



Published in final edited form as:

Stroke. 2012 February ; 43(2): 341–345. doi:10.1161/STROKEAHA.111.633305.

Concurrent Validity and Reliability of Retrospective Scoring of the Pediatric NIH Stroke Scale

Lauren A. Beslow, MD MSCE¹, Scott E. Kasner, MD MSCE², Sabrina E. Smith, MD PhD¹, Michael T. Mullen, MD², Matthew P. Kirschen, MD PhD¹, Rachel A. Bastian, BA¹, Michael M. Dowling, MD PhD MSCS³, Warren Lo, MD⁴, Lori C. Jordan, MD PhD⁵, Timothy J. Bernard, MD⁶, Neil Friedman, MBChB⁷, Gabrielle deVeber, MD⁸, Adam Kirton, MD⁹, Lisa Abraham, MD¹⁰, Daniel J. Licht, MD¹, Abbas F. Jawad, PhD¹¹, Jonas H. Ellenberg, PhD¹², Ebbing Lautenbach, MD MPH MSCE^{13,§}, and Rebecca N. Ichord, MD^{1,§}

¹Division of Neurology, Children's Hospital of Philadelphia, University Pennsylvania School of Medicine, Colket Translational Research Building, 10th Floor, 3501 Civic Center Boulevard, Philadelphia, PA 19104

²Department of Neurology, The Hospital of the University of Pennsylvania, University of Pennsylvania School of Medicine, 3400 Spruce Street, Philadelphia, PA 19104

³Departments of Pediatrics and Neurology, Children's Medical Center Dallas, University of Texas Southwestern Medical Center, 5323 Harry Hines Blvd, Dallas, TX 75390-9063

⁴Department of Neurology, Nationwide Children's Hosp, Ohio State University, EDU 533, 700 Children's Drive, Columbus, OH 43205-2664

⁵Department of Neurology, Johns Hopkins University, 200 N Wolfe Street, Suite 2158, Baltimore, MD 21287

⁶Section of Child Neurology, Denver Children's Hospital, University of Colorado, P.O. Box 6507, Mail Stop F-416, Aurora, CO 80045-0507

⁷Center for Pediatric Neurology/Desh S60, Neurological Institute, Cleveland Clinic, 9500 Euclid Avenue, Cleveland, OH 44195

⁸Department of Pediatrics, Division of Neurology, Hospital for Sick Children, 555 University Ave, Toronto, ON, Canada M5G1X8

⁹Department of Pediatrics and Clinical Neurosciences, Alberta Children's Hospital, University of Calgary, 2888 Shaganappi Trail NW, Calgary, AB, Canada T3B6A8

¹⁰Department of Pediatrics, Schenectady Neurological Consultants, PC, 1401 Union Street, Schenectady, NY 12308

Corresponding author: Lauren A. Beslow MD MSCE, Division of Neurology, Children's Hospital of Philadelphia, Colket Translational Research Building, 10th Floor, 3501 Civic Center Boulevard, Philadelphia, PA 19104, Phone 215-590-4142, Fax 215-590-2950, beslow@email.chop.edu.

[§]Ebbing Lautenbach and Rebecca N. Ichord are co-last authors.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Disclosures

R.N. Ichord: Consultant/Advisory Board, Berlin Heart Clinical Event Committee.

L.C. Jordan: Consultant/Advisory Board, Berlin Heart Clinical Event Committee.

None: L.A. Beslow, S.E. Kasner, S.E. Smith, M.T. Mullen, M.P. Kirschen, R.A. Bastian, M.M. Dowling, W. Lo, T.J. Bernard, N. Friedman, G. deVeber, A. Kirton, L. Abraham, D.J. Licht, A.F. Jawad, J.H. Ellenberg, E. Lautenbach

¹¹Department of Pediatrics, Children's Hospital of Philadelphia, University of Pennsylvania, 3535 Market Street, Philadelphia, PA 19104

¹²Department of Biostatistics and Epidemiology, Center for Clinical Epidemiology and Biostatistics, University of Pennsylvania School of Medicine, 825 Blockley Hall, 423 Guardian Drive, Philadelphia, PA 19104-6021

¹³Division of Infectious Diseases, Department of Medicine, Department of Biostatistics and Epidemiology, Center for Clinical Epidemiology and Biostatistics, University of Pennsylvania School of Medicine, 825 Blockley Hall, 423 Guardian Drive, Philadelphia, PA 19104-6021

Abstract

Background and Purpose—The Pediatric National Institutes of Health Stroke Scale (PedNIHSS), an adaptation of the adult NIH Stroke Scale, is a quantitative measure of stroke severity shown to be reliable when scored prospectively. The ability to calculate the PedNIHSS score retrospectively would be invaluable in the conduct of observational pediatric stroke studies. The study objective was to assess the concurrent validity and reliability of estimating the PedNIHSS score retrospectively from medical records.

