Pielęgniarstwo Neurologiczne i Neurochirurgiczne
THE JOURNAL OF NEUROLOGICAL AND NEUROSURGICAL NURSING

eISSN 2299-0321 ISSN 2084-8021 www.pnn.wshe.pl **Review**

DOI: 10.15225/PNN.2014.3.3.5

The NIHSS for International Neuroscience Nurses

Skala NIHSS do powszechnego zastosowania w neurologii przez pielęgniarki

Janice L. Hinkle

Self Employed Nurse Author & Editor

Abstract

Stroke is a worldwide problem with significant morbidity and mortality. The National Institutes of Health Stroke Scale (NIHSS) is a commonly used assessment tool. The NIHSS is used to assess patient status acutely, to assess treatment efficacy, and predict outcome in patients who speak many different languages. This article provides a review of the current state of research regarding validity and reliability of the English, Chinese, Hindi, and Spanish versions of the NIHSS used by international neuroscience nurses. **(JNNN 2014;3(3):129–135)**
Key Words: assessment, neuroscience nurses, NIHSS

Streszczenie

Udar mózgu stanowi ogólnoświatowy problem charakteryzujący się wysokim stopniem zachorowalności oraz umieralności. Skala The National Institutes of Health Stroke Scale (NIHSS) jest powszechnie stosowanym narzędziem oceny. Wykorzystywana jest do wnikliwej oceny stanu zdrowia pacjenta, efektywności leczenia oraz do przewidywania wyników u pacjentów mówiących różnymi językami. Niniejszy artykuł stanowi przegląd aktualnego stanu badań dotyczących przydatności oraz rzetelności angielskiej, chińskiej, hinduskiej oraz hiszpańskiej wersji Skali NIHSS, wykorzystywanych przez pielęgniarki w dziedzinie nauk neurologicznych na świecie. **(PNN 2014;3(3):129–135)**
Słowa kluczowe: ocena, pielęgniarki nauk neurologicznych, NIHSS

The National Institutes of Health Stroke Scale (NIHSS) is used worldwide by stroke neurologists, neuroscience nurses, and other health care professionals. It is most often used with acute ischemic stroke (AIS) but is also used to assess patients following hemorrhagic stroke or a suspected transient ischemic attack (TIA). The scale is used to assess patient status acutely, to assess treatment efficacy, and predict outcome. Some of the commonly used are English [1,2], Chinese [3], Hindi [4], and Spanish [5] versions of the scale. Other versions exist but are not the focus of this article.

Have you heard this story about reliability and validity? Joe walks into an ice cream shop every night at 5 PM and orders a milk shake. Every night at about 5:05 PM — the soda clerk gives him a milk shake that tastes just like it did the night before. Joe loves this because he knows that he can walk in this ice cream shop and always get the beverage that he likes and that tastes just like it did the night before. One day Joe invites a friend from work to join him and they both go into the shop and order what Joe thinks is the reliable milk shake but when it is delivered, his friend from work tastes the product and declares "this isn't a milk shake — it's an ice cream soda". This story illustrates the two concepts of validity and reliability. While the clerk was reliably delivering the same beverage night after night, he was not delivering what fit the definition of a milk shake therefore stating that the product was a milk shake was not valid [6].

For customers in an ice cream shop it may not make much difference that they were getting an ice cream soda and not a milk shake but for international neuroscience nurses measuring a physical concept such the level of impairment caused by a stroke, the instrument needs to have good validity and reliability. This article provides a review of the current state of research regarding

validity and reliability of the English, Chinese, Hindi, and Spanish versions of the NIHSS used by international neuroscience nurses.

English version

The English version of the NIHSS is used in the United Kingdom, Australia and the USA as well as other English speaking countries (See Table). The NIHSS was introduced as a 15 item scale, later reduced to 14 items developed by stroke neurologists from the University of Cincinnati, the University of Iowa and the NIH-NINDS, both in the USA [1,2,7]. This graded physical examination is a measure of impairment that assesses level of consciousness (LOC) using 3 items, speech, language, cognition, inattention, visual field abnormalities, motor strength (4 items), sensory impairment, and ataxia. It was designed specifically for AIS clinical trials and has been widely used in International clinical trials. Potential scores for the full scale range from 0 to 42.

