

Validation of the National Institutes of Health Stroke Scale, Modified Rankin Scale and Barthel Index in Brazil: The Role of Cultural Adaptation and Structured Interviewing

Carolina Cincura^a Octavio M. Pontes-Neto^c Iuri S. Neville^a
Henrique F. Mendes^c Daniela F. Menezes^a Débora C. Mariano^c
Issana F. Pereira^a Larissa A. Teixeira^c Pedro A.P. Jesus^a Danilo C.L. de Queiroz^c
Davidson F. Pereira^a Elen Pinto^a João P. Leite^c Antonio A. Lopes^b
Jamary Oliveira-Filho^a

^aStroke Clinic and ^bEpidemiology Unit of the Federal University of Bahia, Salvador, and ^cNeurology Service of Ribeirão Preto Medical School, University of São Paulo, São Paulo, Brazil

Key Words

Stroke · National Institutes of Health Stroke Scale · Modified Rankin Scale · Barthel Index · Stroke trials, Brazil

Abstract

Background: We aimed to validate three widely used scales in stroke research in a multiethnic Brazilian population. **Methods:** The National Institutes of Health Stroke Scale (NIHSS), modified Rankin Scale (mRS) and Barthel Index (BI) were translated, culturally adapted and applied by two independent investigators. The mRS was applied with or without a previously validated structured interview. Interobserver agreement (kappa statistics) and intraclass correlation coefficients were calculated. **Results:** 84 patients underwent mRS (56 with and 28 without a structured interview), 57 BI and 62 NIHSS scoring. Intraclass correlation coefficient was 0.902 for NIHSS and 0.967 for BI. For BI, interobserver agreement was good (kappa = 0.70). For mRS, the structured interview improved interobserver agreement (kappa = 0.34 without a structured interview; 0.75 with a structured inter-

view). **Conclusion:** The NIHSS, BI and mRS show good validity when translated and culturally adapted. Using a structured interview for the mRS improves interobserver concordance rates.

Copyright © 2008 S. Karger AG, Basel

Introduction

Stroke is the main cause of death and disability in Brazil, affecting over 200,000 individuals annually [1]. Scales which measure the impact of stroke are important to implement preventive and treatment strategies tailored to a specific population. Stroke investigators in Brazil have used different translated versions of these stroke scales [2–4]. However, simple translation misses cultural peculiarities which may affect scoring a particular scale. Thus, cross-cultural validation and adaptation is frequently necessary before widely using a new scale. This is especially important in multicenter studies using clinical scales as primary endpoints.

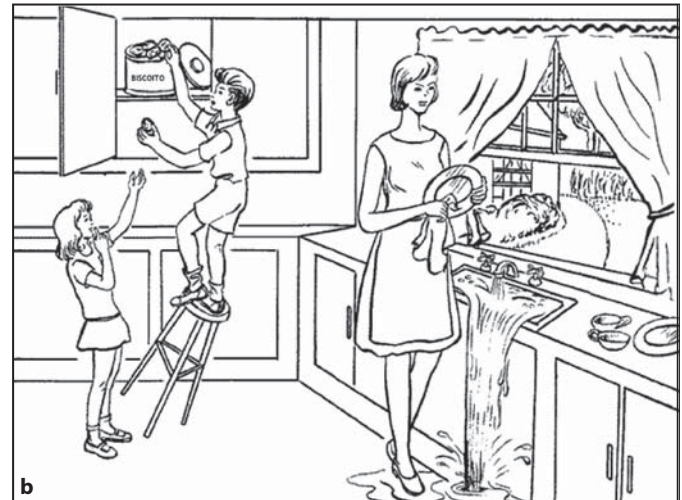


Fig. 1. a, b Material used for evaluation of aphasia.

The National Institutes of Health Stroke Scale (NIHSS), modified Rankin Scale (mRS) and Barthel Index (BI) are widely applied scales in stroke research and have previously been used in the Brazilian literature. In particular, the NIHSS uses standardized figures and phrases to test for aphasia and dysarthria, items especially vulnerable to cultural influences. We therefore aimed to validate translated and culturally adapted versions of the NIHSS, mRS and BI in two tertiary university-based hospitals in Brazil.