Methods—Neurological examinations from medical records of 75 children enrolled in a prospective PedNIHSS validation study were photocopied. Four neurologists of varying training levels blinded to the prospective PedNIHSS scores reviewed the records and retrospectively assigned PedNIHSS scores. Retrospective scores were compared among raters and to the prospective scores.

Results—Total retrospective PedNIHSS scores correlated highly with total prospective scores ($R^2=0.76$). Interrater reliability for the total scores was “excellent” (intraclass correlation coefficient of 0.95, 95% confidence interval 0.94–0.97). Interrater reliability for individual test items was “substantial” or “excellent” for 14 of 15 items.

Conclusions—The PedNIHSS score can be scored retrospectively from medical records with a high degree of concurrent validity and reliability. This tool can be used to improve the quality of retrospective pediatric stroke studies.

Keywords

arterial ischemic stroke; pediatric; NIH stroke scale

Introduction

The National Institutes of Health Stroke Scale (NIHSS), a reliable quantitative measure of acute stroke severity in adults¹, predicts stroke outcome at 7 and 90 days². The adult NIHSS is the primary examination used for adult stroke research and acute treatment trials. This scale has enabled multicenter studies since the clinical stroke severity of various patients can be compared easily, even by practitioners other than neurologists³. Furthermore, the reliability and concurrent validity of estimating the adult NIHSS score from medical records retrospectively was first demonstrated using 39 subjects⁴ and has been replicated^{5–8}. These studies have been extremely useful, allowing researchers to quantify stroke severity when a prospective NIHSS score was not recorded^{9, 10}.

Children were not included in the initial validation study of the NIHSS. A study to determine the validity and utility of a pediatric adaptation of the NIHSS, the Pediatric NIH Stroke Scale (PedNIHSS), closed its enrollment in 2009. Modification of the NIHSS for pediatric use is crucial since the neurological examination of children varies significantly by age and development. The PedNIHSS has the same examination elements as the adult scale

including 11 neurological domains and 15 scored items. The PedNIHSS (for children age 2–18 years) adapts the tasks the child performs so that they are appropriate for the child's age and development. The total score for the PedNIHSS ranges from 0 (least severe) to 42 (most severe)¹¹.

The PedNIHSS validation study was prospective with scores determined during the hospitalization for the acute stroke¹¹. Many studies on pediatric stroke, including population-based studies from which stroke incidences in developed countries have been estimated, are retrospective. Retrospective pediatric stroke studies are limited because comparison of children's clinical stroke severity across centers and even within centers cannot be done quantitatively. The ability to assess the PedNIHSS score retrospectively would allow investigators to use this quantitative measure even when a study design is retrospective or when the PedNIHSS score is missing during prospective studies, thereby improving the validity of study findings. However, the concurrent validity and reliability of calculating the PedNIHSS score retrospectively using medical records are untested.

Materials and Methods

Study design

This was a cross-sectional study comparing PedNIHSS scores derived retrospectively from neurological examinations documented in the medical record to PedNIHSS scores assigned prospectively in the parent study¹¹.

Inclusion criteria

All sites participating in the parent study were invited to participate in the current study, and all subjects enrolled in the parent study with a neurological examination documented in the medical record were eligible for the current study. The methods of the parent study are described elsewhere, and consent from the parent or legal guardian of the child was obtained¹¹. Enrollment in the parent study required definite arterial ischemic stroke based on criteria for the International Pediatric Stroke Study (IPSS). The criteria consist of acute neurological deficit and corresponding imaging findings of focal infarct conforming to an established arterial territory¹². This study was approved by the local IRB at each site.

