Outcome Measures Used in Pediatric Stroke Studies

A Systematic Review

Kyle A. Engelmann, BA; Lori C. Jordan, MD, PhD

Because no gold-standard outcome measure or measures exist to allow comparison of pediatric stroke study outcomes in clinical trials, we designed a systematic review of the literature to survey the current use of pediatric stroke outcome measures. Studies that used at least 1 standardized measure to assess the outcome of children with ischemic or hemorrhagic stroke, from full-term newborn to age 18 years, were included. Although 34 studies were included, an additional 36 studies could not be included because ad hoc, author-generated outcome measures were used. Excluding those measures in neuropsychological batteries, 38 unique outcome measures were used. The Wechsler Intelligence Scales, Pediatric Stroke Outcome Measure, and Bayley Scales of Infant Development were among the most used, but 79% of outcome measures were used by no more than 2 studies. Although many measures used have been validated for use in children with other medical conditions or for adults with stroke, only 1 measure has been specifically validated for use in pediatric ischemic stroke. To maximize comparability of future clinical trial results, agreement regarding a preferred pediatric stroke outcome scale or battery of measures is paramount; these measures should be reliable, responsive to change, and specifically validated for use in children with stroke.

Although pediatric stroke occurs in about 2 or 3 per 100,000 children, treatment is still largely based on low levels of evidence. Three sets of pediatric stroke guidelines exist, but there are no clinical trials to inform treatment outside of sickle cell disease. More clinical trials aiming to improve pediatric stroke treatment are on the horizon, yet no gold-standard outcome measure is available to assess and compare the resulting outcomes.

Several institutions have recently highlighted the importance of validated, reliable outcome measures for patient-oriented research. The National Institutes of Health have begun investing in initiatives such as the Patient-Reported Outcomes Measurement Information System (PROMIS), which aims to develop tools to reliably and validly measure patient-reported outcomes in adults. Similarly, the goal of the common data element project at the National Institute of Neurological Disorders and Stroke is to standardize the collection of investigational data to facilitate comparison of results across studies and more effectively aggregate information into significant metadata sets.

The aim of this systematic review is to assess the standardized outcome measures currently used in pediatric stroke studies, which will serve as a foundation for understanding the appropriate measures for clinical trials in this population.

METHODS

INCLUSION AND EXCLUSION CRITERIA

Eligible studies included children from birth to age 18 years with ischemic stroke, hemorrhagic stroke, or both, had more than 5 subjects, and evaluated children for neurological or functional outcome status with a recog-
The initial search returned 2996 unique studies, of which 30 were suitable for inclusion. Reasons for exclusion are detailed in **Figure 1**. Hand-searching found an additional 4 studies, resulting in 34 included studies. Of note, 36 studies were excluded because ad hoc, descriptive outcome measures were used rather than standardized measures. For example, many studies defined outcome solely by reporting neurological sequelae (e.g., hemiparesis, epilepsy, cognitive impairment, motor deficits), while others used subjective stratifications such as mild, moderate, or severe deficits.

A detailed description of each included study with aim, sample size and characteristics, and outcome measures is shown in supplemental Table 1 (http://kc.vanderbilt.edu/kennedy_pdfs/jordanLori_supp.pdf). Of the 34 studies, 19 were focused on ischemic stroke only, 5 were focused on hemorrhagic stroke only, and 10 included both types of stroke. Infants were exclusively the subjects of 8 studies, 8 studies included children older than 1 year only, and 18 included both age groups. A median of 2 outcome measures were used per study (range, 1-7 outcome measures). More than 1 outcome measure was used in 29 studies (85%).

The most commonly applied outcome measure was the age-appropriate form of the Wechsler Intelligence Scales, used in 34% of studies. The second most prevalent outcome measure was the Pediatric Stroke Outcome Measure (PSOM), used in 7 studies (21%); more prevalence details are provided in the Table. Notably, 24 of 38 outcome measures were used in 1 included study each (63%).

