scope • Not validated for administration by nonneurologists • Limited dynamic range • Outcome values often clumped in 0–3 range
psRRQ[16]	Mean: 4 Range: 4	• Quick to administer • Phone use avoids missing outcomes • Excellent for working with individuals in rural communities who are difficult to access • Reliable estimator of the PSOM	• Validity data limited • Some internal consistency data rates relatively low (ICC = 0.5) • Limited dynamic range
HINE[17–21]	Mean: 4 Range: 4	• Good for identifying motor impairment in infants- can have the caretaker complete the examination—good test-retest and interrater reliability • Predicts walking and independent sitting in children with cerebral palsy • Standardized on large cohorts of children	• Limited age range, which makes it difficult to track recovery over time (ends at 24 months)
PCPC/POPC[22–24]	Mean: 3.33 Range: 2–4	• Does not require training • Does not require the patient to be present if medical chart is available	• No research published in pediatric stroke has used this measure, suggesting it is less commonly used • Not validated in pediatric stroke • Limited dynamic range
King's Outcome Scale for Childhood Head Injury (KOSCHI)[25–27]	Mean: 3 Range: 3	• Wide age range- could be used in younger children if needed • Convergent validity with QoL measures • Similar to the psRRQ and phone version of the PSOM	• Validated in TBI, not pediatric stroke • Items measure different outcomes, grades 1 and 2 reflect physiological function, grades • 2 and 3 reflect awareness and response, and grades 3–5 measure functional independence • Scale places strong emphasis on concentration, behavior, and inhibition—common problems in pediatric stroke. Validity may be impacted in children with comorbid ADHD • Not a great predictor of behavior or emotions • Similar to the psRRQ and phone version of the PSOM • Limited dynamic range
GOS-Peds[28,29]	Mean: 2.33 Range: 2–3	• Good sensitivity to severity ofTBI in children, associated with changes in TBI sequelae over time • Validated for use in infants, toddlers, children, and adolescents	• Not validated in peds stroke • Limited dynamic range
mRS[30]	Mean: 2.33 Range: 2–3	• Well-validated adult disability scale	• Not validated for children • Limited dynamic range • Variable inter-rater reliability across mRS reliability studies
FSS[31]	Mean: 2.33 Range: 2–3	• Quick to use, minimal subjectivity • Applicable to broad age range • Commonly used in hospital environment	• Literature indicates it is not designed to predict outcome and should not be used to assess or predict outcomes for pediatric patients

Abbreviations:

ADHD = Attention-deficit/hyperactivity disorder

FSS = Functional Status Scale

GOS-Peds = Glasgow Outcome Scale-Peds

HINE = Hammersmith Infant Neurological Examination

ICC = Intraclass correlation coefficient

mRS = Modified Rankin Scale

PCPC = Pediatric Cerebral Performance Category

POPC = Pediatric Overall Performance Category

PSOM = Pediatric Stroke Outcome Measure

psRRQ = Pediatric Stroke Recurrence and Recovery Questionnaire

TBI = Traumatic brain injury

When only one number is given for range, all reviewers/raters agreed on that rating. Many are free but require training, refer to Supplemental Table 1 for specified cost.

TABLE 3

Motor Function Measures

Outcome Measure	Ratings (Means and Range)	Pros	Cons
Gross Motor Functional Measure (GMFM)[32–39]	Mean: 4.33 Range: 4–5	• Designed for children with cerebral palsy, also validated in pediatric TBI • Wide age range • Two versions—one that is sensitive to impairments in young children and those with more complex motor disability • Used as an outcome measure in intervention studies in cerebral palsy • Teaching resource available • Reference curves in cerebral palsy available (Hanna et al., 2008)	• Not validated in pediatric stroke • Some users of the GMFM selectively administer only some of the subscales, and the reliability and the validity of the subscale scores are generally not as strong as they are for total score
Beery-Buktenica[40–42]	Mean: 3.66 Range: 3–4	• Measure that can differentiate between motor coordination impairments, perception impairments, and integration impairments • Standardized, excellent validity and reliability	• Upper limb only
AHA[43–48]	Mean: 3.66 Range: 3–4	• Good at detecting small changes in functional ability • Tested in children with cerebral palsy	• Very long and cannot be completed in the clinic; more useful for research purposes or if there is a dedicated physical therapy or occupational therapy program that patients with stroke could incorporate into the program
HAI[49–51]	Mean: 3.66 Range: 3–4	• Requires substantial training to administer	• Only for use in infants (however, can use the AHA for older children)
BOT-2[52,53]	Mean: 3.33 Range: 3–4	• Used as a descriptive research tool • Addresses fine and gross motor function • Short and long forms available	• Cannot administer in clinic • Not used until age 4 years • Fine motor subtest not as strong as the others, needs to be performed by trained individual • BOT-2 test significantly overestimates score compared with the longer form in a healthy population of children
CB&M[54–58]	Mean: 3.33 Range: 3–4	• Validated in peds TBI (subset of participants with stroke), excellent reliability • Reliability estimates similar to adult stroke sample estimates • Can be rated in person or rated using video recording	• Long administration time and requires space to complete assessment • Ratings do not include quality of movement (i.e., patients with ataxia may meet the time requirement resulting in a high school, but the quality of their movement was low) • Arm function seems more important to assess as this tends to be more debilitating and arm function is not assessed • Administration is reliable *if* assessed by physical therapists
Fugl-Meyer Assessment[59–63]	Mean: 3.33 Range: 3–4	• Can calculate score for upper and lower limb function • Brief screening option available	• Difficult to use in the presence of aphasia or spatial neglect • Pain scale is variable • Only validated in adolescents and adults
Six-Minute Walk Test[64–66]	Mean: 3 Range: 3	• Simple test requiring no exercise equipment or advanced training • Reflects functional capacity for daily physical activities • Predictive of morbidity and mortality • Tested in child samples with cerebral palsy	• Highly specific, just related to walking—only looks at lower limb motor function and does not look at sensation, etc. • Does not provide specific information on pulmonary or cardiovascular systems during walking • Inter-rater reliability not great
Action Research Arm Test[67–70]	Mean: 3 Range: 3	Short administration time • Observation-based scoring, need some expertise in rating performance	Only assesses upper limb • Requires a lot of equipment • Not tested in younger children (ages 13+)

