Establishment of the Thai version of National Institute of Health Stroke Scale (NIHSS) and a Validation Study

Yongchai Nilanont MD*, Sucheera Phattharayuttawat PhD**,
Pipat Chiewit MD***, Songkram Chotikanuchit MD*,
Juljak Limsriwilai MD*, Lultrita Chalernpong MD*,
Napaporn Yamkaew RN*, Niramol Lirathpong RN*,
Pornpimol Anprasertporn MD*, Chulaluk Komoltri DrPH****,
Naraporn Prayoonwiwat MD*, Niphon Poungvarin MD, FRCP, FRI*

** Division of Neurology, Department of Medicine, Siriraj Hospital, Mahidol University, Bangkok, Thailand.*
*** Department of Psychiatry, Siriraj Hospital, Mahidol University, Bangkok, Thailand.*
**** Department of Radiology, Faculty of Medicine, Siriraj Hospital, Mahidol University, Bangkok, Thailand.*
***** Department of Research development, Mahidol University, Bangkok, Thailand*

Background: *The National Institute of Health Stroke Scale (NIHSS) is a well validated stroke severity assessment scale. The scale is used to assess neurological deficits, progression and degree of recovery in acute stroke settings. To date, no Thai version exists.*
Objective: *This study aimed to adapt and validate a Thai version of the NIHSS (NIHSS-T).*
Material and Method: *A cross-cultural adaptation of the NIHSS was developed according to the methods recommended by the International Quality of Life Assessment Project Group. Forward and backward translations were performed. A final version of the NIHSS-T was validated against initial MRI infarction volume and modified Rankin Scale (mRS) at 3 months in a consecutive series of acute stroke patients. The patients were prospectively evaluated by 3 different types of health care providers: 2 stroke fellows, 2 internists, and 2 stroke nurses. Mean NIHSS-T scores from all raters were used in the analysis.*
Results: *The study included 32 acute ischemic stroke patients with a mean age (± SD) of 64.53 ± 14.97 years of age. Men comprised 71.9%. Mean NIHSS-T score (± SD) was 7.49 ± 7.02. Intra-observer reliability demonstrated a high agreement with an intraclass correlation (ICC) of 0.98, 0.98, 0.96, 0.98, 0.90 and 0.98 for 2 stroke fellows, 2 internists and 2 stroke nurses respectively. Inter-observer reliability between 6 raters was excellent, i.e.; ICC, 0.99 (0.98, 0.99). Spearman rank correlation coefficients between the initial NIHSS-T score versus initial MRI lesion volume and mRS at 3 months were 0.53 and 0.69 with a p-value of 0.002 and < 0.001 respectively.*
Conclusion: *The Thai version of NIHSS is valid for assessing acute stroke severity. The scale is also reliable when administered in a Thai-speaking setting by trained healthcare professionals.*

Keywords: *National Institute of Health Stroke Scale (NIHSS), Thai version, Stroke severity assessment, Neurological scales*

J Med Assoc Thai 2010; 93 (Suppl. 1): S171-178
Full text. e-Journal: *http://www.mat.or.th/journal*

The National Institute of Health Stroke Scale (NIHSS) is one of the most reliable stroke severity assessment tools[1-3]. It measures 15 neurological items including level of consciousness, eye movements, vision, facial expression, arm and leg strength, limb ataxia, sensation, language ability, speech difficulty, sensory and visual inattention[4,5]. Each item uses a 3- or 4-point scale resulted in a total score of 0-42, with higher scores indicating severe deficits. The score has been widely validated in many prospective and retrospec-

Correspondence to: Nilanont Y, Division of Neurology, Department of Medicine, Faculty of Medicine, Siriraj Hospital, Mahidol University, Bangkok 10700, Thailand. Phone: 0-2419-7101, Fax: 0-2412-3009. E-mail: siysl@mahidol.ac.th

tive studies[5-13]. However, its use is still limited especially in non-english speaking countries due to language and cultural barriers.