Descriptions and psychometric properties of outcome measures used more than once are detailed in the Table, with more detailed information and additional outcome measures provided in supplemental Table 2. A standardized neurological examination, the PSOM, has been validated for infants and children with ischemic stroke. Of the 12 most used outcome measures, 9 (75%) have been validated in children. Interrater reliability data were variable both for specific outcome measures (e.g., the Glasgow Outcome Scale, ranging from 0.31-0.79 depending on the study) and across all outcome measures. Most tools are pediatric measures of cognitive ability (California Verbal Learning Test—Children’s Version, Griffiths Scales of Mental Development, Stanford-Binet Intelligence Scale, Wechsler Intelligence Scales), development (Bayley Scales of Infant Development, Denver Developmental Screening Tests), or overall health (Child Health Questionnaire, Glasgow Outcome Scale, modified Rankin Scale, 36-Item Short Form Health Survey, Vineland Adaptive Behavior Scales).

All included studies using standardized outcome measures were conducted within the previous 2 decades. **Figure 2**A shows graphically that pediatric stroke studies using recognized outcome measures are increasingly prevalent in the literature. The temporal application of the most used outcome measures, those used 3 or more times, is shown in Figure 2B, which demonstrates the increased use of a variety of outcome measures over time. Notably, use of the validated PSOM has increased.

**RESULTS**

**EleCtronIC sEarch MeThod**

Electronic searches of CINAHL, EMBASE, PubMed, and Web of Science were performed in August 2010 using a combination of all relevant PubMed medical subject heading terms and keywords relating to children, stroke, and outcome measures. Identified studies were imported into a reference manager, and duplicates of identical studies were removed.

Two raters agreed on and independently applied inclusion and exclusion criteria to the studies in title, abstract, and full-text reviews. The reference lists of all included articles and of appropriate review articles were examined to identify additional relevant studies.

**sTudy ASseSSMeNT**

Study quality was assessed insofar as each study was screened for clear reporting of methods and data. Information on validity, reliability, and general and psychometric data on standardized outcome measures was retrieved from both the outcome measures' sources and original studies.

**sTaTIsTIcAL ANAlySIs**

Frequency of outcome measure use as well as information on validity and reliability were compiled. This study was designated exempt by the institutional review board.
Currently, there are wide variations in the application of pediatric stroke outcome measures. At this time, the PSOM, a standardized neurological examination, is the best validated outcome measure with direct validation in children aged 0 to 18 years with ischemic stroke. Yet, in pediatric stroke there are many potential domains to assess, including but not limited to adaptive functioning, cognition, emotional health, behavior, and quality of life. Preferably, overall health, cognitive development, and physical development would be assessed both objectively and subjectively, from the patient’s and/or caregiver’s perspective. Agreement among researchers, clinicians, and perhaps patients and their families regarding key outcome domains to measure is necessary prior to undertaking further validation and reliability studies for pediatric stroke outcome tools. We can learn much from other fields in this regard. For example, researchers in neuromuscular disease have pushed for outcome measures that are not only valid and reliable but also responsive to improvement or loss of function so as to capture clinically rel-

