Comparison Between the Original and Shortened Versions of the National Institutes of Health Stroke Scale in Ischemic Stroke Patients of Intermediate Severity

Chun Fan Lee, PhD; Narayanaswamy Venketasubramanian, FRCP;
K.S. Lawrence Wong, MD; Christopher L.H. Chen, FRCP; for the CHIMES Study Investigators

Background and Purpose—The 15-item National Institutes of Health Stroke Scale (NIHSS) has been critiqued for its complexity and variability, and shortened versions have been proposed. This study aimed to compare the measurement properties of the original version with 3 shortened versions with 11, 8, and 5 items, respectively.

Methods—Analyses were performed using data from an international, double-blind randomized controlled trial investigating the efficacy of MLC601 on stroke recovery in patients with ischemic stroke of intermediate severity (Chinese Medicine Neuroaid Efficacy on Stroke recovery [CHIMES]). To compare discriminative ability and responsiveness to change, the effect sizes of the NIHSS scores in relation to modified Rankin Scale, mini-mental status examination, and Barthel index were estimated using regression analysis.

Results—For both discriminative ability and responsiveness to change, the original version exhibited a larger effect size (0.55 and 0.84) in relation to modified Rankin Scale than the other 3 shortened versions (0.35–0.46 and 0.74–0.78).

Conclusions—The original 15-item NIHSS retained information that made it more discriminative and responsive to change than the shortened versions. We recommend future clinical researchers to use the full version NIHSS to evaluate patients' stroke severity.

Clinical Trial Registration Information—URL: http://www.clinicaltrials.gov. Unique identifier: NCT00554723.
   (***Stroke***. 2016;47:236-239. DOI: 10.1161/STROKEAHA.115.011657.)

   Key Words: Barthel index ■ MLC601 ■ modified Rankin scale ■ National Institute of Health Stroke Scale ■ NeuroAiD ■ outcome ■ stroke severity

Methods

Subjects and Assessments

Analyses were performed using data from the Chinese Medicine Neuroaid Efficacy on Stroke recovery (CHIMES) study, an international, randomized, placebo-controlled, double-blind trial investigating the efficacy of MLC601 on stroke recovery in patients with ischemic stroke of intermediate severity in the preceding 72 hours (clinicaltrials.gov NCT00554723).[10,11] Patients were randomly allocated to either MLC601 or placebo for 3 months as add-on to standard stroke care.

Patients were assessed at baseline, day 10 (±2 days), or discharge, whichever was earlier, and month 3 (±1 week). At baseline, demographic and clinical information, as well as prestroke modified Rankin Scale (mRS) score, were ascertained. NIHSS and mini-mental status examination (MMSE) were also performed. At day 10 (or discharge, if earlier) and month 3 visits, NIHSS, mRS, and MMSE were assessed again. In addition, month 3 assessment included the Barthel index (BI).

The National Institutes of Health Stroke Scale (NIHSS) is a widely used scale in assessing neurological deficits in stroke patients. Its measurement properties and accuracy in predicting clinical outcomes have also been extensively evaluated.[1–4] Although it contains only 15 items and takes <8 minutes to perform,[5] it has been critiqued for its complexity and variability.[6–8] As a result, shortened versions of the NIHSS have been suggested.[7–9] Recently, we conducted a large-scale clinical trial that used the NIHSS, as well as other outcome measures commonly used in ischemic stroke patients. Using the data from this trial, the objective of the current study was to compare the measurement properties of the original NIHSS with the shortened versions. Hereafter, the original version is referred to as NIHSS-15, and the 3 shortened versions, 11-item, 8-item, and 5-item, are referred to as NIHSS-11, NIHSS-8, and NIHSS-5, respectively. The 1-item version, however, was not included in the comparison.

