Pediatric RESEARCH

CLINICAL RESEARCH ARTICLE

Validation of the pediatric stroke outcome measure for classifying overall neurological deficit

Mahmoud Slim[1], Christine K. Fox[2], Sharon Friefeld[3], Nomazulu Dlamini[1], Robyn Westmacott[4], Mahendranath Moharir[1], Daune MacGregor[1] and Gabrielle deVeber[1] On behalf of the SIPS Investigators

BACKGROUND: The pediatric stroke outcome measure (PSOM) is a standardized, disease-specific outcome measure. We aimed to validate the overall classification of neurological deficit severity using PSOM.

METHODS: We identified 367 neonates/children with arterial ischemic stroke (AIS) (Derivation Cohort). We analyzed the PSOM subscales (scored as 0 [no deficit], 0.5 [minimal/mild deficit; normal function], 1 [moderate deficit; slowing function], or 2 [severe deficit; missing function]) to derive severity levels using latent class analysis (LCA). We validated a severity classification scheme (PSOM-SCS) in: (a) children who had Pediatric Evaluation of Disability Inventory (PEDI; $n = 63$) and/or the Pediatric Quality-of-Life Inventory (PedsQL; $n = 97$) scored; and (b) an external cohort (AIS; $n = 102$) with concurrently scored modified Rankin Scale (mRS), King's Outcome Scale for Childhood Head-Injury (KOSCHI) and PSOM.

RESULTS: Within the Derivation Cohort, LCA identified three severity levels: "normal/mild," "moderate," and "severe" (83.7%, 13.3%, and 3%, respectively). We developed severity classification based on PSOM subscale scores: "normal/mild"—normal function in all domains or slowing in one domain, "moderate"—slowing in ≥2 domains or missing function in one domain, and "severe"—missing function in ≥2 domains or slowing in ≥1 plus missing in one domain. PEDI and PedsQL both differed significantly across the severity groups. PSOM-SCS displayed high concordance with mRS (agreement coefficient [AC2] = 0.88) and KOSCHI (AC2 = 0.79).

CONCLUSION: The PSOM-SCS constitutes a valid tool for classifying overall neurological severity emphasizing function and encompassing the full range of severity in pediatric stroke.

Pediatric Research (2020) 88:234–242; https://doi.org/10.1038/s41390-020-0842-5

IMPACT:

- Arithmetic summing of the PSOM subscales scores to assess severity classification is inadequate.
- The prior severity classification using PSOM overestimates poor outcomes.
- Three distinct severity profiles using PSOM subscales are identified.
- The PSOM-SCS is in moderate to excellent agreement with other disability measures.
- PSOM-SCS offers a valid tool for classifying the overall neurological deficit severity.

INTRODUCTION

Pediatric arterial ischemic stroke (AIS) represents a serious cause of lifelong neurological disabilities affecting 1.2–8 per 100,000 children/year and one in 2300–5000 live births in neonates.[1–3] Neurologic deficits in sensorimotor function, language production or comprehension, and cognitive and behavioral function affect more than half of pediatric stroke survivors.[4,5] In addition to physical disabilities, pediatric stroke confers significant language, cognitive, emotional, psychological, and socioeconomic burdens.[6,7]

Compared to adults, childhood stroke outcome assessment is challenging because of changing age-specific functions over childhood and across normal development.[8] According to the International Classification of Functioning, Disability, and Health for children and youth developed by World Health Organization, relevant assessment requires thorough consideration of body functions, body structures, and activities, as well as participation in physical, social, and psychological development.[9] Over the past three decades, the pediatric stroke outcome measure (PSOM) has played a key role in evaluating the body function domain in children with pediatric stroke.[10,11]

Several other outcome measures have been applied to assess and classify outcome severity following pediatric stroke including the modified Rankin Scale (mRS) and the Pediatric Evaluation of Disability Inventory (PEDI).[12,13] However, the PSOM remains the only disease-specific outcome measure for this patient population.[10] The PSOM is a standardized neurological examination that includes five equally weighted subscales (sensorimotor right, sensorimotor left, language production, language comprehension,

[1]Department of Pediatrics, Division of Neurology, The Hospital for Sick Children, Toronto, ON, Canada; [2]Departments of Neurology and Pediatrics, University of California, San Francisco, San Francisco, CA, USA; [3]Department of Occupational Science and Occupational Therapy, University of Toronto, Toronto, ON, Canada and [4]Department of Psychology, The Hospital for Sick Children, Toronto, ON, Canada
Correspondence: Gabrielle deVeber (gabrielle.deveber@sickkids.ca)
A list of SIPS Investigators and their affiliations are in the Supplementary Information.

Received: 3 September 2019 Accepted: 6 February 2020
Published online: 16 March 2020

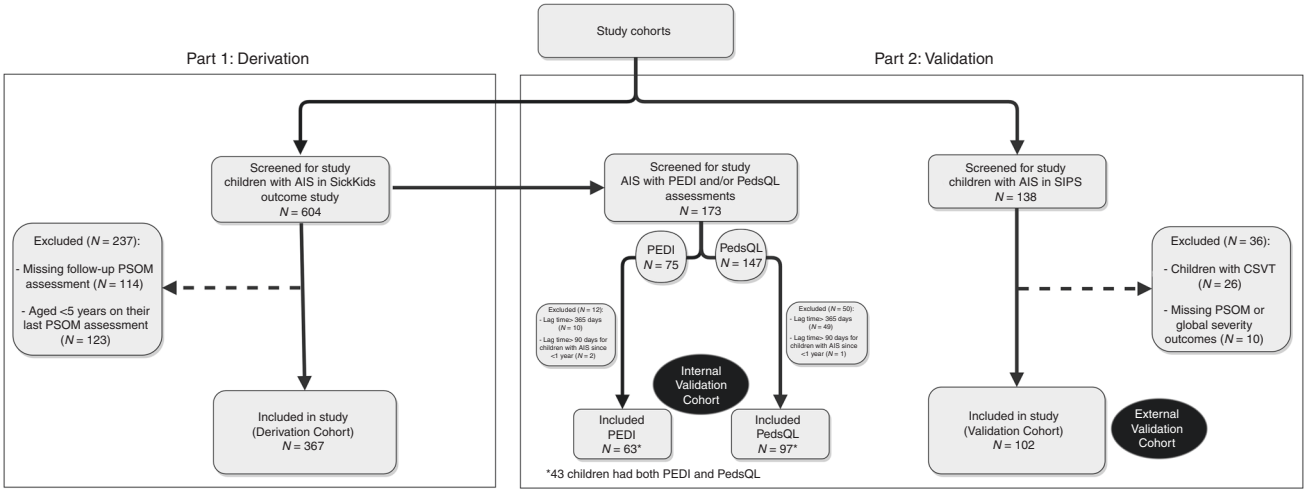

Fig. 1 Flow diagram. The "Derivation Cohort" consisted of children who had PSOM examinations at age ≥5 years old ($N = 367$). The "Internal Validation Cohort" consisted of children who were evaluated using the Pediatric Evaluation of Disability Inventory ($N = 63$) and/or Pediatric Quality of Life Inventory ($N = 97$). The "External Validation Cohort" consisted of unique children with AIS from the multi-institutional Seizures in Paediatric Stroke study ($N = 102$).