Currently, there is a strong need for a well validated stroke severity assessment tool for the Thais. An informal survey performed at the Department of Medicine, Faculty Medicine Siriraj Hospital, Mahidol University, Bangkok, Thailand in November, 2008 showed that only 7.9% from a total of 127 internal medicine residents practice NIHSS to evaluate patients with acute stroke; whereas, 92.1% use Glasgow Coma Scale (GCS) for assessing most neurological conditions including acute stroke. GCS was first introduced by Teasdale and Jennet in 1974[14]. It has gained widespread acceptance around the world as a means of assessing level of consciousness in patients with head injury. However, the scale was not designed for assessing stroke severity. The use of inappropriate clinical tool may lead to an unawareness of clinical deterioration and poor outcomes. Currently, the original NIHSS English version is being performed in very few centers in Thailand. In addition, some items originally designed for testing language and speech abilities were inapplicable for most of the Thais due to cultural and language differences. This study aimed to develop and assess the reliability and validity of the National Institute of Health Stroke Scale, Thai version (NIHSS-T).

Objectives

The purpose of this study is to construct and perform a reliability and validity test of the NIHSS-T.

Material and Method

Scale Development

NIHSS-T was constructed according to methods recommended by the international quality of life assessment project[15-17]. This process composed of 4 steps as follows:

Step 1: A forward translation of the original NIHSS was performed by 2 stroke neurologists (Y.N. and Ni.P.). The initial NIHSS-T version was developed.

Step 2: A backward translation of the first NIHSS-T was performed by 1 stroke neurologist (Na.P.) and 1 psychiatrist (SP) who are fluent in English. Any possible misunderstandings or misinterpretations were detected and corrected by comparing the backward translation and the original NIHSS English version. The second NIHSS-T was then constructed.

Step 3: The second version NIHSS-T was reviewed by 3 independent stroke experts and 1 speech therapist whose native language is Thai to evaluate comprehension and acceptability of the translated version. All translators reviewed the scale, made corrections and additions to produce the third version of the NIHSS-T.

Step 4: The third version was assessed and proof-read to correct any spelling, grammar, and other mistakes. Any disagreements were solved through a consensus meeting of the investigator team.

Concerning the items for testing language ability and speech, a list of Thai words and phrases were created (Fig. 1, 2). Two pages with drawings intended for assessing impairment in naming and visual inattention (neglect) were developed. Some modifications of the original version were made and adjusted according to Thai culture. The NIHSS-T was then finalized and ready to be tested for its reliability and validity.

Fig. 1 Items intended to test for visual inattention and verbal fluency

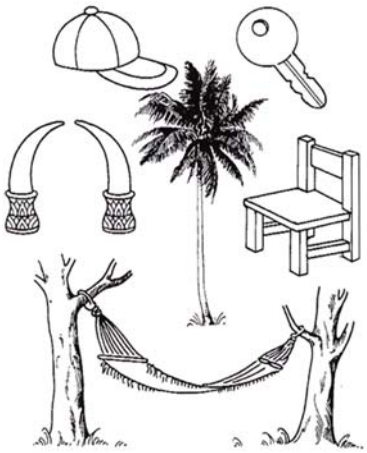

Fig. 2 Items intended to test for naming impairment

Reliability and Validity Test

Study Design

This was a prospective observational study aimed at testing the reliability and validity of the NIHSS-T in acute stroke setting in the Thais.

Population

This study was conducted at the Siriraj Acute Stroke Unit, Division of Neurology, Department of Medicine, Faculty of Medicine Siriraj Hospital, Mahidol University, Bangkok, Thailand. Patient was eligible if the age was $\geq$ 18 years with a diagnosis of acute ischemic stroke within 7 days and a positive brain MRI which was responsible for the index event. Eligible patients hold the Thai nationality and were able to speak and read Thai language fluently prior to the index event. Patients with accompanying other intracranial pathologies (*e.g.* tumors, brain abscess) were excluded from the study.

