Development of the Italian Version of the National Institutes of Health Stroke Scale : It-NIHSS

Francesca Romana Pezzella, Orietta Picconi, Assunta De Luca, Patrick D. Lyden and Marco Fiorelli

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/40/7/2557

Development of the Italian Version of the National Institutes of Health Stroke Scale
It-NIHSS

Francesca Romana Pezzella, MD, PhD; Orietta Picconi, BS; Assunta De Luca, MD;
Patrick D. Lyden, MD, FAAN; Marco Fiorelli, MD, PhD

Background and Purpose—The National Institutes of Health Stroke Scale (NIHSS) is a basic component of the assessment of patients with acute stroke. To foster and standardize the use of the NIHSS among Italian health professionals, we translated the scale, dubbed into Italian the training and test videotapes devised by the National Institutes of Health researchers, and conducted a series of certification courses using the translated videos.

Methods—Translation, text adaptation, video dubbing, and editing of the Italian NIHSS videotapes relied on a team of bilingual stroke neurologists. Three waves of training courses were organized for mixed classes of medical and nonmedical health professionals. The certification test was based on the usual set of 5 videotaped patients. Scoring rules were those provided by the National Institutes of Neurological Disorders and Stroke. Reliability of the Italian NIHSS was assessed using kappa statistics and compared with that of the original NIHSS.

Results—During 3 years, 850 nurses, 460 nonneurologist physicians, and 246 neurologists were trained. Pass rates were respectively 44%, 75%, and 87%, respectively. Overall, 80% of scale items showed moderate to excellent reliability. Independent significant predictors of test failure at multivariate logistic regression were nurse profession (OR, 5.41; 95% CI, 4.07 to 7.20), older age (OR, 1.03; 95% CI, 1.02 to 1.05), and first edition of the course (OR, 3.13; 95% CI, 2.43 to 4.05). The agreement across all items between NIHSS and the Italian NIHSS was 80% (kappa=0.70±0.18, z<0.001).

Conclusions—The Italian translation, supervised by experienced vascular neurologists, did not influence the clinimetric characteristics of the NIHSS. Our findings support the implementation of NIHSS video training in languages other than English. (***Stroke***. 2009;40:2557-2559.)

Key Words: observer variation ■ stroke assessment ■ stroke scales

The National Institute of Health Stroke Scale (NIHSS) is the standard tool for clinical evaluation of acute stroke. With the exposure to a structured training program, the reliability of the scale increases; hence, training and certification have become prerequisite for the use of the NIHSS in therapeutic trials.[1–3] Because the NIHSS is increasingly recommended to quantify stroke severity also in the clinical setting,[4–8] non-English-speaking health professionals are faced with the lack of training aids in their own language.

In 2003, the Department of Neurological Sciences of the Sapienza University of Rome and the Public Health Agency of the Lazio Region promoted a series of integrated initiatives for the improvement of acute stroke care in the community,[9] including the mandate to record the grade of neurological impairment on emergency department arrival using the NIHSS.

In this article, we report our experience with the development of a training and certification system for the NIHSS in Italian.

Materials and Methods

The instructional videotape "The Quick and Easy NIHSS,"[10] and the NIHSS Certification Tape 1 were translated into Italian (It-NIHSS) by 3 bilingual stroke neurologists who checked each other's translation and discussed all the discrepancies. A speech therapist was consulted to build an adequate list of words and sentences to test patients' speech disorders (Item 8 and Item 9). Professional text adaptation, video dubbing, and editing were supervised by a stroke neurologist (F.R.P.) at all times. It-NIHSS training sessions were held for nurses, emergency department physicians (nonneurologist), and neurologists over a 3-year period. For the first edition (2002 to 2003), the It-NIHSS training and certification was part of an 8-hour interactive update course on stroke accredited by the Italian Continuing Medical Education commission, whereas for the second