and cognitive/behavioral) each scored from 0 to 2.[10,14] Taken together, these subscales encompass overall neurological function. The validity and reliability of the five individual PSOM subscales in measuring domain-specific neurological impairment, both prospectively and retrospectively, has been well established against standardized neuropsychological measures.[10]

In most outcome studies, PSOM outcome scores have been dichotomized using arbitrarily "cut-off" values, for example, total score "0 or 0.5" vs. "1 or greater" to define "good" vs. "poor" groups, respectively.[13,14] However, with this approach, "poor" collapses patients with normal function (despite minor neurological findings), slowed but not missing function, a single missing function, and multiple missing functions into one large heterogeneous group. Previous attempts to combine subscale scores of the PSOM to classify overall outcome severity included (1) simple arithmetic sum of the five subscale scores from 0 (no deficit) to 10 (maximum deficit) and (2) a more complex subscale score combination scheme.[14] The arithmetic sum of the PSOM score from 0 to 10 is unsatisfactory as scores can overlap for children with missing function and children with normal function. For instance, a total score of 2 can be obtained from a score of 0.5 on four subscales (no missing function) or from just a score of 2 on one single subscale (missing function). The prior combination scheme published in 2000 demonstrated correlation with quality of life (QOL) measures in pediatric stroke survivors;[14,15] however, it tended to overestimate poor outcomes.[13,16]

There is a clear need for a different approach that helps us identify distinct severity classes based on the PSOM subscales. Global neurological outcome based on combining PSOM subscale scores has not been validated to date and the optimal number of severity classes remains unknown. Latent class analysis (LCA), which identifies hidden, homogeneous sub-groups (subclasses) within larger populations, could be applied to better determine distinct severity profiles among children with stroke.[17] The LCA model has been used in various domains, including behavioral research, diagnostic testing accuracy (in the absence of gold-standard tests), psychology, and education, to identify homogeneous profiles sharing a common trait/abnormality.[18–21] The objective of this study was to develop a new scheme for pediatric stroke outcome classification based on PSOM subscale scores that would (1) accurately reflect the full range of neurological body functions and ability, (2) distinguish between the different severity profiles following AIS, and (3) prove to be valid against standardized function/disability and QOL measures.

METHODS

Study design

We conducted a study in children with AIS to develop and validate an overall severity classification scheme (SCS) for the PSOM using both clinical and statistical approaches. The study comprised two parts: (1) the derivation of a SCS and (2) validation of the SCS against disability and QOL measures in an overlapping internal cohort and against global disability measures in a distinct, external cohort of patients (Fig. 1).

Patient population

We screened neonates and children (term birth to 18 years old) prospectively enrolled in the institutional Children's Stroke Outcome Study at the Hospital for Sick Children between January 1994 and August 2017. The criteria for AIS were: (1) acute focal neurologic deficit or seizure consistent with stroke, and (2) arterial infarct conforming to an arterial territory on computed tomography or magnetic resonance imaging consistent with clinical symptom localization and timing. Infants with presumed perinatal AIS were excluded.

All children had undergone serial PSOM assessments in our Stroke Clinic. We selected all neonates/children who had PSOM examinations performed at age 5 years or older as our Derivation Cohort; we excluded children with their last PSOM below this age because they may later manifest emerging deficits with maturation.

Next, we identified two Validation Cohorts: (1) the Internal Validation Cohort comprised of children who had assessments with the PEDI or Pediatric Quality of Life Inventory (PedsQL) measures at SickKids.[22] In order to account for shifts in severity in the early post-stroke recovery period and ensure a reasonable inter-test time interval, we restricted this subset to children that had had the PSOM ≥5 years of age AND (1) all assessments (PSOM, PEDI, and/or PedsQL) were performed >1 year post-stroke AND a maximum inter-test interval of 365 days, OR (2) Any of PSOM, PEDI, or PedsQL performed 3–12 months after stroke AND a maximum inter-test interval of 90 days; (2) an External Validation Cohort comprised of unique children with AIS from the multi-institutional Seizures in Pediatric Stroke (SIPS) study.[23] All children in the Validation Cohort had undergone concurrent PSOM, King's Outcome Scale for Childhood Head Injury (KOSCHI) and mRS assessments 1 year after stroke onset as part of the SIPS study. Details of the study protocol and procedures have been published elsewhere.[23]

Table 1. Severity classification based on the PSOM subscales scores ($N = 367$).

	PSOM-severity classification scheme (current)	*N* (%)	PSOM-prior severity classification	*N* (%)
Normal	0–0.5 in all subscales	181 (49.3)	0 in all subscales	93 (25.3)
Mild	1 in 1 subscale and <1 in the remaining subscales	84 (22.9)	0.5 in 1 subscale	69 (18.8)
Moderate	1 in ≥2 subscales OR 2 in only 1 subscale and <1 in all remaining subscales	60 (16.4)	0.5 in 2, 3, or 4 subscales OR 1 in 1 subscale OR 1 in 1 subscale and 0.5 in another	92 (25.1)
Severe	2 in only 1 subscale and ≥1 in any of remaining subscales OR 2 in ≥2 subscales	42 (11.4)	0.5 in all subscales OR 1 in 1 subscales and 0.5 in 2 subscales OR 1 in ≥2 subscales OR 2 in ≥1 subscale	113 (30.8)

Study procedures

LCA for development of the PSOM-SCS. Within the Derivation Cohort, we entered the five PSOM subscale scores (each scored 0, 0.5, 1, or 2) into a "LCA" software program to derive the optimal number of inherent distinct outcome classes (between 1 and 10 classes) (statistical derivation). The program empirically identifies patterns of severity by assigning children to a set of discrete, mutually exclusive classes based on the severity of deficit on each of the five PSOM subscales.