### Table. Most Used Outcome Measures in Included Pediatric Stroke Studies

<table>
<thead>
<tr>
<th>Outcome Measure</th>
<th>Descriptiona</th>
<th>Age Range</th>
<th>Duration, min</th>
<th>Stroke Validity</th>
<th>Pediatric Validity</th>
<th>Pediatric Stroke Validity</th>
<th>Interrater Reliability</th>
<th>Studies, No.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bayley Scales of Infant and Toddler Development</td>
<td>Evaluates developmental delay in infants and preschoolers5</td>
<td>1-42 mo5</td>
<td>30-90, age dependent5</td>
<td>No</td>
<td>Yes, concurrent15</td>
<td>No</td>
<td>ICC = 0.47-0.96, scale dependent2</td>
<td>6</td>
</tr>
<tr>
<td>California Verbal Learning Test–Children’s Version</td>
<td>Assesses verbal learning and memory in children and adolescents7</td>
<td>5-16 y11</td>
<td>35-4011</td>
<td>No</td>
<td>Yes, concurrent13</td>
<td>No</td>
<td>NA</td>
<td>4</td>
</tr>
<tr>
<td>Child Health Questionnaire</td>
<td>Assesses child’s physical, emotional, and social well-being from the perspective of parent/guardian or child; can be administered by anyone26</td>
<td>5-18 y13</td>
<td>10-2513</td>
<td>No</td>
<td>Yes, discriminant13</td>
<td>No</td>
<td>NA</td>
<td>2</td>
</tr>
<tr>
<td>Denver Developmental Screening Tests</td>
<td>Determines whether child’s development is within normal range; can be administered by anyone28</td>
<td>&lt;6 y25</td>
<td>2616-17</td>
<td>No</td>
<td>Yes, concurrent16</td>
<td>No</td>
<td>k = 0.7514</td>
<td>2</td>
</tr>
<tr>
<td>Glasgow Outcome Scale</td>
<td>Global measure of outcome with 6 categories27</td>
<td>None specified</td>
<td>519</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Weighted k = 0.31-0.7913</td>
<td>2</td>
</tr>
<tr>
<td>Griffiths Scales of Mental Development</td>
<td>Obtains level of mental development in infants and young children29</td>
<td>0-8 y27</td>
<td>50-6027</td>
<td>No</td>
<td>Yes, construct27</td>
<td>No</td>
<td>Overall k &gt; 0.6022</td>
<td>3</td>
</tr>
<tr>
<td>Modified Rankin Scale</td>
<td>Global measure that focuses on symptoms and disability after stroke30</td>
<td>Designed for &gt;60 y23</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pediatric Stroke Outcome Measure</td>
<td>Neurological assessment tool that measures right and left sensorimotor function, language production and comprehension, and cognitive and behavioral performance24</td>
<td>0-18 y25</td>
<td>NA</td>
<td>Yes, construct25b</td>
<td>Yes, construct25</td>
<td>91%25</td>
<td>7</td>
<td></td>
</tr>
<tr>
<td>36-Item Short Form Health Survey</td>
<td>Generic measure that yields a holistic health profile; can be administered by self, computer, or trained interviewer in person or via telephone26</td>
<td>≥14 y25</td>
<td>5-1026</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>2</td>
</tr>
<tr>
<td>Stanford-Binet Intelligence Scale</td>
<td>Assesses intelligence and cognitive abilities in fluid reasoning, knowledge, quantitative processing, visuospatial processing, and working memory28</td>
<td>≥2 y26</td>
<td>5/subtest27</td>
<td>No</td>
<td>Yes, concurrent27</td>
<td>No</td>
<td>IRR = 0.74-0.9728</td>
<td>2</td>
</tr>
<tr>
<td>Vineland Adaptive Behavior Scales</td>
<td>Assesses personal and social function via communication, daily living skills, socialization, and motor skills29</td>
<td>0-18 y29</td>
<td>20-6029</td>
<td>No</td>
<td>Yes, criterion29</td>
<td>No</td>
<td>IRR = 0.62-0.7829</td>
<td>2</td>
</tr>
<tr>
<td>Wechsler Intelligence Scales</td>
<td>Measures ability to adapt and constructively solve problems in the environment; consist of 3 scales: Wechsler Preschool and Primary Scale of Intelligence (ages 4-6.5 y), Wechsler Intelligence Scale for Children (ages 6-16 y), and Wechsler Adult Intelligence Scale28</td>
<td>4 y to adult31</td>
<td>60-9031</td>
<td>No</td>
<td>Yes, concurrent30</td>
<td>No</td>
<td>NA</td>
<td>15</td>
</tr>
</tbody>
</table>

Abbreviations: ICC, intraclass correlation coefficient; IRR, interrater reliability coefficient; NA, not applicable.

a Unless otherwise stated, all tests are administered by a psychologist or physician.
b Validated for arterial ischemic stroke and cerebral venous sinus thrombosis but not intracerebral hemorrhage.
have also been performed to better use data from clinic-floor and ceiling effects, and increased precision, calibration of items across a broad range to over-level (linear) scaling for better interpretation of variation. Item response theory–based scales have interval-functioning levels are selected, an algorithm whereby only the most informative items. Computerized adaptive tests to measure outcomes more effectively capture function and disability after stroke in children.