Study population and sites

The population for the current study was comprised of 75 children enrolled from the following 9 sites: Children's Medical Center Dallas (Dallas, Texas, 26), The Children's Hospital of Philadelphia (Philadelphia, Pennsylvania, 16), The Hospital for Sick Children (Toronto, Ontario, 13), The Cleveland Clinic Children's Hospital (Cleveland, Ohio, 5), Nationwide Children's Hospital (Columbus, Ohio, 4), The Children's Hospital (Denver, Colorado, 4), The Children's Hospital of Pittsburgh (Pittsburgh, Pennsylvania, 4), Alberta Children's Hospital (Calgary, Alberta, 2), and The Johns Hopkins Children's Center (Baltimore, Maryland, 1). All sites are tertiary care centers with established pediatric stroke programs affiliated with universities. All prospective PedNIHSS scores were performed by a study pediatric stroke neurology attending physician or stroke fellow. The Children's Hospital of Philadelphia (CHOP) was the coordinating center.

Study procedure

Medical record preparation—A study coordinator at each site photocopied the first neurological examination documented by a neurologist or neurologist-in-training. Of the 75 subjects in this study, 36 neurological examinations documented in the medical record (48%) were performed by a general pediatric neurologist who was not the stroke neurologist

performing the prospective PedNIHSS score, and 39 (52%) were performed by the pediatric stroke neurologist who also performed the prospective PedNIHSS. Of the 36 subjects in whom the neurological examination was performed by a general neurologist and not the stroke neurologist who performed the prospective PedNIHSS score, 22 (29% of all subjects) were performed before the subject was enrolled in the parent study. Identifying patient information was removed at the enrolling site. The CHOP study coordinator (R.A.B.) read through each neurological examination, removed any references to the prospectively assigned scores, and photocopied the examination for each of the four raters. A random number generator was used to order the subject neurological examinations, so each rater scored the 75 examinations in the same order to limit variability for each subject due to rater experience. The initial PedNIHSS score from the parent study served as the reference criterion for each subject in the concurrent validity analysis.

Rater training—Four raters from CHOP and The University of Pennsylvania Neurology Departments with varying levels of training were blinded to the prospectively assigned PedNIHSS scores. The raters included a pediatric stroke attending (S.E.S.), an adult stroke attending (M.T.M.), a pediatric stroke fellow (L.A.B.), and a child neurology resident (M.P.K.). The raters attended a training session with the principal investigator of the parent study (R.N.I.). Detailed instructions on how to translate information from the medical record into scores were provided at the training session. If an item was not recorded in the medical record, it was scored as 0 (normal) as was done in past adult studies (see online appendix)⁵. Every rater recorded whether each item was documented. Study data were entered into a database by L.A.B. and were reviewed for accuracy by the CHOP study coordinator (R.A.B.).

Statistical Analyses

Descriptive statistics were performed using frequency distributions and proportions for categorical variables and means with standard deviations or medians with interquartile ranges for continuous variables. Linear regression was performed to determine the correlation of the estimated PedNIHSS scores with the prospectively assigned scores, clustered by subject since the four estimates obtained by the four raters for each subject were not independent. The parameter R^2 was reported as the measure of concurrent validity. Age at time of the stroke and the time difference between the prospective scores and neurological examination recorded in the medical record were examined as covariates. Linear regression analysis was also done for each rater. A sub-analysis was performed to assess the correlation of the estimated PedNIHSS scores with the prospectively assigned scores for the 22 subjects in whom the neurological examination was documented by a non-stroke neurologist before the subject was enrolled in the parent study. Using a pre-determined threshold based on categorizations in the original adult retrospective NIHSS scoring study⁴, the sensitivity and specificity for correctly scoring a total PedNIHSS that had been prospectively scored as ≤ 5 were calculated with their 95% confidence intervals using MedCalc Version 11.5.1 (MedCalc Software, Mariakerke, Belgium). Since the total NIHSS score is generally used as a continuous measure, interrater reliability of the estimated total scores was determined by the calculation of an intraclass correlation coefficient (ICC) using one way analysis of variance. A sub-analysis was performed to assess the reliability of the four raters' retrospective scores for the 22 subjects in whom the neurological examination was documented by a non-stroke neurologist before the subject was enrolled in the parent study. Sub-analyses were performed on the 15 items' scores using a weighted kappa (κ) statistic since the item scores are categorical using SAS 9.2 macro (SAS Institute Inc., Cary, NC). A κ or ICC is considered moderate agreement if 0.41–0.60, substantial agreement if 0.61–0.80, and almost perfect (excellent) if 0.81–1.00¹³. A two-sided probability value of <0.05 was

considered significant. STATA version 11.1 (STATA Corporation, College Station, TX) was used for regression analysis and ICC analysis.