We separately developed criteria for classifying overall neurological outcome severity based on individual subscales scores in order to differentiate children with normal function, decreased/slowed (but not missing) function, and missing functions (disability). The PSOM subscales are each scored as 0 (no deficit), 0.5 (minimal/mild deficit, normal function), 1 (moderate deficit, slowing but not missing function), or 2 (severe deficit, at least one missing function) (Supplemental Table S1 online).[10] The definitions for the classes in the revised PSOM-SCS are summarized in Table 1. Normal function on all subscales (a combination 0's and 0.5's) was classified as overall "normal," while the presence of slowing of function (a score of 1) on only one subscale led to "mild" classification. Slowing function on more than one subscale or missing function (a score of 2) on only one subscale was classified as "moderate," whereas missing function in only one subscale and slowing in any remaining subscales OR missing function in multiple subscales was classified as "severe."

Within the Derivation Study Cohort, we compared the agreement of the PSOM-SCS with the LCA-determined distinct severity groups and the agreement of the prior classification system with the LCA groups.

Validation procedures. In the subset of children at SickKids with PEDI and/or PedsQL scores, we determined the congruence of the PSOM-SCS and the prior classification system with these two measures of disability and QOL. We performed cross-class analysis comparing PEDI and PedsQL continuous scores based on the PSOM-SCS and the prior classification system categories.

We also examined the performance of the PSOM-SCS and the prior classification system in the External "SIPS" Validation Cohort by evaluating the agreement in outcome severity between PSOM-SCS with mRS and KOSCHI. We also compared PSOM-SCS to parental/child responses to the Recovery Recurrence Questionnaire across the PSOM-SCS categories.[24]

Outcome measures

Pediatric stroke outcome measure. The PSOM includes 115 test items encompassing observed functions in behavior, cognition, language, cranial nerve, motor, sensory, cerebellar, and gait sections. In each section, test items are organized sequentially and developmentally from infancy to teenage years. In the "summary of impressions," the examining neurologist scores each of five PSOM subscales (right sensorimotor, left sensorimotor, language expression, language comprehension, cognitive/behavioral) in terms of the child's abilities based on patient history and results of the PSOM neurological examination, as 0 (no deficit), 0.5 (minimal/mild deficit, normal function), 1 (moderate deficit, slowing but not missing function), or 2 (severe deficit, at least one missing function) (Supplementary Table S1 online).

Pediatric Evaluation of Disability Inventory. The PEDI utilizes parental and teacher self-report across 197 questions to measure functional abilities in children.[25] The PEDI also evaluates the need for help provided by caregiver (distributed over 20 questions) for self-care, mobility, and social functions.

Pediatric Quality of Life Inventory. The PedsQL, a multidimensional measure of health-related QOL utilizes child self-report (age 5–18 years) and parental report (age 2–19 years).[26] The 23 questions are distributed over four sub-scales: physical, emotional, social, and school. The PedsQL yields three summary scores: total score, psychosocial health summary score, and physical health score.

King's Outcome Scale for Childhood Head Injury. The KOSCHI, a pediatric adaptation of the Glasgow Outcome Score, measures disability and recovery in children with head injury.[27] KOSCHI includes five main severity groups: normal (5b), mild (5a), moderate (4a and 4b), and severe (3a, 3b, 2, and 1).

Modified Rankin Scale. The mRS, a 6-point ordinal scale, assesses disability in adults with stroke emphasizing motor function. The mRS modified for use in children adds a school performance question and correlates with general functioning and QOL in children with AIS.[13,28] Outcomes are classified as normal (grade 0), mild (grades 1–2), moderate (grade 3), or severe (grades 4–6).[28]

Recurrence and Recovery Questionnaire. The Recurrence and Recovery Questionnaire (RRQ) was derived from the PSOM using a lay language suited to parental survey.[24] The RRQ includes questions relating to clinical recovery (Has your child recovered completely from the stroke?), the need for help in day to day activities (Does your child need extra help with day-to-day activities compared to other children their age?), emotional status (Has the stroke affected your child's emotional state, behavior and feelings about his/herself?), and the use of assistive devices (Does your child use aids or assistive devices?).

This study was approved by the Research Ethics Boards at the Hospital for Sick Children. Written patient or parental informed consent was obtained from all study participants.

Statistical analysis

Continuous variables were presented as mean±standard deviation (SD) or median with interquartile ranges (IQRs), as appropriate. Qualitative variables were described using frequency distributions.

For the LCA, we used the Vuong–Lo–Mendell–Rubin (VLMR) likelihood ratio test to compare models.[29] The VLMR test compares the fit of a K0 class model vs. an alternative K1 -class model; significant p values suggest that the estimated model fits the data better than a model with one less group.[29] LCA was conducted using the statistical package poLCA in R-3.4.1 and MPlus demo Version-7. After selecting the best-fitting model, we assessed concordance between the LCA-generated severity classifications and the PSOM-SCS using the second-order agreement coefficient (AC2).[30] AC2 measures levels of agreement similar to Cohen's κ; however, it has the advantage of being more accurate in the presence of imbalance in the table's marginal totals (skewed distributions) and if an ordered categorical classification such as the PSOM-SCS is being evaluated.[31] We also evaluated the agreement between latent classes with the severity groups originating from the prior classification system.[14]

Group comparisons of PSOM-SCS and the previously published PSOM severity score to PEDI and PedsQL were conducted using Kruskal–Wallis tests and multiple pairwise comparisons were evaluated using Dwass–Steel–Critchlow–Fligner test, a post hoc non-parametric test used to correct for multiple group comparisons.[32] In the Validation Cohort, agreement analyses between PSOM-SCS with mRS and KOSCHI severity classes were assessed using AC2. Based on the published criteria, agreement was defined as poor, fair to moderate, moderate, and excellent for coefficients <0.20, 0.21–0.40, 0.41–0.80, and >0.80, respectively.[33] Differences in RRQ domains between the different severity profiles of the PSOM-SCS were evaluated using χ^2 or Fisher's exact test, as appropriate. Statistical analyses were undertaken using SAS University Edition (SAS Institute, Inc., Cary, NC).