Study Protocol

Baseline evaluation

Index stroke evaluation including history taking and general physical examination were performed within 24 hours of admission by one of the study neurologists. The following data were collected: 1) Demographic information (age, sex, and highest level of education attained); 2) Co-morbidities (diabetes mellitus, hypertension, heart diseases); 3) Stroke subtype according to the Trial of Org 10172 in Acute Stroke Treatment (TOAST) criteria[18]; 4) Disability scale using modified Rankin Scale (mRS)[19,20] at 3 months; 5) MR brain was performed within 48 hours after admission by a 3.0 Tesla MR machine (Philips, Netherlands) in three dimensional AXIAL T1W and AXIAL FLAIR. Volume of brain infarction was quantified by an independent neuroradiologist (PC) who was blinded to clinical information using images from axial FLAIR pulse sequence. Images were transferred to View Forum Extended MR Work Space release 2.5.3.0 2007 by Philips Medical System, Nederland BV, the Netherlands. Areas of cerebral infarction were obtained using cursor that was freely drawn along the margin of cerebral infarct in each slice. After a serial measurement of these axial acquisition slices, automatic volumetric calculation software will provide the infarct volume. This method will facilitate the measurement even though the margin is polygonal shaped.

Scale testing

Eligible patients were evaluated by six raters with different health care expertise: 2 stroke fellows, 2 internal medicine residents, and 2 stroke unit nurses. All raters were trained to administer both the original English version NIHSS and the NIHSS-T. The training consisted of watching video training tools of the original NIHSS, studying and practicing the NIHSS-T under direct supervision by a certified examiner. Only one of the six raters was assigned to administer both scales to the patient within 48 hours after admission using computer-generated random selection. During the administration of the tests, a videotape was recorded using the same protocol. After an initial assessment, the tape was reviewed and scored independently by the rest of the raters (5 raters). The first set of scores was then collected. After 3-5 weeks, all raters were assigned to review the same video tape and independently provided a second set of the NIHSS score. Total and sub-score of NIHSS-T from both the first and second set were used in the analysis.

Follow-up procedure and evaluation

Follow-up visit with neurologist who did not involve in this study was made every month for 3 months. Modified Rankin Scale (mRS) and Barthel Index (BI) were performed at 3 month.

Sample Size Calculation and Statistical Analysis

The target sample size for the primary null hypothesis was 32 patients, which was estimated to have 95% power for detecting a rho (ρ) of 0.7 of the Spearman's rank order correlation coefficient at a 2-sided alpha level of 0.01. The provided sample size has already been accounted for a 10% loss to follow-up.

Baseline characteristics were analyzed using the median or mean $\pm$ SD for continuous variables according to their distribution. Categorical variables were analyzed in terms of their relative frequency. The NIHSS-T total score was calculated according to the methods proposed in the original article.

The Kappa statistics and intraclass correlation coefficient were employed to evaluate inter and intra-rater reliabilities.

The study had been approved by Siriraj Institutional Review Board (COA No. Si 094/2007).

Results

A total of 32 acute ischemic stroke patients were enrolled with a mean age $\pm$ SD of 64.53 $\pm$ 14.97 years (range, 27-91 years). Men comprised of 71.90%. Stroke subtypes according to the TOAST classification were as follows: large vessel atherosclerosis 28.9%,

Table 1. Patients' characteristics

Characteristics	
Mean age ± SD, years (range)	64.53 ± 14.97 (27-91)
Men (%)	71.9
Mean NIHSS-T (range)*	7.49 ± 7.02 (0-31)
Mean time ± SD to complete NIHSS-T (minutes)	8.35 ± 3.12
Stroke Subtype	(%)
Ischemic	100
Cardioembolic	36.80
Large artery atherosclerosis	28.90
Small vessel occlusion	34.20
Stroke Location	(%)
Anterior circulation	78.90
Posterior circulation	21.10

* Mean NIHSS-T was calculated from the sum of total NIHSS-T score from all assessments divided by 12 (the number of evaluation performed).

cardio-embolic 36.8% and small vessel occlusion 34.2%. Anterior circulation stroke accounted for 78.9% whereas 21.1% had posterior circulation ischemia. Diabetes, hypertension, coronary artery disease were diagnosed in 60.5, 60.5, 47.4% of the patients, respectively. Detail baseline characteristics, stroke location and stroke etiology by TOAST classification were shown in Table 1. The NIHSS-T was performed in all patients with a mean total score ± SD of 7.49 ± 7.02. Mean time to complete the NIHSS-T examination ± SD was 8.35 ± 3.12 minutes. Median mRS at 3 months was 3 (95% CI; 1.86-3.30). Mean MRI lesion volume ± SD was 29.43 ± 51.83 cm^3.