Validity

Validity in behavioral measures refers to how well the instrument actually measures the construct it says it is measuring. In the example of the ice cream shop is the customer getting a milk shake or an ice cream soda? With an instrument that is supposed to measure disability, is it really measuring disability or it is measuring impairment? The validity of the NIHSS has been studied in many ways.

The clinical predictive validity of the NIHSS is useful to international neuroscience nurses. There have been a variety of studies in the acute phase of stroke that establish predictive validity. In one study of 643 patients with AIS an initial total NIHSS score of equal to or greater than 8 points was predictive of neurologic deteriorations within the first week of the clinical course of the patient [8]. In the NINDS IV rt-PA trial patients with an NIHSS total score of greater than 22 had a 17% risk of intracranial hemorrhage (ICH) while patients with a total NIHSS score of less than 10 had

Table. Summary of national institutes of health stroke scale (NIHSS)

Category	Description	Score
1a. Level of consciousness (LOC)	Alert	0
	Arousable by minor stimulation	1
	Obtunded, strong stimulation to attend	2
	Unresponsive, or reflexic responses only	3
1b. LOC questions (month, age)	Answers both correctly	0
	Answers one correctly	1
	Both incorrect	2
1c. LOC commands (open, close eyes; make fist, let go)	Obeys both correctly	0
	Obeys one correctly	1
	Both incorrect	2
2. Best gaze (eyes open — patient follows examiner's finger or face)	Normal	0
	Partial gaze palsy	1
	Forced deviation	2
3. Visual (introduce visual stimulus/threat to patient's visual field quadrants)	No visual loss	0
	Partial hemianopsia	1
	Complete hemianopsia	2
	Bilateral hemianopsia	3
4. Facial palsy (show teeth, raise eyebrows and squeeze eyes shut)	Normal	0
	Minor	1
	Partial	2
	Complete	3
5a. Motor; arm — left (elevate extremity to 90° and score drift/movement)	No drift	0
	Drift but maintains in air	1
	Unable to maintain in air	2
	No effort against gravity	3
	No movement	4
	Amputation, joint fusion (explain)	N/A

5b.	Motor; arm — right (elevate extremity to 90° and score drift/movement)	No drift	0
		Drift but maintains in air	1
		Unable to maintain in air	2
		No effort against gravity	3
		No movement	4
		Amputation, joint fusion (explain)	N/A
6a.	Motor; leg — left (elevate extremity to 30° and score drift/movement)	No drift	0
		Drift but maintains in air	1
		Unable to maintain in air	2
		No effort against gravity	3
		No movement	4
		Amputation, joint fusion (explain)	N/A
6b.	Motor; leg — right (elevate extremity to 30° and score drift/movement)	No drift	0
		Drift but maintains in air	1
		Unable to maintain in air	2
		No effort against gravity	3
		No movement	4
		Amputation, joint fusion (explain)	N/A
7.	Limb ataxia (finger-to-nose and heel-to-shin testing)	Absent	0
		Present in one limb	1
		Present in two limbs	2
8.	Sensory (pinprick to face, arm, trunk, and leg — compare side to side)	Normal	0
		Mild to moderate loss	1
		Severe to total loss	2
9.	Best language (name items, describe a picture and read sentences)	No aphasia	0
		Mild to moderate aphasia	1
		Severe aphasia	2
		Mute	3
10.	Dysarthria (evaluate speech clarity by having patient repeat words)	Normal	0
		Mild to moderate dysarthria	1
		Severe dysarthria, mostly unintelligible or worse	2
		Intubated or other physical barrier	N/A
11.	Extinction and inattention (use information from prior testing to score)	No abnormality	0
		Visual, tactile, auditory, or other extinction to bilateral simultaneous stimulation	1
		Profound hemiattention or extinction to more than one modality	2
	Total score		