RESULTS

Patient population

The Derivation Cohort included 367 children with neonatal ($n = 69$) or childhood ($n = 298$) AIS from SickKids who had had a PSOM examination after 5 years of age. Children had a median age of 5 years at stroke onset (IQR: 0.54–9.5) and 60% were male (Table 2). The Internal Validation Cohort consisted of 117 children from SickKids who had PEDI ($n = 63$) and/or PedsQL ($n = 97$) in addition to PSOM and met the inter-test interval criteria. Of these 117 children, 71 were also part of the Derivation Cohort. The External Validation Cohort consisted of 102 children from 14 North American and 7 non-North American SIPS study centers with PSOM, KOSCHI, and mRS scored 1 year after AIS (26 neonatal strokes, median age: 3 years at stroke onset, IQR: 0.03–9.7; 57% males). Figure 1 illustrates the flow diagram for the screening and selection of children in each of the study cohorts.

Development of the novel PSOM-SCS

Latent class analysis. The 3-class model was superior to either the 2-class or the 4-class model. As shown in Fig. 2, the p value for the 2-class model using the VLMR test was <0.001, which indicated that the 2-class model provided a better fit compared to a 1-class model and the 3-class model provided a better fit compared to a 2-class model ($p = 0.0021$); in addition, a 4-class model did not provide a better fit compared to a 3-level model ($p = 0.1113$). The probability of scoring 0, 0.5, 1, or 2 using the 3-class model on each of the subscales is shown in Supplementary Fig. S1 (online). Using the LCA 3-class model, class 1 children ($n = 307$, 83.7%) were more likely to have *normal* function and ability (a score of 0 or 0.5) on all five subscales. Class 2 children ($n = 49$, 13.3%) tended to have *decreased* function. Class 3 included 11 (3%) children who had the highest probability of *missing* function (Supplementary Fig. S1 online).

Severity classification scheme. Based on both/LCA findings and clinical judgment, we developed the PSOM-SCS criteria illustrated in Table 1. Consistent with the 3-class LCA model, we grouped patients as: "normal/mild," "moderate," and "severe." There were 265 (72.2%) children classified as "normal/mild" who were normal or had trivial abnormal physical findings (e.g., tendon reflex asymmetry) or slowing of function in one subscale (e.g., slowed fine finger movements one hand) and no missing function. This category included 93 children who scored 0 on all five subscales, 124 who scored 0.5 on one or more subscales, and 48 with a score of 1 on only one subscale. The "moderate" category included 60 (16.4%) children who scored 1, that is, slowing of function, in two or more subscales, and/or scored 2, that is, missing function in only one subscale. The "severe" category included 42 (11.4%) children who scored 2, i.e. at least one missing function, in at least

Table 2. Demographic and clinical characteristics of children in the Derivation ($N = 367$) and Validation ($N = 102$) Cohorts.

Variable	Derivation Cohort	Validation Cohort
Age at onset, years (25–75 IQR)	5 (0.54–9.5)	3 (0.03–9.7)
Male, *n* (%)	220 (60)	58 (57)
Neonatal stroke, *n* (%)	69 (18.8)	26 (25.5)
Time to last follow-up visit (years)	6.8 (4–10)	1.1 (0.97–1.5)
Age at last PSOM assessment (years)	14.1 (8.8–17.2)	4.5 (1.5–11.5)
PSOM summary score at stroke onset	1 (0.5–2)	NA
PSOM summary score at last follow-up assessment	1 (0–2)	0.5 (0–2)
Time interval between		
Stroke onset and PSOM assessment (years)	6.8 (4–10)	1.1 (0.97–1.5)
PSOM and PEDI assessments (days)	74 (0–160)	NA
Stroke onset and PEDI assessment (years)	3 (2.1–5)	NA
PSOM and PedsQL assessments (days)	0 (0–169)	NA
Stroke onset and PedsQL assessment (years)	3.9 (2–6.3)	NA

Results are expressed as median (IQR: 25–75).
IQR interquartile range, *NA* not available, *PEDI* Pediatric Evaluation of Disability Inventory, *PedsQL* Pediatric Quality of Life Inventory, *PSOM* Pediatric Stroke Outcome Measure.

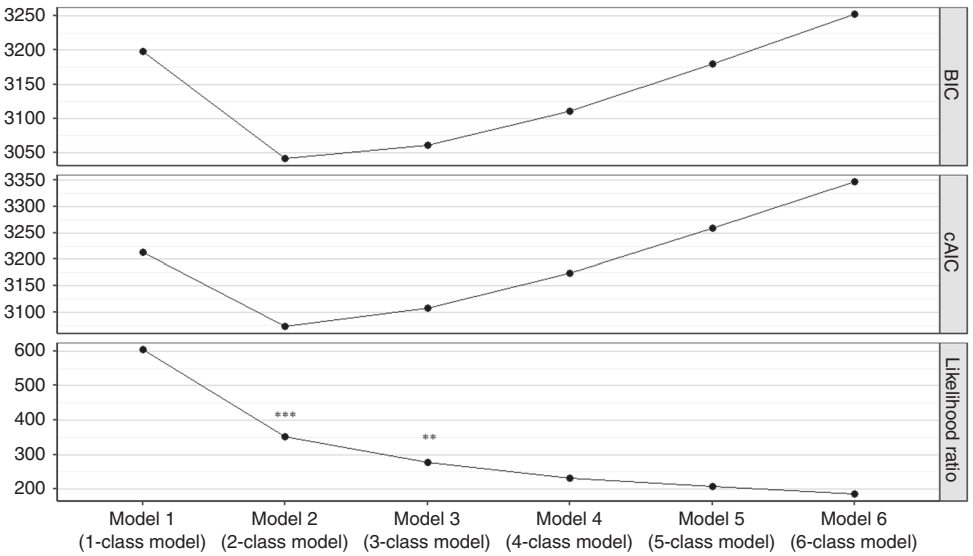

Fig. 2 Fit statistics for the different latent class analysis models. **: Vuong–Lo–Mendell–Rubin $p < 0.01$. P value for comparing the fit of a 3-class model to a 2-class model (null hypothesis: 2-class model is the best-fitting model). ***: Vuong–Lo–Mendell–Rubin $p < 0.001$. *P* value for comparing the fit of a 2-class model to a 1-class model (null hypothesis: 1-class model is the best-fitting model). BIC: Bayesian information criterion; cAIC: consistent Akaike's information criterion.