DOI: 10.1161/STROKEAHA.115.011657

Table 1. Summary of Stroke Measures Classified by Treatment Group

Stroke Measure	All Patients	MLC601	Placebo	P Value
Categorical, N (%)				
mRS at day 10				
0–1	180 (16.8)	84 (15.6)	96 (18.0)	0.327
2–5	891 (83.2)	453 (84.4)	438 (82.0)	
mRS at month 3				
0–1	460 (48.5)	239 (50.0)	221 (46.9)	0.363
2–5	489 (51.5)	239 (50.0)	250 (53.1)	
Barthel index at month 3				
≥85	675 (71.1)	344 (72.0)	331 (70.3)	0.568
<85	274 (28.9)	134 (28.0)	140 (29.7)	
Continuous, mean (SD)/median (IQR)				
MMSE at day 10	25.8 (3.1)/28 (24–30)	25.6 (5.8)/28 (24–30)	26.0 (5.7)/28 (24–30)	0.250
MMSE at month 3	27.0 (4.6)/29 (26–30)	26.9 (4.8)/29 (26–30)	27.1 (4.5)/29 (26–30)	0.465
NIHSS at day 10				
NIHSS-15	6.8 (3.4)/6 (4–9)	6.9 (3.3)/6 (4–9)	6.7 (3.4)/6 (4–9)	0.590
NIHSS-11	4.6 (3.0)/4 (2–7)	4.6 (2.9)/4 (2–7)	4.6 (3.0)/4 (2–6)	0.695
NIHSS-8	3.9 (2.1)/4 (2–5)	3.9 (2.1)/4 (2–5)	3.9 (2.2)/4 (2–5)	0.695
NIHSS-5	2.0 (1.5)/2 (1–3)	2.1 (1.5)/2 (1–3)	2.0 (1.5)/2 (1–3)	0.618
NIHSS at month 3				
NIHSS-15	3.5 (3.1)/3 (1–5)	3.4 (3.0)/3 (1–5)	3.6 (3.3)/3 (1–5)	0.420
NIHSS-11	2.5 (2.7)/2 (1–4)	2.4 (2.5)/2 (1–3)	2.5 (2.8)/2 (1–4)	0.580
NIHSS-8	1.9 (1.8)/2 (1–3)	1.9 (1.7)/2 (1–3)	2.0 (1.9)/1 (1–3)	0.218
NIHSS-5	1.0 (1.3)/1 (0–1)	1.0 (1.2)/1 (0–1)	1.1 (1.4)/1 (0–2)	0.341

IQR indicates interquartile range; MMSE, mini-mental status examination; mRS, modified Rankin Scale; NIHSS, National Institutes of Health Stroke Scale; NIHSS-5, 5-item NIHSS; NIHSS-8, 8-item NIHSS; NIHSS-11, 11-item NIHSS; NIHSS-15, original 15-item NIHSS; and SD, standard deviation.

Statistical Analysis

Subjects recruited for this trial were ischemic stroke patients having a prestroke mRS of 0 or 1 and baseline NIHSS-15 score between 6 and 14; hence, the baseline measures did not vary sufficiently for the evaluation of measurement validity. Therefore, in this study, the data obtained at day 10 and month 3 were analyzed. Patients who died before the assessment at month 3 were excluded from the analysis. Main analyses were based on all eligible patients, whereas subgroup analyses by treatment were also conducted as sensitivity analyses to examine the robustness of the results.

The discriminative ability and responsiveness to change of the 4 NIHSS scores were compared by regressing each score on other

Table 2. Effect Size of the 4 NIHSS Versions in Relation to mRS, MMSE, and BI

NIHSS version	Model R^2	mRS		MMSE		Barthel Index (per 5 Points)	
		Effect Size	Difference (95% CI)	Effect Size	Difference (95% CI)	Effect Size	Difference (95% CI)
Discriminative ability							
NIHSS-15	0.64	0.55		−0.05		−0.16	
NIHSS-11	0.62	0.46	0.09 (0.04–0.13)	−0.04	−0.01 (−0.01–0.00)	−0.17	0.01 (−0.01–0.03)
NIHSS-8	0.59	0.46	0.09 (0.04–0.15)	−0.05	0.00 (−0.01–0.02)	−0.14	−0.02 (−0.04–0.00)
NIHSS-5	0.57	0.35	0.20 (0.12–0.27)	−0.05	−0.00 (−0.02–0.02)	−0.17	0.01 (−0.02–0.03)
Responsiveness to change							
NIHSS-15	0.57	0.84		−0.06			
NIHSS-11	0.55	0.74	0.10 (0.06–0.15)	−0.05	−0.01 (−0.02–0.00)		
\NIHSS-8	0.47	0.78	0.06 (0.00–0.10)	−0.07	0.01 (−0.01–0.02)		
NIHSS-5	0.40	0.77	0.07 (0.01–0.12)	−0.05	−0.01 (−0.02–0.00)		

BI indicates Barthel Index; CI, confidence interval; MMSE, mini-mental status examination; mRS, modified Rankin Scale; NIHSS, National Institutes of Health Stroke Scale; NIHSS-5, 5-item NIHSS;NIHSS-8, 8-item NIHSS; NIHSS-11, 11-item NIHSS; and NIHSS-15, original 15-item NIHSS.