Adapted from the version available at the National Institute of Neurological Disorders and Stroke, National Institutes of Health, Bethesda, MD 20892, www.ninds.nih.gov/doctors/NIH_Stroke_Scale.pdf. It is recommended that the full scale with all instructions be used.

a 3% risk of hemorrhage [9]. One small study looked at the NIHSS items and found that increases in the loss of consciousness and motor limbs total scores were related to neurological deterioration within the first 120 minutes after administration of IV rt-PA [10]. Another study reported that an improvement in the total NHISS of greater than 3 points at 15 minutes or of greater than 5 points at 30 minutes predicts a more favorable outcome and helped identify patients who are not responding to rt-PA [11]. A Swedish study of 347 patients found that for each point on the baseline NIHSS there was an increase in length of stay (LOS) by 0.8 days and an increase

in the total LOS (including rehabilitation) by 3.4 days up to a NIHSS of 19 points [12]. The NIHSS was also better at predicting the presence of dysphagia compared to a nursing dysphagia screening tool in a retrospective cohort study of veterans admitted with AIS [13].

Several studies assessed outcome after hospitalization. In a study of 385 patients 3 months after a stroke a total NIHSS score of equal to or greater than 15 points was strongly associated with the patient being dependent (in a nursing home, chronic home, or substantially dependent on a caregiver) [14]. One study reported that patients with a NIHSS total score of greater than 10 have a more favorable outcome at one year compared to patients with an NIHSS score of greater than 20 [15]. In a community based sample of 377 patients scoring less than a 4 on baseline NIHSS, 75% were independent one year after the stroke, 17% were dependent and 8% were dead [16]. Predicting outcome makes the NIHSS useful for international neuroscience nurses working with families on discharge planning needs [8,17,18].

Factor Analysis is a statistical process that is used to establish how individual items cluster around a dimension. Two factors were found in factor analysis of the NIHSS corresponding to left and right brain function in a sample of mild to moderate stroke [19] and another of patients with large strokes [20]. In both studies patients with left-brain strokes score 4 points higher on the NIHSS compared to right brain strokes [19,20].

Reliability

An important aspect of reliability of the NIHSS is the ability of the measure to produce the same results when used by different individuals. This is called inter-rater or inter-observer reliability sometimes reported using an intra-class correlation coefficient (ICC). The kappa statistic or kappa coefficient is also commonly used to indicate the magnitude of agreement between observers. A kappa of 1 indicates complete agreement while a kappa of 0 indicates the agreement is due to chance. A kappa can be reported as weighted or unweighted [21].

An important aspect of inter-rater reliability is agreement on total scores. The overall reliability of the NIHSS in the context of stroke clinical trials is well established [22]. One study assessed the reliability of the NIHSS when administered by research nurses [23]. Using the NIHSS 31 consecutive patients who had a stroke were assessed by 2 neurologists as well as 1 of 2 trained research nurses. There was a high level of agreement for total scores between the 2 neurologists (ICC=0.95) and between each neurologist and research nurse (ICC =0.92 and 0.96). These researchers concluded that in both hospital and community settings trained research nurses can administer the NIHSS with reliability similar to stroke-trained neurologists [23].

Another study investigated the reliability of the NIHSS in a representative sample of raters who had completed videotaped certification examinations on the National Stroke association website between 1998 and 2004. There were 7,405 unique raters and more than half (54%) of them were registered nurses (RNs) [24]. The nurses agreed with the most common response on scoring more frequently than physicians ($p<0.0001$) and 91% of the RNs passed the certification examination [24].

An additional aspect of inter-rater reliability is agreement on individual item scores within the scale. One study reported moderate to excellent agreement (weighted *kappa* greater than 4) overall between neurologists and research nurses on the majority of the NIHSS items but poor agreement on the one item of limb ataxia [22]. Another study with a total of 38, 148 individual NIHSS item responses reported poor agreement on 4 items [23]. In a large sample of clinicians the item of limb ataxia, agreement was extremely low while the three items assessing gaze, aphasia, and facial weakness were low using an unweighted kappa statistic [24]. Another study using updated cases reported low agreement on the ataxia and facial weakness items [25].