Outcome Measure	Ratings (Means and Range)	Pros	Cons
QUEST[71–73]	Mean: 3 Range: 3	• Designed specifically for cerebral palsy • Excellent reliability and validity in cerebral palsy • Assesses quality of movement unlike most measures that assess whether or not movement can be completed • Useful for assessing spasticity • Free	• Need to have experienced therapist administer the measure—does not have structured guidelines • Long administration time (3045 minutes) • QUEST is an assessment of quality; a change in score may not equate to a change in function/skill level • Age range is small • No validity data in acquired brain injury • No demonstration of predictive utility
PBS[58]	Mean: 3 Range: 3	• Good test-retest and inter-rater reliability • Short administration time • Validated in cerebral palsy Predictive of gross motor function	• Not yet validated in pediatric stroke • Only addresses balance skills so would have to do other tests to address other motor functions • Floor and ceiling effects • Correlates with age
Swallowing (oromotor)			
Dysphagia Disorder Survey[74–76]	Mean: 3 Range: 3	• Excellent for assessing swallowing and feeding behaviors • Well validated over broad age range • Relatively short • One of the few available measures for assessing physiologic limitations in swallowing and feeding sensory-motor function • Developed for persons with developmental disability so it is appropriate for use in individuals with varying intellectual capacities	• Highly specific, not really generalizable • Needs to be administered by trained speech and language pathologist, occupational therapist, or physical therapist No literature on utility in assessing improvement over time

Abbreviations:

AHA = Assisting Hand Assessment

BOT-2 = Bruininks-Oseretsky Test of Motor Proficiency, 2nd edltlon

CB&M = Community Balance & Mobility Scale

GMFM = Gross Motor Functional Measure

HAI = Hand Assessment for Infants

PBS = Pediatric Berg Balance Scale

QUEST = Quality of Upper Extremity Test

TBI = Traumatic brain injury

When only one number is given for range, all reviewers/raters agreed on that rating.