Results

Subject characteristics

Nine of 15 centers (60%) participating in the parent prospective study agreed to enroll their subjects in the current study. All subjects from the 9 centers were enrolled in the current study with the exception of one subject for whom the neurological examination in the medical record could not be located. Therefore, 75 of 113 children from the parent study (66.4%) were included. Forty-five (60%) subjects were male. Racial distribution was 56 white (8 Hispanic), 11 black or African American, 6 Asian, 1 mixed race, and 1 of unknown race (Hispanic). The mean and median ages of the participants were 9.7 years (standard deviation [SD] 5.5 years) and 9.2 years (interquartile range [IQR] 4.5–15.6 years), respectively. There were no significant differences between the 75 subjects enrolled in the study and the 38 subjects who were not enrolled on the basis of median prospective PedNIHSS total score, age at presentation, or sex (data not shown).

Concurrent validity of retrospective scores with prospective scores

The mean and median total prospective and retrospective PedNIHSS scores were not different for the 75 subjects ($p=0.49$ Student's *t*-test; $p=0.37$ Wilcoxon rank-sum). The mean prospective total PedNIHSS score for the 75 participating children was 8.2 (SD 7) with median 6 (IQR 3–12), and the mean total retrospective PedNIHSS score was 7.6 (SD 7) with median 5 (IQR 3–11). In regression analysis, R^2 for the retrospective estimations of the total scores and the prospectively assigned total scores was 0.76, slope 0.87 (95% CI 0.77–0.97), intercept 1.62, and $p<0.001$. The mean and median time difference between the performance of the prospective PedNIHSS score and the neurological examination documented in the medical record were 6.8 hours (SD 12.8 hours) and 1 hour (IQR 0.25–9 hours), respectively. Neither age at time of the stroke nor time difference between the prospective score and the neurological examination recorded in the medical record altered the observed R^2 (not shown). The R^2 for each of the 4 raters' retrospectively estimated total scores versus the prospectively assigned total scores ranged from 0.73–0.83. Figure 1 demonstrates a scatter plot of the prospective total PedNIHSS scores versus the retrospective total PedNIHSS scores by rater. Two hundred sixty-eight of 300 retrospective total scores (89%) were within 5 points of the prospectively scored totals. The sensitivity and specificity for the raters correctly scoring a total PedNIHSS that had been prospectively scored as ≤ 5 were 87% (95% confidence interval [CI] 81–92%) and 81% (95% CI 74–87%), respectively. In the sub-analysis on the 22 subjects whose neurological examination in the chart was performed prior to parent study enrollment and in whom the neurological examination was performed by a general pediatric neurologist who was not the pediatric stroke neurologist who performed the prospective PedNIHSS, the R^2 was 0.85, slope 0.88 (95% CI 0.77–1.00), intercept 1.44, and $p<0.001$.

Interrater reliability of retrospective scores

For the total score, the ICC was 0.95 (95% exact CI 0.94–0.99). Figure 2 demonstrates box plots of the distribution of the retrospectively estimated total PedNIHSS scores for each rater, indicating that the four raters' total scores were comparable. The weighted κ for the item scores ranged from 0.47 to 0.93 (Table 1). The ICC was 0.95 (95% exact CI 0.90–0.98) in the sub-analysis on the 22 patients in whom the neurological examination was documented prior to parent study enrollment by a non-stroke neurologist different from the pediatric stroke neurologist who performed the prospective PedNIHSS.

Discussion

Retrospective studies have been critical in the field of pediatric stroke, but a limitation has been the inability to measure and then adjust for initial clinical stroke severity. The results of this study demonstrated that the PedNIHSS score can be assessed retrospectively from neurological examinations found in the medical record with excellent concurrent validity and interrater reliability. Our ICC for total PedNIHSS score among the four raters was 0.95, which compares favorably to the ICC of 0.82 previously reported for retrospective assessment of the NIHSS in adults⁴ and to the ICC of 0.99 obtained in the prospective pediatric study¹¹. Compared to the weighted κ for the item scores from the prospective pediatric study, our weighted κ for all items were similar except for the visual item (3)¹¹. There was 100% agreement on the visual performance item in the prospective study whereas in the retrospective study, this item had poorest agreement ($\kappa = 0.47$). Upon inspection of the 4 raters' scores for this item, there were 6 instances in which 1 rater recorded a 3 (bilateral hemianopia) while the other raters recorded 0 (normal). These were subjects who had severe strokes with poor mental status. This scenario was perhaps more difficult to score since the directions for the prospective PedNIHSS score indicate that a comatose subject should be scored as a 3 for visual fields while our algorithm indicated that items not clearly documented should be scored as 0¹¹. While the reliability for the visual item was only moderate, the ICC for the overall score was excellent. We therefore would not alter the retrospective scoring method to improve the interrater reliability of this single item.