Inter-rater Reliability

All patients were evaluated by six raters with different health care expertise: 2 board certified neurologists, 2 internal medicine residents, and 2 stroke unit nurses. Inter-observer reliability of the total NIHSS-T score between 6 raters was excellent with an intraclass correlation coefficient (ICC) of 0.99 (0.98, 0.99). Table 2 showed interobserver reliability among 6 raters for the 15 tested items.

Test-Retest/Intra-rater Reliability

All patients were graded twice within a 3-5 week interval by the same raters using a video tape which was recorded during the initial NIHSS-T evaluation. Intra-rater reliability demonstrated a high agreement of the total NIHSS-T score with an intraclass correlation (ICC) of 0.98, 0.98, 0.96, 0.98, 0.98 and 0.90 for 2 neurologists, 2 internal medicine residents and 2 stroke nurses respectively Table 3.1. Intra-rater reliability of each NIHSS-T subscore was demonstrated in percentage of crude agreement and weighted kappa as was shown in Table 3.2.

Construct Validity

All patients completed the 3-month follow-up. The construct validity of the NIHSS-T was substantial as was demonstrated by the Spearman rank correlation coefficient between the initial NIHSS-T total score and the 3-month modified Rankin Score of 0.69 ($p < 0.001$). The score also had a significant positive correlation with the initial brain infarction volume (correlation coefficient of 0.53, $p = 0.002$).

Discussion

Our study showed excellent inter- and intra-rater reliabilities of the total NIHSS-T score with an intraclass correlation coefficient of more than 0.9 in all levels of healthcare providers. The NIHSS-T was valid when being used by Thai-speaking health care professionals examining Thai-speaking patients with an acute stroke as shown by a rho (ρ) of 0.69, $p < 0.001$, and 0.53, $p = 0.002$ (when compared the initial score with the 3-month mRS and volume of infarction respectively).

This study involved raters with different healthcare expertise including stroke nurses, internists and neurologists in order to test whether the NIHSS-T scale could be reliably used in all level of Thai medical personnel. An excellent intra-rater reliability ($ICC \geq 0.9$) of the total NIHSS-T score was consistently found among 6 raters. Intra- rater reliability for each subscale

Table 2. Interobserver reliability among 6 raters as demonstrated by Intraclass Correlation Coefficient (ICC)

Item	ICC (95% CI)	
	Observe #1	Observe #2
1a. Level of Consciousness (LOC)	0.15 (0.04, 0.31)	0.09 (-0.01, 0.24)
1b. LOC, questions	0.85 (0.77, 0.91)	0.83 (0.74, 0.90)
1c. LOC, commands	0.97 (0.95, 0.98)	0.97 (0.95, 0.98)
2. Best gaze	0.18 (0.06, 0.35)	0.34 (0.19, 0.52)
3. Visual fields	0.49 (0.34, 0.66)	0.47 (0.32, 0.64)
4. Facial palsy	0.43 (0.28, 0.61)	0.33 (0.18, 0.51)
5a. Motor (left)	0.82 (0.72, 0.89)	0.94 (0.91, 0.97)
5b. Motor (right)	0.94 (0.89, 0.96)	0.99 (0.99, 0.99)
6a. Motor (left)	0.76 (0.65, 0.85)	0.89 (0.83, 0.94)
6b. Motor (right)	0.91 (0.86, 0.95)	0.95 (0.92, 0.97)
7. Limb ataxia	0.03 (-0.04, 0.16)	0.19 (0.07, 0.36)
8. Sensory	0.62 (0.48, 0.76)	0.59 (0.44, 0.73)
9. Best language	0.82 (0.72, 0.89)	0.89 (0.84, 0.94)
10. Dysarthria	0.58 (0.44, 0.73)	0.54 (0.39, 0.69)
11. Extinction and Inattention	0.23 (0.11, 0.41)	0.35 (0.21, 0.53)