stroke measures in a linear regression model. For discriminative ability, the NIHSS scores at month 3 were regressed on mRS and MMSE, and BI divided by 5 at month 3, adjusting for baseline characteristics and risk factors. The resulting regression coefficients can be interpreted as the change in NIHSS score per unit increment in mRS and MMSE and every 5-point increment in BI, respectively. Because the 4 NIHSS scores are not of the same metric and not directly comparable, standardization by dividing these coefficients by the residual standard deviation of the regression model was performed to obtain the effect size. This effect size may be regarded as a signal-to-noise ratio reflecting measurement precision: the larger the signal (numerator) and the smaller the noise (denominator), the larger the effect size.[12] A score with a larger effect size is desired because it requires a smaller sample size to achieve the same research purpose. The 95% confidence interval for the difference between the effect sizes of the original and shortened versions was estimated using 1000 bootstrap replications.[13] Similarly, responsiveness to change from day 10 to month 3 was also compared. Confirmatory factor analysis and receiver operating characteristic curve analysis were also performed (please see the online-only Data Supplement). All analyses were performed using SAS v9.3.

Results

The overall baseline characteristics and study flow of patients in the CHIMES study were previously described.[11] In this study, we used the data from 1071 patients at day 10 and 949 patients at month 3 who were alive and had complete NIHSS assessments. Their mRS, MMSE, BI, and NIHSS scores are summarized in Table 1. There was no significant difference in these scores between the 2 groups.

The upper panel of Table 2 shows the effect size of the 4 NIHSS scores for detecting a difference in mRS, MMSE, and BI at month 3 adjusting for baseline characteristics and risk factors, together with the bootstrap confidence interval for the difference with NIHSS-15. For mRS, the original NIHSS-15 score had a significantly larger effect size (0.55) than the other 3 shortened versions (0.35–0.46). However, the effect sizes were similar among the 4 versions and of small magnitude for MMSE (–0.05 to –0.04) and BI (–0.17 to –0.14). Similarly, the lower panel of Table 2 compares the effect size of the NIHSS scores for detecting a change from day 10 to month 3 in mRS and MMSE. Again, NIHSS-15 achieved a larger effect size (0.84) for mRS than the other 3 versions (0.74–0.78), whereas the 4 versions had similar effect size for MMSE (–0.06 to –0.06).

A set of sensitivity analyses stratified by treatment group were also performed for the above analyses. Results were not qualitatively different from that based on all patients and, thus, are not repeated here. Results for confirmatory factor analysis and receiver operating characteristic curve analysis were summarized in online-only Data Supplement (Table I, Figure I, and Figure II in the online-only Data Supplement).

Discussion

Our comparison of the original version with the shortened versions of the NIHSS showed that the original 15-item NIHSS exhibited the largest effect size in discriminative ability and responsiveness to change in relation to mRS. However, regardless of the version used, the NIHSS scores seemed to be only weakly to moderately associated with the MMSE and BI, suggesting that these instruments may measure domains and functions not addressed by the NIHSS and, hence, function as complementary outcomes.

Our study does have clinical and research implications. Busy clinicians are often tempted by shortened versions of clinical scales to allow more efficient use of their time with the patient. Researchers would prefer fewer data points as this reduces data entry time and occasions for errors. However, the shortened NIHSS, although easier and quicker to perform than the full version, comes at a cost of lowered discriminatory value and responsiveness to change, which may impact negatively on the ability to predict outcome, be it in clinical or research settings.

A limitation of this study was the relatively stringent inclusion criteria of the trial, which recruited patients with narrow ranges of prestroke mRS and baseline (original version) NIHSS. Hence, our findings may not be extrapolated beyond the population included in the study, that is, more severe strokes and patients dependent before the index stroke. The shortened versions may be more discriminative and useful in this group of more severe stroke patients. The strengths of this study include that it is a multicentre study performed by experienced stroke trialists, which would provide reliable data. The study involved a large number of subjects, the vast majority of whom had complete NIHSS and functional outcome data.

In summary, despite being critiqued for its complexity and variability, the original 15-item NIHSS, which can be performed easily over a few minutes, retained information that made it more discriminative and responsive to change than the shortened versions. We recommend that clinical researchers use the full version NIHSS to evaluate patients' stroke severity.