TABLE 4

Behavior, Adaptive Functioning, Quality of Life, and Mood

Outcome Measure	Ratings (Means and Range)	Pros	Cons
Behavior and adaptive functioning			
ABAS[77–79]	Mean: 4.66 Range: 4–5	• Excellent validity and reliability • Good at identifying deficits in clinical populations • Some validation in developmentally delayed children	• Need computer program to score it efficiently, easy to make errors when hand scoring • Expertise required • Only available in two languages (English and Spanish) • Not sensitive to TBI-related impair- ments/issues • Intended for the assessment of older children rather than younger children
CASP[80–84]	Mean: 4.33 Range: 4–5	• Affordable • Short administration time • Concurrent validity with other measures evaluated • Excellent reliability • Widely available • Validated in ABI	• Moderate face validity • Some reports of ceiling effects in ABI • Item ratings compare child with same- age peers, may be less responsive to change than measures that are not age- referenced • Does not differentiate between the extent of participation and extent to which child is able to participate
WeeFIM[85–88]	Mean: 4.33 Range: 4–5	• Strong psychometric properties in cerebral palsy and children with disabilities • Predicts longitudinal functional recovery in children with disability • Short administration time • Broad age range	• Developmental differences in children less than 3 years old may create a floor effect • Expensive • Requires licensing to score
ASQ[89–92]	Mean: 4 Range: 4	• Has good predictive validity for gross motor development • Ease of administration • Shorter administration time than the Bayley	• Only available for birth-6 years. Limited age range—unable to track children over time • Not as sensitive for infants with extremely low birth weight
BASC[93–97]	Mean: 4 Range: 4	• Broad age range • Teacher, parent, child versions • Strong psychometric properties (excellent validity and reliability)	• Longer administration time • Requires psychologist for interpretation
CBCL[98–102]	Mean: 4 Range: 4	• Broad age range • Teacher, parent, and self-report forms • One of the most commonly administered measures of behavioral functioning • Corresponds to DSM-V criteria • Extensively validated in children across a broad age span and cross culturally	• Not validated in pediatric stroke or cerebral palsy • Some of the items are not relevant to younger children
VABS-3rd edition[103,104]	Mean: 3.66 Range: 3–4	• Good validity • Excellent reliability	• Test-retest reliability is inconsistent, susceptible to responder bias • Longer administration time • Responder bias
Vanderbilt Assessment Scale (3rd edition)[105–110]	Mean: 3.66 Range: 3,5	• Parent and teacher form—useful for collecting the same information from multiple sources • Used in clinical setting • Screens for other disorders in addition to ADHD • Free	• Low inter-rater reliability • No evidence found for discriminant validity • Should be used as a screening tool only • Items are more relevant for school- aged children than younger children

Outcome Measure	Ratings (Means and Range)	Pros	Cons
		• Utility for screening for comorbid disorders • Teacher scale correlates highly with diagnosis of ADHD	
QoL			
PSQLM[111,112]	Mean: 4.66 Range: 4–5	• Only validated QoL measure in pediatric stroke • Developed specifically for pediatric stroke • Takes into account what is important to the family in calculation of QoL score	• Not widely used, newer measure
Neuro-Quality of Life Measure (Neuro QoL Pediatrics)[113–115]	Mean: 4 Range: 4	• Consists of both generic and targeted domains, which allows investigators to compare children's health status with other disease groups (e.g., TBI, cerebral palsy)	• Upper and lower extremity function were not validated in the Neuro QoL pediatrics version
PROMIS[116]	Mean: 4 Range: 4	• Validated in pediatric epilepsy and muscular dystrophy populations • Item banks developed for children and separate item banks for adults • Psychometrically sound	• Pediatric ABI and TBI not included in neuro QoL development, and it has not been validated in these populations
PedsQL/PedsQL Infant[117,118]	Mean: 4 Range: 3–5	• Not validated in pediatric stroke but commonly used	• Quality of life item measures assume certain level of functional ability
Mood			
RCADS[119–121]	Mean: 4 Range: 4	• Screens for anxiety and depression • Broad age range • Free • Available in 25 languages	• Some mixed findings on convergent validity
Child Depression Inventory[122–125]	Mean: 3.66 Range: 3–4	• Broad age range • Easy to score	• Depression only • Cost

Abbreviations:

ABAS = Adaptive Behavior Assessment System (3rd edition)

ABI = Acquired brain injury

ASQ = Ages and Stages Questionnaires

BASC = Behavioral Assessment System for Children

CASP = Child and Adolescent Scale of Participation

CBCL = Child Behavior Checklist

DSM-V = Diagnostic and Statistical Manual of Mental Disorders, fifth edition

PedsQL = Pediatric Quality of Life Inventory

PROMIS = Patient-Reported Outcomes Measurement System

PSLQM = Pediatric Stroke Quality of Life Measure

QoL = Quality of Life

RCADS = Revised Children's Anxiety and Depression Scale

TBI = Traumatic brain injury

VABS = Vineland Adaptive Behavior Scale

WeeFIM = Functional Independence Measure

When only one number is given for range, all reviewers/raters agreed on that rating.