Table 3.1. Intra-rater reliability

	Intraclass correlation	95% confidence interval
Fellow A	0.98	0.96-0.99
Fellow B	0.98	0.95-0.99
Resident A	0.96	0.93-0.98
Resident B	0.98	0.96-0.99
Nurse A	0.89	0.80-0.95
Nurse B	0.98	0.95-0.99

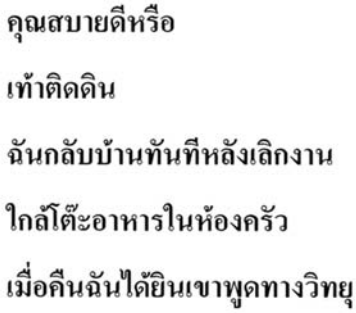

Fig. 3 Items intended to test for dysarthric speech

was assessed by calculating the kappa statistics. The kappa statistics corrects for the chance agreement between the first and second evaluations by the same rater[21,22]. The degree of intra- observer agreement based on the kappa statistic may be interpreted with the following scale: < 0, poor agreement; 0-0.20, slight agreement; 0.21-0.40, fair agreement; 0.41-0.60, moderate agreement; 0.61-0.80, substantial agreement; 0.81-1.00, almost perfect agreement[23]. Substantial to almost perfect agreement was found in 11 of 15 tested NIHSS-T items. Three of the 15 items were found to have fair to moderate agreement. The kappa agreement was lowest which indicated poor agreement in the item of limb ataxia. This finding was consistent in all groups of rater which was similar to the previous studies of their original English[5] and Chinese version[24].

Similar result was found when the NIHSS-T being assessed for the inter-rater reliability in which agreement was substantial to almost perfect (ICC > 0.60) in 8 of 15 items. Poor to slight agreement (ICC < 0.4) was found on level of consciousness (LOC), best gaze, inattention and limb ataxia. The least reliable item was limb ataxia as shown with an ICC of 0.03 and 0.19 for the first and second evaluation among 6 raters. This was also consistent with prior studies[1,5,24].

In this study, we have increased both inter and intra-observer variabilities and avoid any biases as much as possible by the following methods. Firstly, the initial administration of the NIHSS-T was recorded and the same video tape was reviewed by all raters to grade the second set of the score. An interval of at least 3-5 weeks after the first assessment was required for subsequent evaluation in order to prevent any possible recalled bias. By using a videotape in the second

Table 3.2. Intra-rater reliability between two repeated measurements from the same observer

Item	% Crude agreement, Weighted kappa					
	Fellow		Resident		Nurse	
	A	B	A	B	A	B
1a. LOC	96.8, 0.65	100, *	96.9, #	96.9, 0.65	100, *	100, 1
1b. LOC, Q	93.8, 0.80	81.3, 0.82	93.8, 0.95	90.6, 0.91	84.4, 0.82	87.5, 0.91
1c. LOC, C	100, 1	96.9, 0.96	100, 1	100, 1	90.6, 0.89	96.9, 0.96
2. Best gaze	87.5, 0.67	96.9, #	81.3, 0.09	90.6, 0.19	84.4, 0.60	81.3, 0.56
3. Visual fields	90.6, 0.37	96.9, 0.84	87.5, 0.62	87.5, 0.77	87.5, 0.80	84.4, 0.57
4. Facial palsy	56.3, 0.54	65.6, 0.23	65.6, 0.83	56.3, 0.46	65.6, 0.59	71.9, 0.47
5a. Motor (Lt.)	90.6, 0.78	96.9, 0.99	90.6, 0.86	96.9, 0.98	96.9, 0.85	96.9, 0.99
5b. Motor (Rt.)	93.8, 0.95	90.6, 0.97	90.6, 0.87	96.9, 0.97	90.6, 0.98	93.8, 0.98
6a. Motor (Lt.)	87.5, 0.95	81.3, 0.90	93.8, 0.92	93.8, 0.96	84.4, 0.61	93.8, 0.98
6b. Motor (Rt.)	87.5, 0.96	84.4, 0.94	81.3, 0.94	90.6, 0.96	84.4, 0.79	87.5, 0.97
7. Limb ataxia	87.5, -0.07	81.3, 0.49	87.5, -0.05	96.9, #	75.0, 0.37	81.3, 0.62
8. Sensory	90.6, 0.79	84.4, 0.73	84.4, 0.76	84.4, 0.80	90.6, 0.83	96.9, 0.95
9. Language	93.8, 0.97	93.8, 0.96	87.5, 0.90	81.3, 0.91	78.1, 0.89	81.3, 0.91
10. Dysarthria	62.5, 0.67	81.3, 0.80	78.1, 0.75	75.0, 0.67	81.3, 0.84	81.3, 0.90
11. Extinction and Inattention	87.5, 0.56	93.8, 0.47	81.3, 0.29	87.5, 0.76	71.9, 0.50	75.0, 0.63