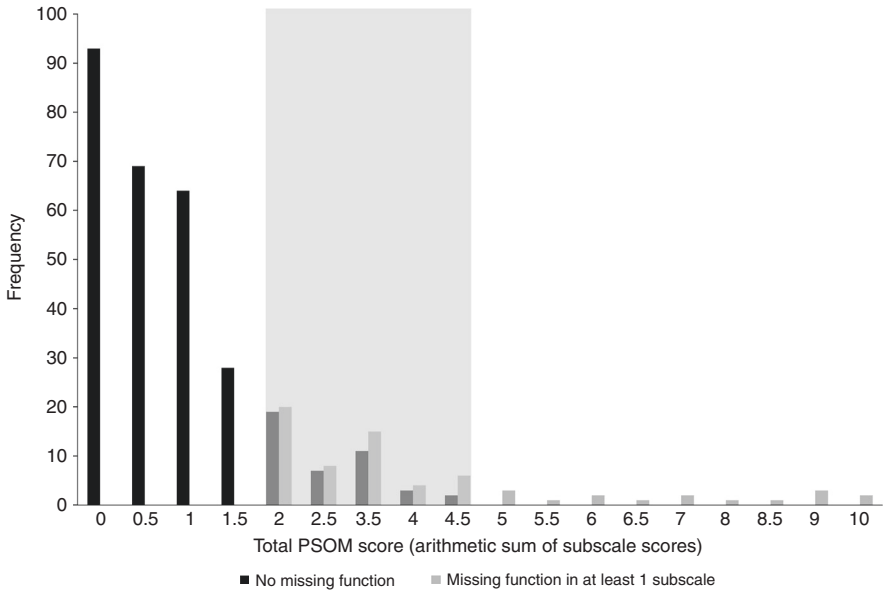

Fig. 3 Frequency distribution of the total PSOM scores in patients with or without missing neurologic function. Gray area: overlapping total PSOM scores for children with no missing function or with at least one missing function.

one subscale plus 1, i.e. slowing of function, in at least one other domain OR at least one missing function in ≥2 suscales; therefore, most likely to be disabled.

Agreement between current PSOM-SCS and prior PSOM classification system with the LCA model. The concordance between LCA and the PSOM-SCS in classifying pediatric stroke severity yielded an AC2 of 0.81 (95% confidence interval [CI]: 0.76–0.85); classification was concordant in 77.6% of cases. On the other hand, the agreement between LCA and the prior classification system resulted in an AC2 of 0.40 (95% CI: 0.31–0.49); classification was concordant in 48.7% of cases.

Mathematical summing of PSOM subscale scores. This yielded three categories of score ranges when examined based on the presence of missing function(s) (Fig. 3). Subjects with no missing function and those with missing function were clearly identified by summary scores 0–1.5 and 5–10, respectively. However, summary scores 2–4.5 contained both subjects with ($n = 53$) and without ($n = 42$) missing function, indicating the inadequacy of this approach to total severity classification on the PSOM.

Validation of the PSOM-SCS

Internal Validation Cohort: Comparison of PEDI and PedsQL scores across the different severity groups of the prior classification system vs. the PSOM-SCS. Among children screened for the study, an eligible subset of 117 children were assessed using both PEDI and PedsQL (43 children), PEDI alone (20 children), or PedsQL alone (54 children). PEDI assessments were done at a median 3 years after stroke and at intervals from the PSOM of maximum 82 days in the

Table 3. Comparison of PEDI and PedsQL domain scores among the severity groups of the new severity classification scheme and the prior severity classification (tested using Kruskal–Wallis and post hoc Dwass–Steel–Critchlow–Fligner test).

PEDI	PSOM-severity classification scheme					PSOM-prior severity classification			
	Total sample, $N = 63$	Normal/mild, $N = 49$	Moderate, $N = 7$	Severe, $N = 7$	P value	Normal/mild, $N = 16$	Moderate, $N = 32$	Severe, $N = 15$	P value
PEDI—Functional skills scale (mean ± SD)									
Self-care	74.9 ± 19.8	78 ± 19.2*	69 ± 12.1	56.4 ± 20.2	0.017	83.6 ± 17.2*^	66.6 ± 18.2	65.1 ± 19.2	0.002
Mobility	86.8 ± 14.9	88.3 ± 14.8	85.3 ± 14	78.2 ± 15.7	0.184	92.4 ± 10.9	78.8 ± 17.9*	83 ± 15	0.01
Social function	72.7 ± 20	76 ± 20	64.5 ± 11.4	57.7 ± 19.8	0.034	78.3 ± 18.2*	70 ± 22.3	63.7 ± 18.3	0.02
PEDI—Caregiver assistance scale (mean ± SD)									
Self-care	75.3 ± 21.04	78.1 ± 20.5	67.7 ± 15.4	58.7 ± 27	0.090	82.5 ± 18.5^	67.8 ± 21	66.7 ± 21.9	0.02
Mobility	86.6 ± 15.9	88.5 ± 15.4	80.9 ± 16.2	77.9 ± 17.3	0.176	91.6 ± 12	81.1 ± 19.6	81 ± 16.4	0.048
Social function	79.7 ± 18.4	83.2 ± 16.3*	73.1 ± 10.4	58.9 ± 27.6	0.019	87.4 ± 13.5*	75.5 ± 18.9	66.8 ± 19.8	0.01
PedsQL (mean ± SD)	Total sample, $N = 97$	Normal/mild, $N = 69$	Moderate, $N = 17$	Severe, $N = 11$		Normal/mild, $N = 21$	Moderate, $N = 47$	Severe, $N = 29$	
Total score	71.1 ± 17.9	76.4 ± 16.5*^	58.7 ± 13.7	56.5 ± 16.2	<0.0001	77.1 ± 17.4*	75.9 ± 14.2*	57.8 ± 14.2	<0.0001
Psychosocial health	69.2 ± 17.7	74 ± 16.9*^	58.4 ± 14.1	55.6 ± 13.6	<0.0001	75 ± 18.1	73 ± 14*	57.3 ± 13.5	<0.0001
Physical health	67.3 ± 18.9	70.6 ± 19.4*	61.3 ± 17.2	55.9 ± 10	0.01	71.7 ± 20.9*	68.6 ± 16.4	59.4 ± 14.6	0.01

PEDI Pediatric Evaluation of Disability Inventory, *PedsQL* Pediatric Quality of Life Inventory, *SD* standard deviation.
*$P < 0.05$ for post hoc pairwise comparisons of normal/mild vs. severe or moderate vs. severe.
^$P < 0.05$ for post hoc pairwise comparisons of normal/mild vs. moderate.

first post-stroke year and maximum 350 days in the subsequent post-stroke years. The PEDI results correlated with the PSOM-SCS classification. Children classified as normal/mild had the highest scores on the PEDI domains (least disability), followed by children classified with moderate severity by the PSOM-SCS. Patients classified as severe by the PSOM-SCS had the lowest scores (most disability) (Table 3). Differences in self-care (parent-report) and social function but not mobility scores were statistically significant across the PSOM-SCS classes ($p < 0.05$). In comparison, although the overall score differences were statistically significant, the prior PSOM severity classification did not differentiate between the moderate and severe classifications (Table 3).