Appendix–CHIMES Investigators

CHIMES Study Investigators: Philippines—Jose C. Navarro, Herminigildo H. Gan, Annabelle Lao, Alejandro Baroque II, Johnny Lokin, John Harold B. Hiyadan, Ma. Socorro Sarfati, Randolph John Fangonillo, Neil Ambasing, Carlos Chua, Ma. Cristina San Jose, Joel Advincula, Eli John Berame, Maria Teresa Canete. Singapore—Narayanaswamy Venketasubramanian, Sherry H. Young, Marlie Jane Mamauag, San San Tay, Shrikant Pande, Umapathi Thirugnanam, Rajinder Singh, Hui Meng Chang, Deidre Anne De Silva, Bernard P.L. Chan, Vijay Sharma, Teoh Hock Luen. Thailand—Niphon Poungvarin, Sombat Muengtaweepongsa, Somchai Towanabut, Nijasri Suwanwela, Songkram Chotickanuchit, Siwaporn Chankrachang, Samart Nitinun. Sri Lanka—H. Asita de Silva, Udaya Ranawake, Nirmala Wijekoon. Hong Kong—K.S. Lawrence Wong. Malaysia—Gaik Bee Eow.

Disclosures

None.

References

1. Goldstein LB, Bertels C, Davis JN. Interrater reliability of the NIH stroke scale. *Arch Neurol*. 1989;46:660–662.
2. Schmülling S, Grond M, Rudolf J, Kiencke P. Training as a prerequisite for reliable use of NIH Stroke Scale. *Stroke*. 1998;29:1258–1259.
3. Albanese MA, Clarke WR, Adams HP, Jr, Woolson RF. Ensuring reliability of outcome measures in multicenter clinical trials of treatments for acute ischemic stroke. The program developed for the Trial of Org 10172 in Acute Stroke Treatment (TOAST). *Stroke*. 1994;25:1746–1751.
4. Lyden P, Lu M, Jackson C. Underlying structure of the National Institutes of Health Stroke Scale: results of a factor analysis. *Stroke* 1999; 30:2347–2354.
5. 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.
6. Dewey HM, Donnan GA, Freeman EJ, Sharples CM, Macdonell RA, McNeil JJ, et al. Interrater reliability of the National Institutes of Health Stroke Scale: rating by neurologists and nurses in a community-based stroke incidence study. *Cerebrovasc Dis*. 1999;9:323–327. doi: 16006.
7. Lyden PD, Lu M, Levine SR. A modified National Institutes of Health Stroke Scale for use in stroke clinical trials: preliminary reliability and validity. *Stroke* 2001; 32:1310–1317.
8. Meyer BC, Hemmen TM, Jackson CM, Lyden PD. Modified National Institutes of Health Stroke Scale for use in stroke clinical trials: prospective reliability and validity. *Stroke*. 2002;33:1261–1266.
9. Tirschwell DL, Longstreth WT, Jr, Becker KJ, Gammans RE Sr, Sabounjian LA, Hamilton S, et al. Shortening the NIH Stroke scale for use in the prehospital setting. *Stroke*. 2002;33:2801–2806.
10. Venketasubramanian N, Chen CL, Gan RN, Chan BP, Chang HM, Tan SB, et al; CHIMES Investigators. A double-blind, placebo-controlled, randomized, multicenter study to investigate Chinese Medicine Neuroaid Efficacy on Stroke recovery (CHIMES Study). *Int J Stroke*. 2009;4:54–60. doi: 10.1111/j.1747-4949.2009.00237.x.
11. Chen CL, Young SH, Gan HH, Singh R, Lao AY, Baroque AC II, et al; CHIMES Study Investigators. Chinese medicine neuroaid efficacy on stroke recovery: a double-blind, placebo-controlled, randomized study. *Stroke*. 2013;44:2093–2100. doi: 10.1161/STROKEAHA.113.002055.
12. Fayers PM, Machin D. *Quality of Life: Assessment, Analysis and Interpretation*. England: John & Sons; 2000:4.
13. Efron B, Tibshirani R. *An Introduction to the Bootstrap*. New York, NY: Chapman & Hall; 1993.

Stroke

American Heart Association

American Stroke Association ®

Comparison Between the Original and Shortened Versions of the National Institutes of Health Stroke Scale in Ischemic Stroke Patients of Intermediate Severity

Chun Fan Lee, Narayanaswamy Venketasubramanian, K.S. Lawrence Wong, Christopher L.H. Chen and for the CHIMES Study Investigators

The online version of this article, along with updated information and services, is located on the
World Wide Web at:
http://stroke.ahajournals.org/content/47/1/236