The current study has several potential limitations. While not all subjects in the parent study were enrolled in this ancillary study, subjects who were enrolled did not differ from those not enrolled with respect to age, sex, and prospective PedNIHSS total score. All study subjects presented to tertiary care hospitals and were enrolled in a single clinical trial in which the PedNIHSS was tested as a research tool. It is not possible to determine if documentation practices in the medical record were altered based on enrollment. Some neurologists may have documented examinations either more or less carefully due to subject participation in the parent study. Notably, the PedNIHSS score form for the study was not a part of the medical record and was not documented in the medical record in most cases. Nonetheless, if participation in the study increased the degree of documentation in the medical chart compared to documentation practices in patients not enrolled in the study, our reliability estimates or the concurrent validity estimate may have been increased. However, our excellent reliability estimate (ICC=0.95) and concurrent validity estimate ($R^2=0.85$) from the sub-analysis on 22 subjects whose neurological examination was documented prior to enrollment in the parent study and by a general pediatric neurologist who was not the pediatric stroke neurologist who performed the prospective PedNIHSS score suggests that documentation practices in the chart may largely have been unchanged. Furthermore, the fact that the neurological examination corresponding to three different PedNIHSS score items was not documented in the medical records of greater than 15% of the subjects suggests that documentation practices in the medical record may not have been altered by subject participation in the parent study. Most pediatric strokes are mild to moderate (median prospective total score 6), and only 7 subjects had prospective total PedNIHSS scores ≥ 20 . Therefore, extrapolation of the concurrent validity and reliability of the retrospective scoring method to scores in the upper range for the total PedNIHSS score may need additional study. Furthermore, the study findings cannot be extrapolated to raters who are non-neurologist physicians, nurses, and research coordinators or to neurological examinations documented by non-neurologists. Replication of the study involving such raters who have other training, in patients whose neurological examinations were documented by practitioners other than neurologists, and in patients who are not enrolled in the prospective PedNIHSS validation study would be useful in order to expand the use of the retrospective estimation of the PedNIHSS score. Nevertheless, our study results are

strengthened by including raters at various stages in their neurology training and by our inclusion of subjects from 9 geographically diverse centers. Both of these factors increase the generalizability of our results. Despite possible limitations, our results demonstrate that the PedNIHSS score can be scored retrospectively from medical records.

Conclusions

Stroke in children is uncommon^{14, 15}, with diverse and often multiple etiologies¹⁶, and the predictors of outcome are poorly understood. Although we hope that the Pediatric NIH Stroke Scale score will become part of the routine examination for childhood stroke patients, scoring the PedNIHSS from medical records retrospectively is valid and reliable. The ability to assess and control for initial stroke severity will facilitate potential retrospective studies and even prospective studies with missing data on PedNIHSS score that could provide essential information about pathophysiology, clinical course, and outcome of pediatric stroke.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The manuscript is the Master of Science in Clinical Epidemiology thesis for L.A. Beslow from The Center for Clinical Epidemiology and Biostatistics at the University of Pennsylvania.

Funding sources

Lauren Beslow: NIH-T32-NS007413, L. Morton Morley Funds of The Philadelphia Foundation

Sabrina Smith: NIH-K12-NS049453

Michael Dowling: NIH-KL2-RR024983, The First American Real Estate Services, Inc., Doris Duke Charitable Foundation

Lori Jordan: NIH-K23-NS062110

Timothy Bernard: NIH-K23-HL096895

Gabrielle deVeber: NIH-R01-NS062820

Adam Kirton: Heart/Stroke Foundation Canada, Heart/Stroke Foundation Alberta, Alberta Children's Hospital Foundation and Research Institute, Hotchkiss Brain Institute

Daniel Licht: NIH-K23-NS52380, Dana Foundation, June and Steve Wolfson Family Fund for Neurological Research

Abbas Jawad: NIH-R01-NS050488

Ebbing Lautenbach: NIH-K24-AI080942

Rebecca Ichord: NIH-R01-NS050488, K23-NS062110

References

1. Brott T, Adams HP Jr, Olinger CP, Marler JR, Barsan WG, Biller J, et al. Measurements of acute cerebral infarction: A clinical examination scale. *Stroke*. 1989; 20:864–870. [PubMed: 2749846]
2. Adams HP Jr, Davis PH, Leira EC, Chang KC, Bendixen BH, Clarke WR, et al. Baseline nih stroke scale score strongly predicts outcome after stroke: A report of the trial of org 10172 in acute stroke treatment (toast). *Neurology*. 1999; 53:126–131. [PubMed: 10408548]