Ninety-seven children were assessed with the PedsQL at median 3.9 years from stroke onset. PedsQL total and summary scores were significantly higher in the normal/mild group compared to moderate or severe groups using PSOM-SCS ($p < 0.05$) for total score, psychosocial health, and physical health (Table 3). Differences between the moderate and severe groups were not statistically significant. Similarly, differences across the severity groups of the PSOM-prior severity classification were statistically significant, but post hoc pairwise comparisons did not show any significant differences between normal/mild vs. moderate groups.

External Validation Cohort: Agreement of prior vs. current PSOM-SCS with disability measure scores. In the validation cohort, 86%, 11%, and 13% of children were classified as normal/mild, moderate, and severe respectively according to the PSOM-SCS. On the other hand, using the PSOM-prior severity classification, 45%, 29%, and 26% were classified as normal/mild, moderate, and severe, respectively. Agreement between the PSOM-SCS and KOSCHI was moderate at 75.5% (AC2 = 0.79; standard error [SE] = 0.04); however, 15.7% of children classified by PSOM-SCS as "normal/mild" were classified as "moderate" on the KOSCHI. For the mRS, agreement with the PSOM-SCS was excellent at 84.4% (AC2 = 0.88; SE = 0.03) and 5% of children classified as "normal/mild" on PSOM-SCS were "moderate" on the modified Rankin (Table 4). On the other hand, the agreement between the prior severity classification with KOSCHI and mRS was fair to moderate with less agreement compared to PSOM-SCS. Agreement was 63.8% (AC2 = 0.61) and 71.6% (AC2 = 0.68) of cases, respectively (Table 4).

Based on the RRQ survey, children with PSOM-SCS severe were more likely to require help in day-to-day activities (100%), use assistive devices (80%), and suffer from poor emotional status (50%) ($p < 0.001$) (Supplementary Fig. S2 online). While 69% of patients with normal/mild status reported complete recovery, only one child classified as moderate and none from the severe group reported complete recovery ($p < 0.001$).

DISCUSSION

We developed and validated an evidence-based overall severity classification based on PSOM scoring in children with AIS. We ascertained three distinct severity profiles: normal/mild, moderate, and severe and validated them against multiple standardized disability and QOL measures. By employing LCA, we demonstrated the superiority of a three-level severity classification over other models. Our corresponding severity classification definitions provide categorization that is meaningful to patients, families, and healthcare providers, with capacity to distinguish between normal function, decreased or slowing of function, and disability (multiple missing functions). The exclusion of children with PSOM performed aged <5 years in the Derivation Cohort reduced the potential bias that could have been attributed to the LCA model by including children most likely to grow into a later deficit.

We previously validated the five component PSOM subscales against relevant standardized neuropsychological measures.[10] However, validation of the global/overall neurological outcome based on the total PSOM has been lacking. Prior approaches for classifying the overall severity have been problematic. We demonstrated that arithmetic summing of the five subscale scores resulted in significant overlap in total PSOM scores for children with and without missing neurological function. Despite the historical tendency to dichotomize outcomes, we believe that a three-level classification approach as provided by the PSOM-SCS

Table 4. Agreement between PSOM-SCS and the prior classifications with the KOSCHI and mRS classification of disability.

	PSOM-severity classification scheme			PSOM-prior severity classification		
	Normal/mild	Moderate	Severe	Normal/mild	Moderate	Severe
KOSCHI						
Normal/mild (scores 5a and 5b), *N* (%)	61 (59.8)	3 (2.9)	0	31 (30.4)	2 (1.9)	1 (1)
Moderate (scores 4a and 4b), *N* (%)	16 (15.7)	10 (9.8)	6 (5.9)	15 (14.7)	28 (27.5)	19 (18.6)
Severe (scores 1–3), *N* (%)	0	0	6 (5.9)	0	0	6 (5.9)
AC2 (SE), 95% CI	0.79 (0.04), 0.71–0.87			0.61 (0.05), 0.5–0.72		
mRS						
Normal/mild (grades 0 and 2), *N* (%)	72 (70.6)	6 (5.9)	0	39 (38.2)	3 (2.9)	2 (1.9)
Moderate (grades 3), *N* (%)	5 (4.9)	7 (6.9)	5 (4.9)	7 (6.9)	27 (26.5)	17 (16.7)
Severe (grades 4–6), *N* (%)	0	0	7 (6.9)	0	0	7 (6.9)
AC2 (SE), 95% CI	0.88 (0.03), 0.82–0.94			0.68 (0.05), 0.57–0.79		

AC2 second-order agreement coefficient, *CI* confidence interval, *KOSCHI* King's Outcome Scale for Childhood Head Injury, *PSOM-SCS* Pediatric Stroke Outcome Measure-severity classification scheme, *mRS* modified Rankin Scale, *SE* standard error.

provides a more clinically relevant approach. On the spectrum of outcome from normal to severely disabled, the importance of distinguishing among children with intact function, minor limitations, and more severe limitations including disability is evident. In an era of aggressive and risky but potentially life-saving therapeutic interventions, including endovascular treatment,[34] the adult stroke studies rely on death/disability vs. alive, and non-disabled for decision making. It is important to utilize a similar "cut-point" now available in our PSOM-SCS if we are going to migrate such treatments down to children. Future studies employing the PSOM-SCS can employ the three-level approach or dichotomize at different levels, that is, normal/abnormal or non-disabled/disabled, depending on the study purpose.

The PSOM was designed to describe cognitive/behavioral, language, and sensori-motor deficits after a stroke. It was intended to provide clinically meaningful results. The PSOM is essentially a neurological examination and the presence of an abnormal finding on the neurological examination itself may or may not constitute a meaningful deficit, as exemplified by abnormal deep tendon reflexes, a positive Babinski, slightly asymmetric index finger tapping, and so on in a child without a corresponding functional consequences. The presence of such findings, however, does not fit the label "normal" as with a "normal neurological examination."