Chinese Version

The English version of the NIHSS has been in use for several decades [1,2]. The Chinese National Institute of Health Stroke Scale (C-NIHSS) was first translated and back translated in 2006 [3]. One study used 6 clinicians (3 nurses and 3 doctors) to assess 48 patients using the C-NIHSS and test the validity and reliability of the scale.

Validity

Ten experts evaluated the appropriateness of the translated items. For the item on motor arm and leg movement, the translation of the word “drift” into “swaying” received a low from three experts and was, thus, changed to “shaking” in Chinese. Based on the experts’ assessment, the index of C-NIHSS content validity reached 1.00 following minor revisions [3].

Reliability

Inter-rater agreement was tested among three paired groups of physicians. The kappa statistic for all items but two in all three physician groups was high (greater than 0.60). Facial palsy and limb ataxia received kappa values of 0.47–0.58 in two of the paired groups. Nurses were

paired into three groups to test inter-rater agreement as well. Only one item, limb ataxia, received a kappa value less than 0.59 (*kappa*=0:52) in one group. Kappa values on other items were all above 0.60. When all six raters were randomly grouped into pairs, nine groups were formed. One item, facial palsy, received consistently low kappa values. Eight groups had low kappa values (0.28–0.59) and of these two groups had low values (0.39). The consistently lowest item (extinction and inattention), had kappa values below 0.59 in all 9 groups, 7 of which were below 0.39 [3].

Hindi Version

India has a population of over one billion people; therefore a Hindi version of the NIHSS is very appropriate. One article about the HV-NIHSS was located and it explained how it was developed for a stroke clinical trial. Sentences and words in items 9 and 10 were substituted with Hindi words and phrases. The material used to evaluate aphasia and dysarthria were culturally adapted. The pictures in the English version of the hammock, cactus, feather and glove were replaced with pictures of a balloon, lock, ball, spoon, tiger, bus, telephone, tire, and rose. The cookie jar picture was replaced with a picture of a lady cooking over a traditional stove [4].

Validity

There was good correlation between the English version of the NIHSS and the HV-NIHSS in 38 patients who knew both English and Hindi. Prasad and colleagues also assessed the construct validity and predictive validity of the HV-NIHSS in a group of patients admitted to a referral hospital in Delhi who had a stroke. Higher baseline scores on the HV-NIHSS predicted poorer functional outcome at 3 months. The HV-NIHSS had negative correlations with both the Barthel at 3 months and the Glasgow Coma Scale (GCS) at baseline indicating concurrent construct validity [4].

Reliability

Prasad and colleagues tested the inter-rater reliability of the HV-NIHSS by having 2 investigators rate 107 patients. The Pearson *r* ranged from 0.72 to 0.99 on the 15 items of the scale. The total score of the HV-NIHSS was 0.995, (95% CI=0.993–0.997)[4].

Spanish Version

The Spanish language is another widely used language. Therefore a Spanish version of the NIHSS (SV-NIHSS) has appropriately been adapted and tested.

Validity

Construct validity was adequate when tested in 102 patients. The SV-NIHSS had a negative correlation with baseline GC S (Spearman coefficient =-0.574, P<.001) and with Barthel index at 3 months (Spearman coefficient =-0.658, P<.001). Patients with different Rankin scores at 3 months also had significantly different baseline SV-NIHSS scores, from a mean of 4.29±2.21 for Rankin score of 0 to a mean of 29.40±3.97 for Rankin score of 6 (P<.001) [5].

Reliability

Interrater reliability was independently evaluated for 98 patients, showing a high agreement with Kappa's of 0.77 to 0.99 for the 15 items [5]. The interrater correlation coefficient was 0.991 (95% confidence interval, 0.987–0.994). Intrarater reliability was excellent with a Kappa statistic ranging from 0.86 to 1.00 for the 15 items. The mean intrarater correlation coefficient was 0.994 (95% confidence interval, 0.991–0.996) between the 8 physicians who took part in the study [5]. A weakness of this study was that no nurses were included.