3. Goldstein LB, Samsa GP. Reliability of the national institutes of health stroke scale. Extension to non-neurologists in the context of a clinical trial. *Stroke*. 1997; 28:307–310. [PubMed: 9040680]
4. Kasner SE, Chalela JA, Luciano JM, Cucchiara BL, Raps EC, McGarvey ML, et al. Reliability and validity of estimating the nih stroke scale score from medical records. *Stroke*. 1999; 30:1534–1537. [PubMed: 10436096]
5. Williams LS, Yilmaz EY, Lopez-Yunez AM. Retrospective assessment of initial stroke severity with the nih stroke scale. *Stroke*. 2000; 31:858–862. [PubMed: 10753988]
6. Barber M, Fail M, Shields M, Stott DJ, Langhorne P. Validity and reliability of estimating the scandinavian stroke scale score from medical records. *Cerebrovasc Dis*. 2004; 17:224–227. [PubMed: 14707426]
7. Stavem K, Lossius M, Ronning OM. Reliability and validity of the canadian neurological scale in retrospective assessment of initial stroke severity. *Cerebrovasc Dis*. 2003; 16:286–291. [PubMed: 12865618]
8. Bushnell CD, Johnston DC, Goldstein LB. Retrospective assessment of initial stroke severity: Comparison of the nih stroke scale and the canadian neurological scale. *Stroke*. 2001; 32:656–660. [PubMed: 11239183]
9. Chitravas N, Dewey HM, Nicol MB, Harding DL, Pearce DC, Thrift AG. Is prestroke use of angiotensin-converting enzyme inhibitors associated with better outcome? *Neurology*. 2007; 68:1687–1693. [PubMed: 17502550]
10. Hsia AW, Sachdev HS, Tomlinson J, Hamilton SA, Tong DC. Efficacy of iv tissue plasminogen activator in acute stroke: Does stroke subtype really matter? *Neurology*. 2003; 61:71–75. [PubMed: 12847159]
11. Ichord RN, Bastian R, Abraham L, Askalan R, Benedict S, Bernard TJ, et al. Interrater reliability of the pediatric national institutes of health stroke scale (pednihss) in a multicenter study. *Stroke*. 2004; 35:613–617. [PubMed: 21317270]
12. Sebire G, Fullerton H, Riou E, deVeber G. Toward the definition of cerebral arteriopathies of childhood. *Curr Opin Pediatr*. 2004; 16:617–622. [PubMed: 15548922]
13. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics*. 1977; 33:159–174. [PubMed: 843571]
14. Fullerton HJ, Wu YW, Zhao S, Johnston SC. Risk of stroke in children: Ethnic and gender disparities. *Neurology*. 2003; 61:189–194. [PubMed: 12874397]
15. Giroud M, Lemesle M, Gouyon JB, Nivelon JL, Milan C, Dumas R. 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*. 1995; 48:1343–1348. [PubMed: 7490597]
16. Roach ES, Golomb MR, Adams R, Biller J, Daniels S, Deveber G, et al. Management of stroke in infants and children: A scientific statement from a special writing group of the american heart association stroke council and the council on cardiovascular disease in the young. *Stroke*. 2008; 39:2644–2691. [PubMed: 18635845]