Table 1. Interrater Agreement of It-NIHSS Scores by Profession

Scale Item	Nurses		Non-Neurologist Physicians		Neurologists		All	
	Kappa	P	Kappa	P	Kappa	P	Kappa	P
Level of consciousness (LOC)	0.461	0.000	0.615	0.000	0.451	0.000	0.511	0.000
LOC questions	0.760	0.000	0.764	0.000	0.776	0.000	0.763	0.000
LOC commands	0.011	0.000	0.005	0.000	0.050	0.021	0.009	0.000
Gaze	0.038	0.000	0.017	0.000	0.029	0.000	0.031	0.000
Visual fields	0.523	0.000	0.662	0.000	0.784	0.000	0.506	0.000
Facial paresis	0.571	0.000	0.506	0.000	0.470	0.000	0.522	0.000
Motor arm right	0.824	0.000	0.861	0.000	0.847	0.000	0.837	0.000
Motor arm left	0.825	0.000	0.797	0.000	0.861	0.000	0.811	0.000
Motor leg right	0.622	0.000	0.613	0.000	0.762	0.000	0.620	0.000
Motor leg left	0.804	0.000	0.736	0.000	0.716	0.000	0.775	0.000
Limb ataxia	0.308	0.000	0.237	0.000	0.375	0.000	0.267	0.000
Sensory	0.676	0.000	0.792	0.000	0.794	0.000	0.714	0.000
Language	0.482	0.000	0.506	0.000	0.639	0.000	0.475	0.000
Dysarthria	0.697	0.000	0.787	0.000	0.741	0.000	0.723	0.000
Hemi-inattention	0.761	0.000	0.781	0.000	0.744	0.000	0.746	0.000

(2003 to 2004) and third editions (2005 to 2006), only the 2 It-NIHSS videotapes were presented to groups of no more than 35 people at a time.

The answers to the certification test were archived in a database along with age, profession, and site of work (university versus community hospital) of the rater, date of session, and course edition. Using SPSS (Version 8.0; Chicago, Ill), we performed a multivariate analysis to evaluate the influence of participants' characteristics on the probability to pass the certification test.

The agreement among raters was evaluated using the unweighted kappa statistics[11–15] (Version 6.0; Stata, College Station, Texas) for the case of multiple raters who examine a few subjects. To ensure comparability of results, we adopted the same statistical methodology for the estimation of the agreement as that published by the National Institute of Neurological Diseases and Stroke investigators.[1] Concordance was considered poor for kappa <0.40, moderate for kappa between 0.40 and 0.75, and excellent for kappa >0.75.

Results

From October 2002 to January 2006, 1556 healthcare professionals were exposed to the training and attempted the certification. There were 850 nurses, 460 nonneurologist physicians, and 246 neurologists. Course participants' average age was 41 years with neurologists being the oldest professional group (47 years). Pass rate (ie, the proportion of raters having scored <5 outlier responses across the 5 videotaped patients, an outlier being a response given by ≤12% of the original cohort of National Institute of Neurological Diseases and Stroke investigators[1]), differed significantly among the speciality subgroups: 44% of nurses, 75% of nonneurologist physicians, and 87% of neurologists were certified on the first attempt (*P*<0.05). For comparative analyses, we considered the performance of the latter group as the gold standard. The multivariate logistic regression showed that nurse profession (OR, 5.41; 95% CI, 4.07 to 7.20), older age (OR, 1.03; 95% CI, 1.02 to 1.05), and having attended the first edition of the course (OR, 3.13; 95% CI, 2.43 to 4.05) were significant independent predictors of test failure. The area under the receiver operating characteristic curve was 0.73 for this model.

The computed kappa and its probability value for each item by professional group is shown in Table 1. Overall, the agreement was poor for 3 items (gaze, level of consciousness commands, and ataxia), moderate for 8 items, and excellent for the remaining 4 items (motor function and orientation). Table 2 shows the overall agreement beyond chance between the original NIHSS and the It-NIHSS obtained by comparing the historical sample of NIHSS raters[13] with the performance of the Italian neurologists. The observed agreement was 80% with a kappa of 0.70 (z<0.001).

Table 2. Agreement Between the Original NIHSS and the It-NIHSS

Scale Item	It-NIHSS	NIHSS
Level of consciousness (LOC)	Moderate	Moderate
LOC questions	Excellent	Moderate
LOC commands	Poor	Poor
Gaze	Poor	Poor
Visual fields	Excellent	Excellent
Facial paresis	Moderate	Poor
Motor arm right	Excellent	Excellent
Motor arm left	Excellent	Excellent
Motor leg right	Excellent	Excellent
Motor leg left	Moderate	Moderate
Limb ataxia	Poor	Poor
Sensory	Excellent	Excellent
Language	Moderate	Poor
Dysarthria	Moderate	Moderate
Hemi-inattention	Moderate	Moderate

Agreement: 80.00%; expected agreement: 33.33%; kappa: 0.7000; SE: 0.1789.