* 1x1 table, # 1x2, 2x1 table; Kappa cannot be computed since the table is not a 2x2 table.

Fig. 4 Items intended to test for dysarthric speech and alexia

evaluation, a possibility of having any changes in neurological signs that might occur overtime can be eliminated. Secondly, a training and practice session was given to all raters prior to the study. This can greatly increase the inter-rater agreement[25]. Finally, this research employed a random sampling method in that the first examiner was randomly selected to perform the NIHSS-T examination.

There are certain limitations in our study. Firstly, only acute ischemic stroke patients were included in the study; therefore, the scale may not be applicable in hemorrhagic stroke. Secondly, the score indicates severity of stroke which may depend on the location and extent of lesion. Ischemic stroke patients with different infarct locations may get a significantly different score. Patient with a small cerebral infarct located around periventricular area may receive a low score on NIHSS-T while patients with the same amount of infarct at brainstem may have a very high score due to different fiber tract involvement. For this reason, the relationship between MRI lesion volume and the NIHSS-T score may not be perfectly correlated as shown in this study with a rho (ρ) of 0.53, $p = 0.002$. However, the score was shown to have a substantial correlation with patients' disability level at 3 months. Thirdly, arranging an MRI brain to be performed at the same time interval after stroke onset was not practical for most of the patients in our center due to its availability. This may be another possible explanation for a lower correlation coefficient ($\rho = 0.53$) when compared to the relationship between initial NIHSS-T and disability level (mRS) at 3 months ($\rho = 0.69$).

Although further validation study of the NIHSS-T in other stroke settings, *i.e.*; intracerebral hemorrhage is needed, we recommend integrating this scale into a routine neurological evaluation in acute stroke setting as it provides a valid and reliable tool that can serve as an outcome predictor after stroke. This study

demonstrated that Thai-speaking healthcare professionals trained with the NIHSS-T administration can be confident in applying this clinical assessment tool with Thai-speaking patients in assessing stroke severity.

Conclusion

The NIHSS, Thai version is reliable and valid for assessing acute stroke severity.

Acknowledgements

The authors would like to thank Dr. Surat Boonyakarnkul, Dr. Roekchai Tulyapronchote, Dr. Nijasri Charnnarong Suwanwela, Dr. Suchat Hanchaiphiboolkul, Dr. Samart Nidhinandana for their comments concerning the Thai version of the National Institute of Health Stroke Scale. The authors thank all patients who participated in this study and Ms. Pimpa Thepphawan for technical and logistic support.