The prior PSOM severity classification over-weighted the finding of isolated neurological testing abnormalities in a child with completely normal function (subscale scores 0.5) by placing such a child into a more severe category. Therefore, in the current study we attempted to improve the PSOM summary score to make it more functionally meaningful by translating findings on the neurological examination into categories that are weighted towards relevant function. For these reasons, using the current PSOM-SCS we have classified the 0.5 scores along with normal, thereby changing the definition for the other categories by removing them from the other more severe categories. For children with a single domain having slowed function, but no loss of function (score of 1), we have classified them as "mild" and using the 3-class scheme, we categorized them along with "normal" such that "normal/mild" is the most favorable of the three outcome classes. This resulted in our shifting patients into the normal/mild category compared with prior classification.

Although refining the severity classification of the neurological function using PSOM-SCS does not negate the usefulness of previous outcome studies conducted in AIS, which distinguished children with deficit on one single domain vs. the aforementioned heterogeneous groups, we believe that the PSOM-SCS provides a new opportunity to better characterize neurological function among children with AIS and to be useful in future studies aiming to identify clinical and radiological predictors of outcomes.

We found excellent concordance between the PSOM-SCS and the LCA severity profiles. Although the PSOM-SCS is a measure of functional outcome rather than QOL, we also demonstrated significant differences of PedsQL scores across the PSOM-SCS classes. The latter scores are comprised of self-reports, parent reports and clinician reports, supporting validity over a wide range of perspectives. In accordance with international classification of functioning, disability, and health in children and youth,[9] our validation approach using proxy disability measures assessed body functions across a wide spectrum of domains, including physical, motor, speech, language, cognitive, and emotional ability, as well as activities and participation, including communication, mobility, self-care, and social function ensuring the relevance of the PSOM and PSOM-SCS.

Unlike the prior severity classification which does not differentiate between PEDI scores in the moderate and severe groups, we found a clear pattern of worsening in the PEDI functional skills and caregiver assistance scales scores across increasing levels of severity of the PSOM-SCS. However, the PEDI mobility domain was not significantly different across PSOM-SCS levels. This is likely due to the overall high PEDI scores (mean 86) indicating preserved mobility across our full cohort. This agrees with prior research showing that children with AIS have preserved mobility relative to other functional domains.[5] Differences in the physical health summary scores of the PedsQL across the severity groups of the PSOM-SCS were less pronounced than differences in the psychosocial health summary scores. These results align with prior findings that altered behavior and cognition have a greater impact than altered physical function on the global well-being of children after stroke.[35,36]

We demonstrated moderate to excellent agreement between the PSOM-SCS and the KOSCHI, and modified Rankin, whereas agreement of these disability measures with the prior severity classification was fair to moderate. While our previously published severity classification approach overestimated deficit severity compared to the modified Rankin,[13,14,16] the current PSOM-SCS appears to slightly underestimate severity compared with the KOSCHI and modified Rankin. Similar to PSOM, KOSCHI considers age-specific functions, cognitive functions, and neurological sequelae affecting function;[27] however, it has not been validated for use in children with AIS. The PSOM has advantages over modified Rankin and KOSCHI in that it accounts for age-specific evolving functions in longitudinal serial assessments and its

subscales represent global function distributed over 40% sensorimotor and 60% language/cognition domains.

Our study has several limitations. This was a retrospective evaluation of prospectively enrolled cohort in the institutional Children's Stroke Outcome Study. In addition, the study was limited by the lack of systematic evaluation of children using the disability measure (PEDI) in the Derivation Cohort, which was only available in a subset of 63 children. Nevertheless, the comparative analysis of PEDI scores among the different severity profiles was regarded as a confirmatory step following the LCA validation, which was conducted using a relatively larger sample size. In addition, despite excluding children who were very young at assessment and limiting the inter-test interval, since PEDI and PedsQL measures were not done on the same day as PSOM, deficits may still have emerged or recovered between assessments, which could have resulted in underestimating the PSOM-SCS validity in the PEDI and PedsQL samples. Finally, the aforementioned limitations of the mRS and KOSCHI disability measures in the validation constitutes another limitation in our study.

In clinical practice, assessing outcomes in terms of limitation in activity, participation, and environmental domains is crucial to ensure thorough and rigorous evaluation of health and well-being following pediatric stroke. The use of the PSOM coupled with disability measures that evaluate the impact of stroke on activities and participation, as well as environmental factors remains the cornerstone for the comprehensive evaluation of children with stroke. The development and validation of the PSOM-SCS has established the ability of the PSOM to measure and classify global neurological deficit and its validity for outcomes of relevance, including function and disability after pediatric stroke. The three-level outcome classification demonstrates promising discriminatory capacity based on disability and QOL measures and will enable the PSOM to be of greater value in future studies of pediatric stroke outcome, including clinical trials of therapies.

ACKNOWLEDGEMENTS

The study was funded by Auxilium Foundation, Pediatric Epilepsy Research Foundation (Grant-112010-007), and National Institutes of Health (2K12NS001692-11, KL2TR000143). These funders were not involved in study design, data analysis, or publication decisions. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

AUTHOR CONTRIBUTIONS

M.S., C.K.F., and G.d.V. contributed to the conception and design of the study, analysis and interpretation of data, drafting of the article, and final approval; S.F., N.D., R.W., M.M., and D.M. contributed to the acquisition of data, critical revision of the intellectual content, and final approval.

ADDITIONAL INFORMATION

The online version of this article (https://doi.org/10.1038/s41390-020-0842-5) contains supplementary material, which is available to authorized users.

Competing interests: The authors declare no competing interests.