Subjects and Methods

Patients with a diagnosis of stroke were prospectively recruited from two university-based academic centers (Federal University of Bahia and São Paulo University at Ribeirão Preto, both in Brazil) between July 2005 and July 2006. Patients were selected from the outpatient clinics as well as from the admitting wards. Stroke was defined as a new focal neurological deficit lasting over 24 h and confirmed by head CT imaging. Patients were evaluated by two independent investigators on the same day. For the mRS and BI, one investigator was a physical therapist and the other was either a medical student or a neurology resident. For the NIHSS, one investigator was a neurologist and the other either a medical student or a neurology resident. The order of evaluations was random.

Each investigator was trained and certified for NIHSS application through the American Stroke Association's Online NIH Stroke Scale Training Program (www.asa.trainingcampus.net/UAS/Modules/TREES/windex.aspx). The NIHSS [5] was translated from English to Portuguese and from Portuguese to English to allow for detection of inconsistencies. The material used for evaluation of aphasia and dysarthria (figures and phrases) was

culturally adapted while maintaining the particular phonemes for dysarthria evaluation. For example, the hammock, the feather, the cactus, the key and the glove figures were replaced by pictures of a banana, a car, a horse, scissors and a coconut palm for aphasia evaluation (fig. 1–3). A videotaped training version was created for the Portuguese version of the NIHSS (PV-NIHSS) and used in both centers before starting data collection.

For the mRS and BI, a meeting was performed with all investigators to standardize scoring criteria. For the BI, scores were compared in categories of complete independence, slight dependence, moderate dependence, severe dependence and complete dependence, as previously defined [6]. For a group of patients, a structured interview was used for mRS scoring [7], but scale certification was not readily available or required for investigators to participate in the study. Thus, for the mRS, interobserver agreement was calculated for evaluations with and without a structured interview.

For statistical analyses, we used kappa statistics to compare interobserver agreement for each scale. For the PV-NIHSS, agreement for each scale item was calculated separately using kappa statistics. For interobserver agreement rating, we considered <0.20 as a poor association; 0.20–0.39 as fair; 0.40–0.59 as moderate; 0.60–0.79 as good, and >0.8 as excellent. We calculated intraclass correlation coefficients for each scale and considered the following scores: <0.75 = poor to moderate reliability; 0.75–0.90 = good reliability; >0.90 = excellent reliability. To investigate the correlation between each scale, we used the Spearman test.

Results

Over a 1-year period, 132 patients were evaluated. Of these patients, 84 underwent mRS scoring (28 without a structured interview and 56 with a structured interview),

2

Você sabe como fazer.
De volta pra casa.
Eu cheguei em casa do trabalho.
Próximo da mesa, na sala de jantar.
Eles ouviram o Pelé falar no rádio.

3

Mamãe
Tic-Tac
Paralelo
Obrigado
Estrada de ferro
Jogador de futebol

Fig. 2. Phrases in Portuguese used for evaluation of aphasia.

Fig. 3. Words in Portuguese for evaluation of dysarthria.

Table 1. Interobserver agreement for NIHSS items

Scale item	kappa	95% CI
LOC	0.97	0.75–1.00
LOC questions	0.91	0.72–1.00
LOC commands	0.94	0.73–1.00
Conjugate gaze	0.92	0.71–1.12
Visual fields	0.58	0.39–0.77
Facial palsy	0.26	0.05–0.45
Motor strength upper limbs	0.70	0.50–0.89
Motor strength lower limbs	0.74	0.55–0.92
Ataxia	0.70	0.49–0.89
Sensory	0.51	0.30–0.71
Language/aphasia	0.68	0.48–0.86
Dysarthria	0.64	0.40–0.84
Extinction or inattention	0.49	0.43–0.84

LOC = Level of consciousness.

57 underwent BI scoring and 62 PV-NIHSS scoring (some patients were scored on more than one scale). Most patients were evaluated at outpatient clinics, while only 9 patients in the admitting wards, within 7 days of stroke onset. Mean age was 61 ± 12 years and 50.8% were male.

For PV-NIHSS evaluation, the intraclass correlation coefficient was 0.902 (95% CI = 0.84–0.94). Interobserver agreement for each scale item is shown in table 1. Interobserver agreement was considered excellent for level of consciousness and conjugate gaze; good for motor strength, ataxia, language and dysarthria; moderate for visual fields, sensory testing and extinction/neglect evaluation, and fair for facial palsy.

In the BI analysis, the intraclass correlation coefficient was 0.967 (95% CI = 0.94–0.98). When considering the

five categories of the scale, weighted kappa was 0.70 (95% CI = 0.61–0.96).