References

1. D'Olhaberriague L, Litvan I, Mitsias P, Mansbach HH. A reappraisal of reliability and validity studies in stroke. Stroke 1996; 27: 2331-6.
2. Adams HP, Jr., Davis PH, Leira EC, 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; 13: 126-31.
3. Lyden PD, Hantson L. Assessment scales for the evaluation of stroke patients. J Stroke Cerebrovasc Dis 1998; 7: 113-27.
4. 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: 864-70.
5. Goldstein LB, Bertels C, Davis JN. Interrater reliability of the NIH stroke scale. Arch Neurol 1989; 46: 660-2.
6. 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: 1534-7.
7. 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: 656-60.
8. Williams LS, Yilmaz EY, Lopez-Yunez AM. Retrospective assessment of initial stroke severity with the NIH Stroke Scale. Stroke 2000; 31: 858-62.
9. De Haan R, Horn J, Limburg M, Van Der Meulen J, Bossuyt P. A comparison of five stroke scales with measures of disability, handicap, and quality of life. Stroke 1993; 24: 1178-81.
10. Wang S, Lee SB, Pardue C, Ramsingh D, Waller J, Gross H, et al. Remote evaluation of acute ischemic stroke: reliability of National Institutes of Health Stroke Scale via telestroke. Stroke 2003; 34: e188-92.
11. 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: 2801-6.
12. 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: 1261-6.
13. Kasner SE, Cucchiara BL, McGarvey ML, Luciano JM, Liebeskind DS, Chalela JA. Modified National Institutes of Health Stroke Scale can be estimated from medical records. Stroke 2003; 34: 568-70.
14. Teasdale G, Jennett B. Assessment of coma and impaired consciousness. A practical scale. Lancet 1974; 2: 81-3.
15. Gandek B, Ware JE Jr. Methods for validating and norming translations of health status questionnaires: the IQOLA Project approach. International Quality of Life Assessment. J Clin Epidemiol 1998; 51: 953-9.
16. Bullinger M, Alonso J, Apolone G, Lepl□ge A, Sullivan M, Wood-Dauphinee S, et al. Translating health status questionnaires and evaluating their quality: the IQOLA Project approach. International Quality of Life Assessment. J Clin Epidemiol 1998; 51: 913-23.
17. Ware JE Jr, Keller SD, Gandek B, Brazier JE, Sullivan M. Evaluating translations of health status questionnaires. Methods from the IQOLA project. International Quality of Life Assessment. Int J Technol Assess Health Care 1995; 11: 525-51.
18. Adams HP, Jr., Bendixen BH, Kappelle LJ, Biller J, Love BB, Gordon DL, et al. Classification of subtype of acute ischemic stroke. Definitions for use in a multicenter clinical trial. TOAST. Trial of Org 10172 in Acute Stroke Treatment. Stroke 1993; 24: 35-41.
19. Rankin J. Cerebral vascular accident in patients over the age of 60: prognosis. Scottish Med J 1957; 2: 200-15.
20. Banks JL, Marotta CA. Outcomes validity and reliability of the Modified Rankin Scale: Implications for stroke clinical trials: A literature review and synthesis. Stroke 2007; 38: 1091-6.

21. Cohen J. Coefficient of agreement for nominal scales. Educ Psychol Meas 1960; 20: 37-46.
22. Fleiss JL. Statistical methods for rates and proportions. New York: John Wiley & Sons Inc; 1981: 229-232.
23. Landis RJ, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977; 33; 159-74.
24. Sun TK, Chiu SC, Yeh SH, Chang KC. Assessing reliability and validity of the Chinese version of the stroke scale: scale development. Int J Nurs Stud 2006; 43: 457-63.
25. Lyden P, Brott T, Tilley B, Welch KM, Mascha EJ, Levine S, et al. Improved reliability of the NIH Stroke Scale using video training. NINDS TPA Stroke Study Group. Stroke 1994; 25: 2220-6.