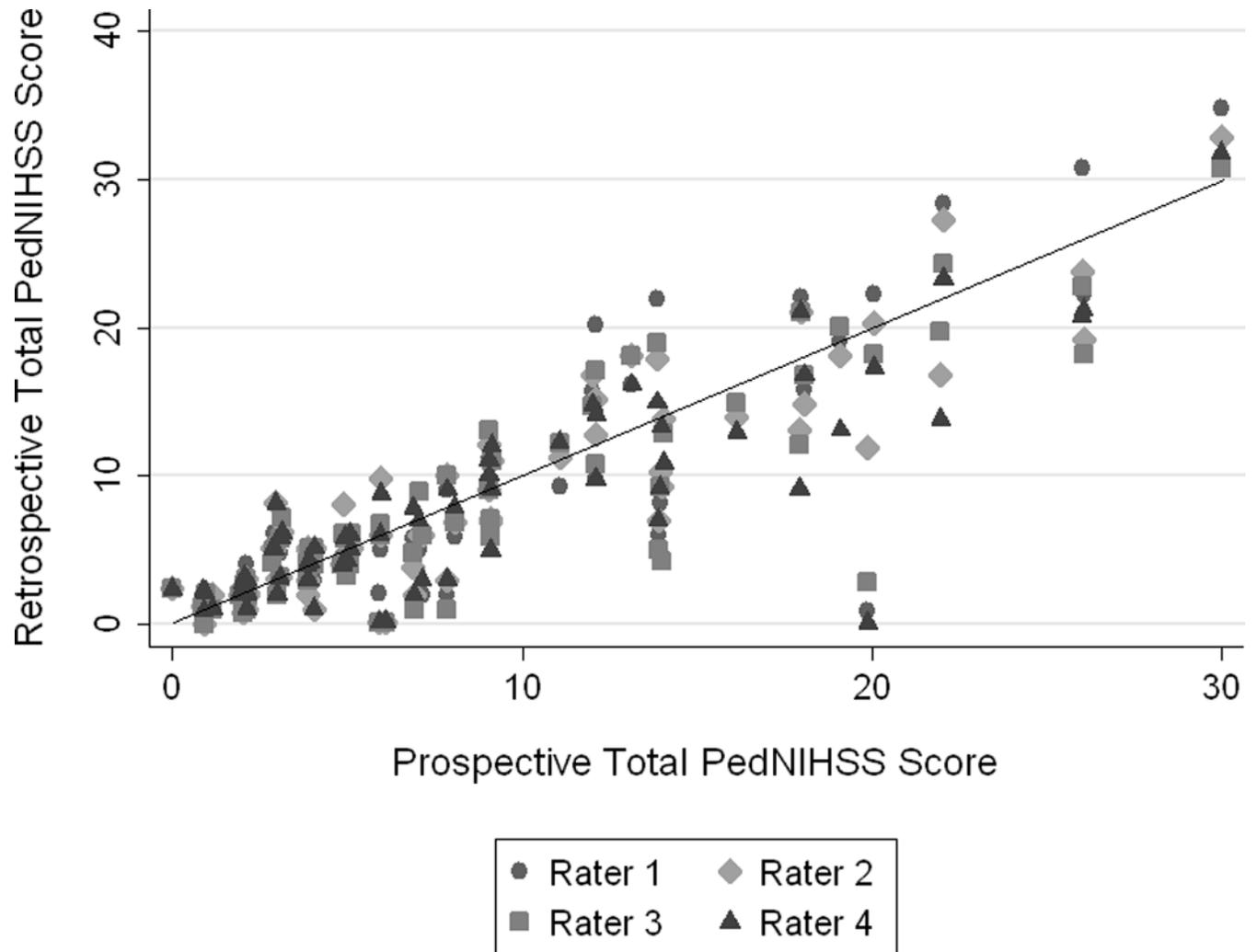


Figure 1. Scatter plot of prospective total Pediatric NIH Stroke Scale (PedNIHSS) score versus retrospective total PedNIHSS score for the four raters. Line represents reference with slope of 1.