Learning the NIHSS

Training and certification in use of the NIHSS is recommended for International neuroscience nurses. The English version of the NIHSS has been taught in person, with videotapes, DVDs and there are professionally filmed cases for training in at least two places (www.strokeassociation.com or www.ninds.gov). Instruction takes 2 to 3 hours. There is an International Electronic Education Network (IEEN)® located at (www.HealthCarePoint.com) that includes instruction on the English, Portuguese, Italian and Spanish versions of the NIHSS. The company, Genentech (http://www.activase.com) has laminated cards in English and Spanish, for order, free of charge for use by international neuroscience nurses.

The best method of learning the scale is still an area of debate and ongoing research. One study compared the use of videotaped instruction to an interactive computer assisted instruction to teach the C-NIHSS [26]. The nurses with less experience in neurological nursing performed a better assessment following the interactive

computer assisted instruction compared to nurses who were taught with an instructor led program using the videotape [26]. More research is needed to determine the best method of instruction for other versions of the NIHSS.

Conclusion

International neuroscience nurses need to base interventions on evidence and to do so it is important to use valid and reliable assessment scales. The English version of the NIHSS has been in use the longest and has good validity and reliability. The C-NIHSS, the HV-NIHSS and the SV-NIHSS all have good beginning research to establish validity and reliability but there are gaps in the research that could be filled by international neuroscience nursing research. The SV-NIHSS, for example, was only tested with physicians. Another area for further research is the best method of education to use with practitioners learning the scale.

References

[1] Brott T.G., Adams H.P., Olinger C.P. et al., Measurements of acute cerebral infarction: A clinical examination scale. *Stroke*. 1989;20(7):864–870.

[2] Hinkle J.L. Reliability and validity of the National Institutes of Health Stroke Scale for neuroscience nurses. *Stroke*. 2014;45(3);e32–34.

[3] Sun T.K., Chiu S.C., Yeh S.H., Chang K.C. Assessing reliability and validity of the Chinese version of the stroke scale: scale development. *International Journal of Nursing Studies*. 2006;43(4):457–463.

[4] Prasad K., Dash D., Kumar A. Validation of the Hindi version of National Institute of Health Stroke Scale. *Neurology India*. 2012;60(1):40.

[5] Domínguez R., Villa J.F., Augustovski F. et al. Spanish Cross-Cultural Adaptation and Validation of the National Institutes of Health Stroke Scale. *Mayo Clinic Proceedings.* 2006;81(4):476–480.

[6] Buelow J.M., Hinkle J.L. Why are reliability and validity important to neuroscience nurses? *Journal of Neuroscience Nursing*. 2008;40(6):369–70,372.

[7] Brott T.G., Reed R.L. Intensive care for acute stroke in the community hospital setting. The first 24 hours. *Stroke*. 1989;20(5):694–697.

[8] Miyamoto N., Tanaka Y., Ueno Y. et al. Demographic, clinical, and radiologic predictors of neurologic deterioration in patients with acute ischemic stroke. *Journal of Stroke and Cerebrovascular Diseases*. 2013; 22(3):205–210.

[9] Summers D., Leonard A., Wentworth D. et al. American Heart Association Council on Cardiovascular Nursing and the Stroke Council, Comprehensive overview of nursing and interdisciplinary care of the acute ischemic stroke patient: a scientific statement from the American Heart Association. *Stroke*. 2009;40(8):2911–2944.

[10] Nanri Y., Yakushiji Y., Hara M. et al. Stroke scale items associated with neurologic deterioration within 24 hours after recombinant tissue plasminogen activator therapy. *Journal of Stroke and Cerebrovascular Diseases*. 2013;22(7): 1117–1124.

[11] Takagi T., Kato T., Sakai H., Nishimura Y. Early Neurologic Improvement Based on the National Institutes of Health Stroke Scale Score Predicts Favorable Outcome within 30 Minutes after Undergoing Intravenous Recombinant Tissue Plasminogen Activator Therapy. *Journal of Stroke and Cerebrovascular Diseases*. 2014;23(1): 69–74.