TABLE 5

Cognition and Language

Outcome Measure	Ratings (Means and Range)	Pros	Cons
Cognition			
CVLT-Child / CVLT-33126–137	Mean: 4.66 Range: 4–5	• Short administration time • Well normed and studied in pediatric stroke and TBI • Child and adult versiondgood for tracking over time	• Cost but still less expensive than most other cognitive tests • Requires training to administer and interpret
WIAT-III[138–142]	Mean: 4.66 Range: 4–5	• Well standardized • Frequently used in research and clinically in pediatric stroke	• Long administration time
BRIEF-2[143–153]	Mean: 4.33 Range: 4–5	• Quick • Available in 40 languages • Solid psychometric properties • Easy to administer • Child and adult versiondgood for tracking over time	• Cost, parent-teacher inter-rater agreement was only moderate but was indicated to be consistent with expectation for different environmental settings
Wechsler Intelligence Tests (WISC-V/WPPSI-IV/WAIS-IV)[154–162]	Mean: 4.33 Range: 3–5	• Normed in >2000 children • Excellent psychometric properties • Most commonly used cognitive measure in North America • Can track patients from childhood to adulthood using same measure (different version depending on age) • Validated for use remotely (i.e., online administration)	• Cost • Requires significant training • Lengthy • Practice effects—frequent assessments are not feasible • Although available in several different languages, validation processes are of varying quality—suggested cautious use of Spanish version
Conner's Scale for ADHD Assessment, third edition[163,164]	Mean: 4.33 Range: 4–5	• Easy to • No information on sensitivity to change or cross-cultural validity	•No information on sensitivity to change or cross-cultural validity
NIH Toolbox Cognitive Battery[165–170]	Mean: 4 Range: 4	• Shorter testing than most cognitive screening batteries • Solid psychometric properties • Has been assessed across multiple conditions, including pediatric stroke • Child and adult version—good for tracking over time	• Cost (less than some) • Experience needed in psychologyand requires training to administer
DKEFS[171–178]	Mean: 3.66 Range: 3–4	• 9 subtests—can administer individually or as a group • Sensitive in the detection of frontal lobe function • Developed for populations with brain • Sensitive in distinguishing between different clinical groups (focal frontal lesions vs fetal alcohol syndrome vs healthy controls) • • Alternative forms—reduces practice	• Validity is lower in youngest age groups (greater variability in scores in younger groups during standardization) • Only valid for children ages 8+ • Adequate reliability for some subtests • Test instructions can be complex and repetitive effects
TEA-Ch[178–181]	Mean: 3.66 Range: 3–4	• Tests a number of different types of attention • Scores are age-sensitive • 3 versions available—risk of practice effects is lower • Child and adult version—good for tracking over time	• Clinicians tend to find the most recent version (second version) cumbersome • Children with ADHD have trouble undergoing assessment • Poor test-retest • Expensive • Scoring instructions are unclear
WMS-IV[182–185]	Mean: 3.33 Range: 3–4	• Good but longer administration time for same cost as similar tests • Decent reliability	• Not studied in peds stroke • Only for children 16+ • Different depending on age, which makes it concerning for assessing longitudinal effects

Outcome Measure	Ratings (Means and Range)	Pros	Cons
			• Practice effects • Low scores are common in healthy children and adolescents, interpret with some caution
ChAMP[186–190]	Mean: 3.33 Range: 3–4	• Screening index as well as longer battery available • Ecologically valid • Shorter than other memory tests • Does not rely heavily on other neurocognitive domains (e.g., visual-motor integration) • Developed and validated in children with motor impairments	• Ongoing expense for electronic scoring • Less sensitive to memory deficits than the CVLT • Screening index less reliable than subtest measures • Practice effects observed if tested again <45 days later • Tasks can be monotonous and boring for younger children for visual memory subtests—sustained attention is required Instructions subtest may be culturally relevant for some children leading to inflated scores • Low scores are common in healthy children and adolescents, interpret with some caution
Speech and language			
FOCUS[191–192]	Mean: 4 Range: 4	• Good validity and reliability • Parents can administer it • Low cost	• Relatively long delay between assessment points • Age range is small • Lengthy • Costly
EVT, third edition, and PPVT, fifth edition[197,198]	Mean: 3.66 Range: 3–4	• Reliable and valid. • Modifications available for children with disabilities and young children • Some cross-cultural validity • Can compare expressive to receptive language on two measures that were co-normed • Accessible for children with low intellectual ability	• Long administration time • High cost • Limited in its complexity as it is a naming task and does not evaluate higher-order aspects oflanguage (i.e., inference or comprehension of sentences/instructions)
Children's Acquired Aphasia Screening Test[199]	Mean: 1.33 Range: 1–2	• Developed to evaluate aphasia in children	• No information available within the literature

Abbreviations:

ADHD = Attention-deficit/hyperactivity disorder

BRIEF-2 = Behavior Rating Inventory of Executive Function, Second Edition

CELF = Clinical Evaluation of Language Fundamentals

ChAMP = Child and Adolescent Memory Profile

CVLT-3 = California Verbal Learning Test—Child, Third Edition

DKEFS = Delis Kaplan Executive Function System

EVT = Expressive Verbal Test

FOCUS = Focus on the Outcome of Communication Under Six

PPVT = Peabody Picture Vocabulary Test

TBI = Traumatic brain injury

TEA-Ch = Test of Everyday Attention-Child

WAIS-IV = Weschler Adult Intelligence Scale-IV

WIAT-III = Wechsler Intellectual Achievement Test, Third Edition

WISC-V = Weschler Intelligence Scale for Children-V

WMS-IV = Wechsler Memory Scale, Fourth Edition

WPPSI- IV = Welschler Preschool and Primary Scale of Intelligence-IV

When only one number is given for range, all reviewers/raters agreed on that rating.