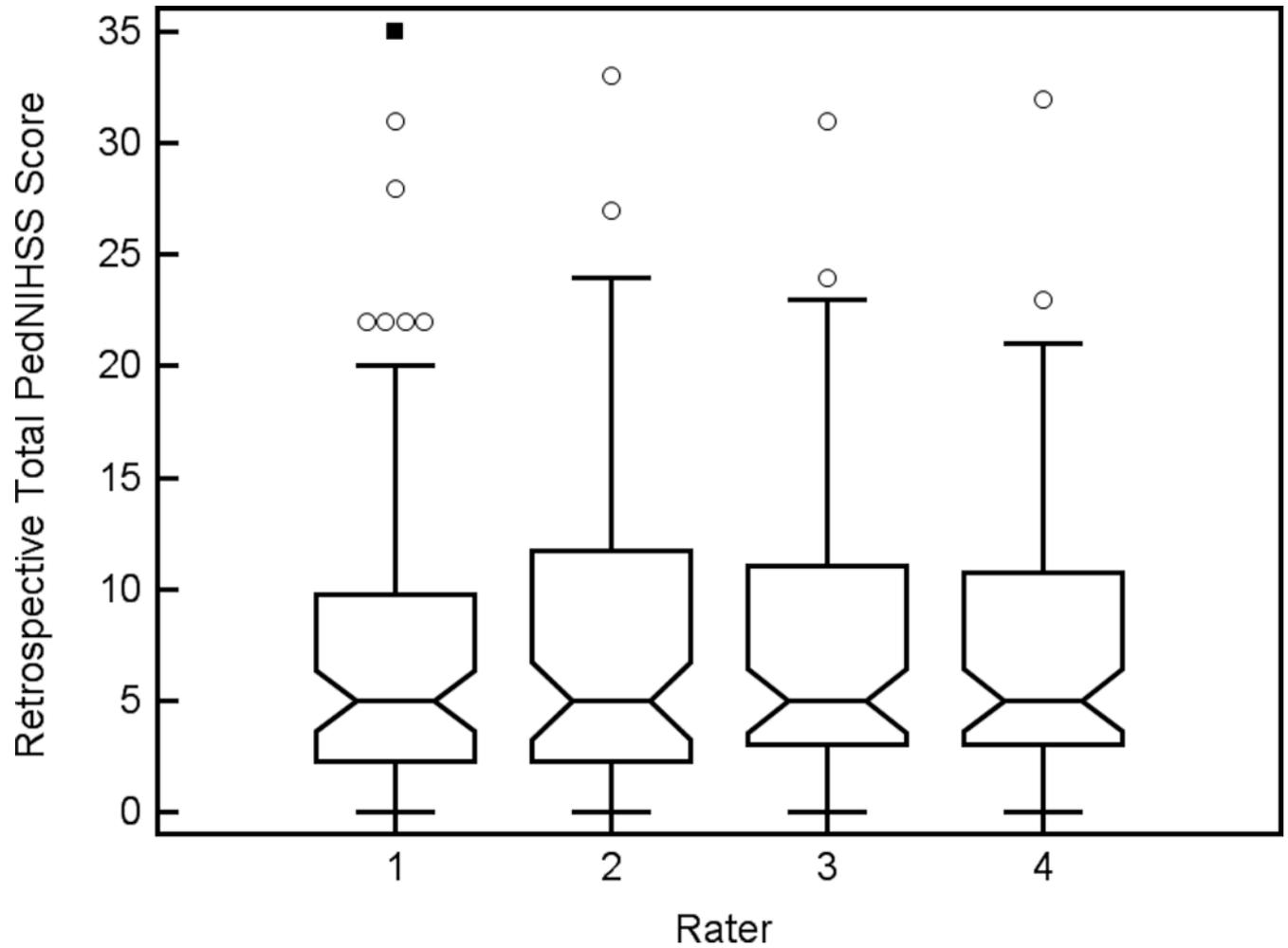


Figure 2.

Box plots of distributions of retrospective total PedNIHSS scores for each rater. Lower and upper boundaries represent the 25th and 75th percentiles. The central line represents the median. The circles represent outliers, defined as those scores farther away from the box than 1.5 times the interquartile range. The square is an extreme outlier, defined as a score farther away from the box than 3 times the interquartile range. Whiskers represent the largest and smallest non-outlying scores.

Table 1

Agreement among four raters by individual retrospectively scored PedNIHSS item

Item	Weighted κ	95% CI for κ	N (%) item undocumented*
1a - LOC	0.82	0.76–0.87	1 (1.3%)
1b - LOC Questions	0.77	0.69–0.83	3 (4%)
1c - LOC Commands	0.71	0.62–0.79	6 (8%)
2 - Best Gaze	0.83	0.78–0.88	6 (8%)
3 - Visual	0.47	0.35–0.59	18 (24%)
4 - Facial Palsy	0.84	0.79–0.89	2 (2.7%)
5a - Motor Leg Left	0.93	0.90–0.95	2 (2.7%)
5b - Motor Leg Right	0.93	0.91–0.95	0 (0%)
6a - Motor Arm Left	0.89	0.85–0.92	3 (4%)
6b - Motor Arm Right	0.91	0.87–0.94	0 (0%)
7 - Limb Ataxia	0.87	0.83–0.91	11 (14.7%)
8 - Sensory	0.72	0.63–0.79	8 (10.7%)
9 - Best Language	0.83	0.77–0.88	9 (12%)
10 - Dysarthria	0.86	0.80–0.90	14 (18.7%)
11 - Extinction/Inattention	0.77	0.69–0.83	31 (41.3%)

κ , weighted kappa; CI, confidence interval; N, number;

* Number of charts in which item not documented (total charts = 75); LOC, level of consciousness