การศึกษาความเที่ยงตรงของแบบวัดความรุนแรงโรคหลอดเลือดสมอง National Institute of Health Stroke Scale ภาคภาษาไทย (NIHSS-T)

ยงชัย นิละนนท์, สุชีรา ภัทรายุตวรรตน์, พิพัฒน์ เชี่ยววิทย์, สงคราม โชติกอนุชิต, จุลจักร ลิ่มศรีวิไล, ลัลธริตา เจริญพงษ์, นภาพร แย้มแก้ว, นิรมล ลิรัฐพงค์, พรพิมล อันประเสริฐพร, จุฬาลักษณ์ โกมลตรี, นาราพร ประยูรวิวัฒน์, นิพนธ์ พวงวรินทร์

ภูมิหลัง: *National Institute of Health Stroke Scale (NIHSS) เป็นแบบวัดความรุนแรงของโรคหลอดเลือดสมองที่ได้รับการศึกษาความเที่ยงตรงและเป็นที่ยอมรับกันทั่วโลก ในประเทศไทยยังไม่มี NIHSS ภาคภาษาไทย*

วัตถุประสงค์: *เพื่อดัดแปลงและศึกษาความเที่ยงตรงของแบบวัดภาคภาษาไทย (NIHSS-T)*

วัสดุและวิธีการ: *คณะผู้วิจัยได้ดัดแปลงแบบวัด NIHSS เป็น NIHSS-T ตามคำแนะนำของ International Quality of Life Assessment Project Group โดยมีการแปลย้อนไปและกลับของสองภาษาจากนั้นทำการทดสอบความเที่ยงตรงของแบบวัดโดยเปรียบเทียบคะแนน NIHSS-T กับปริมาตรเนื้อสมองที่ขาดเลือด ซึ่งวัดจากภาพคลื่นแม่เหล็กไฟฟ้าสมองขณะแรกรับ และความพิการที่ 3 เดือนด้วย modified Rankin Scale (mRS) ผู้ป่วยทุกรายได้รับการประเมินด้วย NIHSS-T จากบุคคลากรทางการแพทย์ 3 กลุ่มคือ แพทย์เฟลโลว์ระบบประสาท 2 คน แพทย์ประจำบ้านอายุรศาสตร์ 2 คน และพยาบาลประจำหอผู้ป่วยโรคหลอดเลือดสมอง 2 คน จากนั้นวิเคราะห์ค่าคะแนนเฉลี่ยของ NIHSS-T จากผู้ประเมินทุกคน*

ผลการศึกษา: *ผู้ป่วยไทยที่เป็นโรคหลอดเลือดสมองตีบหรืออุดตันเฉียบพลันเข้าร่วมการศึกษา 32 คนอายุเฉลี่ย $\pm$ ค่าความเบี่ยงเบนมาตรฐานคือ 64.53 $\pm$ 14.97 ปี เพศชายร้อยละ 71.9 ค่าเฉลี่ยคะแนน NIHSS-T $\pm$ ค่าเบี่ยงเบนมาตรฐานเท่ากับ 7.49 $\pm$ 7.02 ความเชื่อมั่นของแบบวัดในผู้ประเมินคนเดียวกัน (Intra-observer reliability) คำนวณจาก Intra-class correlation (ICC) เท่ากับ 0.98, 0.98, 0.96, 0.98, 0.90 และ 0.98 สำหรับแพทย์เฟลโลว์ระบบประสาท, แพทย์ประจำบ้านอายุรศาสตร์ และพยาบาลประจำหอผู้ป่วย โรคหลอดเลือดสมองตามลำดับ ค่าความเชื่อมั่นระหว่างผู้ประเมิน (Inter-observer reliability) ทั้ง 6 คนคำนวณจาก ICC เท่ากับ 0.99 (0.98, 0.99) ค่าเที่ยงตรงของแบบวัดคำนวณจากค่าสัมประสิทธิ์สหสัมพันธ์สเปียร์แมน (Spearman rank correlation coefficients) ระหว่างคะแนน NIHSS-T กับปริมาตรเนื้อสมองที่ขาดเลือด และระดับความพิการที่ 3 เดือนเท่ากับ 0.53 ($p = 0.002$) และ 0.69 ($p < 0.001$) ตามลำดับ*

สรุป: *แบบวัด NIHSS-T มีค่าความเที่ยงตรงและความเชื่อมั่นสูงในการประเมินความรุนแรงของโรคหลอดเลือดสมองในผู้ป่วยไทย โดยบุคลากรทางการแพทย์ที่ได้รับการฝึกฝนแล้ว*