Informed consent: Parental and (where appropriate) patient consent were obtained for all patients included in this study.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

REFERENCES

1. Giroud, M. et al. Cerebrovascular disease in children under 16 years of age in the city of Dijon, France: a study of incidence and clinical features from 1985 to 1993. *J. Clin. Epidemiol.* **48**, 1343–1348 (1995).
2. Fullerton, H. J., Wu, Y. W., Zhao, S. & Johnston, S. C. Risk of stroke in children: ethnic and gender disparities. *Neurology* **61**, 189–194 (2003).
3. Raju, T. N. et al. Ischemic perinatal stroke: summary of a workshop sponsored by the National Institute of Child Health and Human Development and the National Institute of Neurological Disorders and Stroke. *Pediatrics* **120**, 609–616 (2007).
4. deVeber, G. A. et al. Epidemiology and outcomes of arterial ischemic stroke in children: the Canadian Pediatric Ischemic Stroke Registry. *Pediatr. Neurol.* **69**, 58–70 (2017).
5. Galvin, J., Hewish, S., Rice, J. & Mackay, M. T. Functional outcome following paediatric stroke. *Dev. Neurorehabil.* **14**, 67–71 (2011).
6. Lo, W., Stephens, J. & Fernandez, S. Pediatric stroke in the United States and the impact of risk factors. *J. Child Neurol.* **24**, 194–203 (2009).
7. Anderson, V. et al. Social competence following pediatric stroke: contributions of brain insult and family environment. *Soc. Neurosci.* **9**, 471–483 (2014).
8. Kirton, A. & deVeber, G. Paediatric stroke: pressing issues and promising directions. *Lancet Neurol.* **14**, 92–102 (2015).
9. World Health Organization. *International Classification of Functioning, Disability and Health: Children and Youth Version: ICF-CY* (World Health Organization, 2007).
10. Kitchen, L. et al. The pediatric stroke outcome measure: a validation and reliability study. *Stroke* **43**, 1602–1608 (2012).
11. Gordon, A. L. Functioning and disability after stroke in children: using the ICF-CY to classify health outcome and inform future clinical research priorities. *Dev. Med. Child Neurol.* **56**, 434–444 (2014).
12. Engelmann, K. A. & Jordan, L. C. Outcome measures used in pediatric stroke studies: a systematic review. *Arch. Neurol.* **69**, 23–27 (2012).
13. Bulder, M. M. et al. Measuring outcome after arterial ischemic stroke in childhood with two different instruments. *Cerebrovasc. Dis.* **32**, 463–470 (2011).
14. deVeber, G. A., MacGregor, D., Curtis, R. & Mayank, S. Neurologic outcome in survivors of childhood arterial ischemic stroke and sinovenous thrombosis. *J. Child Neurol.* **15**, 316–324 (2000).
15. Neuner, B. et al. Health-related quality of life in children and adolescents with stroke, self-reports, and parent/proxies reports: cross-sectional investigation. *Ann. Neurol.* **70**, 70–78 (2011).
16. Vazquez Lopez, M. et al. Outcome of arterial ischemic stroke in children with heart disease. *Eur. J. Paediatr. Neurol.* **21**, 730–737 (2017).
17. Muthen, B. O. in *New Developments and Techniques in Structural Equation Modeling* (eds Marcoulides, G. A. & Schumacker, R. E.) 1–33 (Lawrence Erlbaum, Hillsdale, 2001).
18. Pepe, M. S. & Janes, H. Insights into latent class analysis of diagnostic test performance. *Biostatistics* **8**, 474–484 (2007).
19. Kim, Y., Barreira, T. V. & Kang, M. Concurrent associations of physical activity and screen-based sedentary behavior on obesity among US Adolescents: a latent class analysis. *J. Epidemiol.* **26**, 137–144 (2016).
20. Evans-Polce, R., Lanza, S. & Maggs, J. Heterogeneity of alcohol, tobacco, and other substance use behaviors in U.S. college students: a latent class analysis. *Addict. Behav.* **53**, 80–85 (2016).
21. Fitzpatrick, K. E., Gray, R. & Quigley, M. A. Women's longitudinal patterns of smoking during the pre-conception, pregnancy and postnatal period: evidence from the UK infant feeding survey. *PLoS ONE* **11**, e0153447 (2016).
22. Fiume, A. et al. Development and validation of the Pediatric Stroke Quality of Life Measure. *Dev. Med. Child Neurol.* **60**, 587–595 (2018).
23. Fox, C. K. et al. Prolonged or recurrent acute seizures after pediatric arterial ischemic stroke are associated with increasing epilepsy risk. *Dev. Med. Child Neurol.* **59**, 38–44 (2017).
24. Lo, W. D. et al. The Pediatric Stroke Recurrence and Recovery Questionnaire: validation in a prospective cohort. *Neurology* **79**, 864–870 (2012).
25. Haley, S. M., New England Medical Center H, Group PR. *Pediatric Evaluation of Disability Inventory (PEDI): Development, Standardization and Administration Mmanual* (New England Medical Center Hospital, PEDI Research Group, Boston, 1992).
26. Varni, J. W., Burwinkle, T. M., Seid, M. & Skarr, D. The PedsQL 4.0 as a pediatric population health measure: feasibility, reliability, and validity. *Ambul. Pediatr.* **3**, 329–341 (2003).
27. Crouchman, M., Rossiter, L., Colaco, T. & Forsyth, R. A practical outcome scale for paediatric head injury. *Arch. Dis. Child* **84**, 120–124 (2001).
28. Bigi, S. et al. Acute ischemic stroke in children versus young adults. *Ann. Neurol.* **70**, 245–254 (2011).
29. Nylund, K. L., Asparouhov, T. & Muthen, B. O. Deciding on the number of classes in latent class analysis and growth mixture modeling: a Monte Carlo simulation study. *Struct. Equ. Model.* **14**, 535–569 (2007).
30. Gwet, K. L. Computing inter-rater reliability and its variance in the presence of high agreement. *Br. J. Math. Stat. Psychol.* **61**, 29–48 (2008).
31. Wongpakaran, N., Wongpakaran, T., Wedding, D. & Gwet, K. L. A comparison of Cohen's Kappa and Gwet's AC1 when calculating inter-rater reliability coefficients: a study conducted with personality disorder samples. *BMC Med. Res. Methodol.* **13**, 61 (2013).

32. Douglas, C. E. & Michael, F. A. On distribution-free multiple comparisons in the one-way analysis of variance. *Commun. Stat. Theory Methods* **20**, 127–139 (1991).
33. Landis, J. R. & Koch, G. G. The measurement of observer agreement for categorical data. *Biometrics* **33**, 159–174 (1977).
34. Rivkin, M. J. et al. Thrombolysis in pediatric stroke study. *Stroke* **46**, 880–885 (2015).
35. Friefeld, S., Yeboah, O., Jones, J. E. & deVeber, G. Health-related quality of life and its relationship to neurological outcome in child survivors of stroke. *CNS Spectr.* **9**, 465–475 (2004).
36. Friefeld, S. J., Westmacott, R., Macgregor, D. & Deveber, G. A. Predictors of quality of life in pediatric survivors of arterial ischemic stroke and cerebral sinovenous thrombosis. *J. Child Neurol.* **26**, 1186–1192 (2011).