[12] Appelros P. Prediction of length of stay for stroke patients. *Acta Neurologica Scandinavica*. 2007;116(1):15–19.

[13] Bravata D.M., Daggett V.S., Woodward-Hagg H. et al. Comparison of two approaches to screen for dysphagia among acute ischemic stroke patients: nursing admission screening tool versus National Institutes of Health stroke scale. *Journal Of Rehabilitation Research And Development*. 2009;46(9):1127–1134.

[14] Johnston K.C., Wagner D.P. Relationship between 3-month National Institutes of Health Stroke Scale score and dependence in ischemic stroke patients. *Neuroepidemiology*. 2006:27(2):96–100.

[15] Kwiatkowski T.G., Libman R.B., Frankel M. et al. Effects of tissue plasminogen activator for acute ischemic stroke at one year. National Institute of Neurological Disorders and Stroke Recombinant Tissue Plasminogen Activator Stroke Study Group. *New England Journal of Medicine*. 1999;340(23):1781–1787.

[16] Appelros P., Terent A. Characteristics of the National Institute of Health Stroke Scale: results from a population-based stroke cohort at baseline and after one year. *Journal of Stroke and Cerebrovascular Diseases*. 2004;17(1): 21–27.

[17] Schlegel D., Kolb S.J., Luciano J.M. et al. Utility of the NIH Stroke Scale as a predictor of hospital disposition. *Stroke*. 2003;34(1):134–137.

[18] Schlegel D., Tanne D., Demchuk A.M., Levine S.R., Kasner S.E. Prediction of hospital disposition after thrombolysis for acute ischemic stroke using the National Institutes of Health Stroke Scale. *Archives of Neurology*. 2004;61(7):1061–1064.

[19] Lyden P., Lu M., Jackson C. et al. Underlying structure of the National Institutes of Health Stroke Scale — Results of a factor analysis. *Stroke*. 1999;30(11):2347–2354.

[20] Lyden P., Claesson L., Havstad S., Ashwood T., Lu M. Factor analysis of the National Institutes of Health Stroke Scale in patients with large strokes. *Archives of Neurology.* 2004;61(11):1677–1680.

[21] Viera A.J., Garrett J.M. Understanding interobserver agreement: the kappa statistic. *Family Medicine*. 2005; 37(5):360–363.

[22] Goldstein L.B., Samsa G.P. Reliability of the National Institutes of Health Stroke Scale — Extension to non-neurologists in the context of a clinical trial. *Stroke*. 1997; 28(2):307–310.

[23] Dewey H.M., Donnan G.A., Freeman E.J. et al. Inter-rater reliability of the National Institutes of Health Stroke Scale: Rating by neurologists and nurses in a com-

munity-based stroke incidence study. *Cerebrovascular Diseases*. 1999;9(6):323–327.
[24] Josephson S.A., Hills N., Johnston S.C. NIH Stroke Scale reliability in ratings from a large sample of clinicians. *Cerebrovascular diseases*. 2006;22(5–6):389–395.
[25] Lyden P., Raman R., Liu L. et al. NIHSS training and certification using a new digital video disk is reliable. *Stroke*. 2005;36(11):2446–2449.
[26] Chiu S., Cheng K.Y., Sun T.K. et al. The effectiveness of interactive computer assisted instruction compared to videotaped instruction for teaching nurses to assess neurological function of stroke patients: a randomized controlled trial. *International journal of nursing studies*. 2009;46(12):1548–1556.

Corresponding Author:
Janice L. Hinkle
641 Massachusetts Ave N.E.
Washington DC. 20002, USA
e-mail: janhinkle@hotmail.com

Conflict of Interest: None
Funding: None
Author Contributions: Janice L. Hinkle[A, B, C, E, F]
(A — Concept and design of research, B — Collection and/or compilation of data, C — Analysis and interpretation of data, E — Writing an article, F — Search of the literature
Received: 17.06.2014
Accepted: 11.08.2014