For the mRS, we compared the results with and without a structured interview. Without the structured interview, the interobserver agreement was only fair (kappa = 0.34; 95% CI = 0.09–0.59). However, when using the structured mRS, we observed an improvement in interobserver agreement (kappa = 0.75; 95% CI = 0.57–0.92).

The correlation between the PV-NIHSS and the other two scales was good [$r_s = -0.438$ between the PV-NIHSS and the BI ($p = 0.003$); $r_s = 0.607$ between the PV-NIHSS and the mRS ($p < 0.001$)]. Agreement rates for all three scales did not significantly differ among different health care professionals.

Discussion

The present study demonstrated good interobserver agreement rates for the PV-NIHSS, BI and mRS with a structured interview, which validates all three scales in a Portuguese-speaking population. Our main finding was an inappropriately low concordance rate for the mRS without the structured interview. This finding is important because many multicenter studies use the mRS as a primary endpoint without any particular structured interview. Although the mRS is fast and easy to apply, each scale item description is broad and subject to wide interpretations. Based on this observation, a structured interview was proposed which decreased interobserver agreement rates, but had not been externally validated [7]. Our study results provide external validation for the mRS with a structured interview and suggest that this strategy should always be performed with the mRS. An alterna-

tive strategy has been applied in clinical trials, using a certification process with video-based training, but formal interobserver agreement rates have not been calculated with this approach [8].

The NIHSS was translated and adapted to the Brazilian culture and was easily applied by different health care professionals. Cultural adaptation of scale items was feasible and well tolerated by each patient. Most scale items performed favorably, showing good concordance rates. However, facial palsy rating was extremely variable among health care professionals, a finding which has previously been reported [9]. Other scale items have also

shown low concordance rates in other studies, such as level of consciousness, aphasia, ataxia and dysarthria [9, 10]. These items exhibited good to excellent concordance rates in our study, maybe due to our sample population, which included mostly outpatients. Therefore, the PV-NIHSS should be tested in other Portuguese-speaking populations and more specifically in the acute setting in order to further validate our results.

In conclusion, the NIHSS, BI and mRS are well suited to study patients from ethnically diverse cultures. The mRS showed low concordance rates which improved after applying a previously validated structured interview.

References

- 1 Mansur AP, de Souza MFM, Favarato D, Avakian SD, Machado CLA, Aldrigui JM, Franchini RJA: Stroke and ischemic heart disease mortality trends in Brazil from 1979 to 1996. *Neuroepidemiology* 2003;22:179–183.
- 2 Cabral NL, Moro C, Silva GR, Scola RH, Werneck LC: Study comparing the stroke unit outcome and conventional ward treatment: a randomized study in Joinville, Brazil. *Arq Neuropsiquiatr* 2003;61:188–193.
- 3 Sociedade Brasileira de Doencas Cerebrovasculares (SBDCV): Brazilian consensus for thrombolysis in acute ischemic stroke. *Arq Neuropsiquiatr* 2002;60:675–680.
- 4 de Caneda MA, Fernandes JG, de Almeida AG, Mugnol FE: Confiabilidade de escalas de comprometimento neurológico em pacientes com acidente vascular cerebral. *Arq Neuropsiquiatr* 2006;64:690–697.
- 5 Brott T, Adams HP Jr, Olinger CP, et al: Measurements of acute cerebral infarction: a clinical examination scale. *Stroke* 1989;20:864–870.
- 6 Shah S, Vanclay F, Cooper B: Improving the sensitivity of the Barthel Index for stroke rehabilitation. *J Clin Epidemiol* 1989;42:703–709.
- 7 Wilson JT, Hareendran A, Hendry A, Potter J, Bone I, Muir KW: Reliability of the modified Rankin Scale across multiple raters: benefits of a structured interview. *Stroke* 2005;36:777–781.
- 8 Quinn TJ, Lees KR, Hardemark HG, Dawson J, Walters MR: Initial experience of a digital training resource for modified Rankin scale assessment in clinical trials. *Stroke* 2007;38:2257–2261.
- 9 Josephson SA, Hills NK, Johnston SC: NIH Stroke Scale reliability in ratings from a large sample of clinicians. *Cerebrovasc Dis* 2006;22:389–395.
- 10 Lyden PD, Lu M, Levine SR, Brott TG, Broderick J, NINDS rtPA Stroke Study Group: A modified National Institutes of Health Stroke Scale for use in stroke clinical trials: preliminary reliability and validity. *Stroke* 2001;32:1310–1317.