Discussion

To standardize and foster the use of the NIHSS among Italian health professionals, we considered various approaches. We opted for video training because it is convenient and has a well-documented efficacy. Rather than developing a new Italian video, we deemed more cost-effective translating into Italian the original video training system. This study included more than 1500 individuals and is the largest NIHSS group of raters examined so far for educational research purposes. Compared with the original scale, the It-NIHSS exhibited good validity and similar reliability. Poor interrater agreement was limited to the same items that showed low values of kappa in the original NIHSS assessment.

Rates of success at certification test were good for physicians, but less so among nurses. Unlike physicians, Italian nurses have a weak theoretical background in neurology. For the vast majority of them, NIHSS examination was a previously unheard-of tool. However, nurse certification rates appeared to be improving with ongoing training.

Pass rates increased across the 3 waves of courses. During the first edition, the It-NIHSS was part of a complex and tiring 1-day 8-hour course, suggesting that raters may have been fatigued when performing the test. In the second and third edition, classes were less numerous, and the course was shorter and more focused on the It-NIHSS certification. Because pass rates were significantly higher, we now only train and certify in this way.

Further research should be directed at new training strategies to increase the probability of health professionals to use the NIHSS correctly. The use of this scale also needs to be expanded globally. A translation supervised by experienced vascular neurologists did not seem to have a negative impact on the reliability and the validity of the NIHSS. Although the implications of our experience cannot be regarded as universal, our data support the use of NIHSS video training in languages other than English.

Sources of Funding

Disclosures

References

1. Lyden P, Brott T, Tilley B, Welch KM, Mascha EJ, Levine S, Haley EC, Grotta J, Marler J. Improved reliability of the NIH Stroke Scale using video training. *Stroke*. 1994;25:2220–2226.
2. Schmulling S, Grond M, Rudolf J, Kiencke P. Training as a prerequisite for reliable use of NIH Stroke Scale. *Stroke*. 1998;29:1258–1259.
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.
4. Kasner SE, Chalela JA, Luciano JM, Cucchiara BL, Raps EC, McGarvey ML, Conroy MB, Localio AR. Reliability and validity of estimating the NIH Stroke Scale score from medical records. *Stroke*. 1999;30: 1534–1537.
5. Williams LS, Yilmaz EY, Lopez-Yunez AM. Retrospective assessment of initial stroke severity with the NIH Stroke Scale. *Stroke*. 2000;31: 858–862.
6. 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.
7. Albers GW, Bates VE, Clark WM, Bell R, Verro P, Hamilton SA. Intravenous tissue-type plasminogen activator for treatment of acute stroke: the Standard Treatment with Alteplase to Reverse Stroke (STARS) study. *JAMA*. 2000;283:1145–1150.
8. Grond M, Stenzel C, Schmulling S, Rudolf J, Neveling M, Lechleuthner A, Schneweis, Heiss WD. Early intravenous thrombolysis for acute ischemic stroke in a community-based approach. *Stroke*. 1998;29:1544–1549.
9. De Luca A, Agabiti N, Fiorelli M, Sacchetti ML, Tancioni V, Picconi O, Cardo S, Guasticchi G. Implementation of a surveillance system for stroke based on administrative and clinical data in the Lazio Region (Italy): methodological aspects. *Ann Ig*. 2003;15:207–214.
10. 'Quick & Easy' NIH Stroke Scale and Tutorial Program. SNR Resources, Division of The Stroke Group, Inc; 2000.
11. Cyr L, Francis K. Measures of clinical agreement for nominal and categorical data: the kappa coefficient. *Comput Biol Med.* 1992;22: 239–246.
12. Gross ST. The kappa coefficient of agreement for multiple observer when the number of subjects is small. *Biometrics*. 1986;42:883–893.
13. Cohen J. A coefficient of agreement for nominal scales. *Educ Psychol Meas*. 1960;20:37–46.
14. Fleiss JL. *Statistical Methods for Rates and Proportions*, II ed. New York: Wiley; 1981.
15. Byrt T. How good is that agreement? *Epidemiology*. 1996;7:561.