The creation of outcome assessment guidelines will facilitate appropriate outcome measure selection for future studies as well as communication and comparison of treatment results. If a gold-standard pediatric stroke outcome assessment is not established, the comparability of pediatric stroke trial results will be undermined, potentially delaying the effective treatment of pediatric patients with stroke for years to come.

**Figure 2.** Number of pediatric stroke studies using validated outcome measures (A) and outcome measures used most over time (B). WAIS indicates Wechsler Adult Intelligence Scale; WISC, Wechsler Intelligence Scale for Children; and WPPSI, Wechsler Preschool and Primary Scale of Intelligence.

Advanced statistical methods have also been performed to better use data from clinical trials in neurorehabilitation and multiple sclerosis. To chart the way forward in pediatric stroke, collaboration among pediatric stroke professionals, clinical trialists, and experts in statistics and clinimetrics is needed. There are many competing issues; for the purposes of clinical trials, investigators would prefer a single composite measure of global outcome such as the modified Rankin Scale used in adult stroke that does not require scoring by a physician or psychologist. A battery of measures is more costly and complex but in theory would better capture function and disability after stroke in children. Finally, patient-reported outcomes have become increasingly important. The PROMIS uses modern psychometric methods, including item response theory, to construct question banks that may be used to create computerized adaptive tests to measure outcomes more efficiently and precisely. Computerized adaptive tests use an algorithm whereby only the most informative items targeting an individual's functioning levels are selected, thus reducing the burden of traditional fixed-length questionnaires that may force patients to answer irrelevant items. Item response theory–based scales have interval-level (linear) scaling for better interpretation of variation, calibration of items across a broad range to overcome floor and ceiling effects, and increased precision to allow more sensitivity to change. In traditional ordinal scale–based outcome measures such as the modified Rankin Scale, a 1-point change from 5 (severe disability) to 4 (moderately severe disability) is not the same distance as a 1-point change from slight to no disability, making change over time more difficult to interpret.

While the PROMIS has significant advantages, it is important to remember that in young children, patient-reported outcomes are measured via parental responses or proxy. In studies of the most widely used pediatric quality-of-life measure, the Pediatric Quality of Life Inventory (supplemental Table 2), only moderate correlation was found between self- and proxy-report in older children; parents consistently underestimated their child's health-related quality of life, perhaps due to anxiety. Better correlation was found in children with chronic health conditions (correlations ranging from 0.5-0.61) and for physical rather than psychological and social proxy-reports. Given these concerns, patient-reported outcomes should not be the only outcomes in children. Particularly in young children who are difficult to assess, developmental measures may still be needed.

**Accepted for Publication:** May 18, 2011.

**Correspondence:** Lori C. Jordan, MD, PhD, Department of Neurology, Division of Pediatric Neurology and Stroke, Vanderbilt University Medical Center, 2200 Children’s Way, DOT 11242, Nashville, TN 37232 (lori.jordan@vanderbilt.edu).

**Author Contributions:** Mr Engelmann and Dr Jordan had full access to all the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis. Study concept and design: Engelmann and Jordan. Acquisition of data: Engelmann and Jordan. Analysis and interpretation of data: Engelmann and Jordan. Drafting of the manuscript: Engelmann and Jordan. Critical revision of the manuscript for important intellectual content: Engelmann and Jordan. Statistical analysis: Engelmann and Jordan. Administrative, technical, and material support: Engelmann. Study supervision: Jordan.

**Financial Disclosure:** Dr Jordan has served as a consultant for Berlin Heart.

**Funding/Support:** Dr Jordan is supported by grant K23NS062110 from the National Institute of Neurological Disorders and Stroke.

**Role of the Sponsor:** The funding agency had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; or preparation, review, or approval of the manuscript.

**Additional Contributions:** Victoria H. Goode, MLIS, assisted with the design and conduct of the study.
REFERENCES

34. Hobart JC, Cano SJ, Thompson AJ. Effect sizes can be misleading: is it time to change the way we measure change? J Neurol Neurosurg Psychiatry. 2010; 81(9):1044-1048.