Published in final edited form as:
Int J Stroke. 2009 August ; 4(4): 267–273. doi:10.1111/j.1747-4949.2009.00294.x.

The Modified National Institutes of Health Stroke Scale (mNIHSS): Its Time Has Come

Brett C. Meyer, M.D. and
Department of Neurosciences, UCSD School of Medicine, Stroke Center (8466) 3rd Floor, OPC, Suite #3, 200 West Arbor Drive. San Diego, CA. 92103-8466

Patrick D. Lyden, M.D., FAAN
Department of Neurosciences, UCSD School of Medicine, and Research Division, Department of Veteran's Affairs, Stroke Center (8466) 3rd Floor, OPC, Suite #3, 200 West Arbor Drive. San Diego, CA. 92103-8466

Abstract

The National Institutes of Health Stroke Scale (NIHSS) is a well known, reliable and valid stroke deficit scale. The NIHSS is simple, quick, and has shown significant reliability in diverse groups, settings, and languages. The NIHSS also contains items with poor reliability and redundancy. Recent investigations (include assessing a new training DVD, analyzing web-based or videotape certifications, and testing foreign language versions) have further detailed reliability issues. Items recurrently shown to have poor reliability include Level of Consciousness, Facial Palsy, Limb Ataxia, and Dysarthria.

The modified NIHSS (mNIHSS) minimizes redundancy and eliminates poorly reliable items. The mNIHSS shows greater reliability in multiple settings and cohorts, including scores abstracted from records, when used via telemedicine, and when used in clinical trials. In a validation of the mNIHSS against the NIHSS, the number of elements with excellent agreement increased from 54% to 71%, while poor agreement decreased from 12% to 5%. Overall, 45% of NIHSS items had less than excellent reliability vs. only 29% for the mNIHSS.

The mNIHSS is not the ideal stroke scale, but it is a significant improvement over the NIHSS. The mNIHSS has shown reliability at bedside, with record abstraction, with telemedicine, and in clinical trials. Since the mNIHSS is more reliable, it may allow for improved practitioner communication, improved medical care, and refinement of trial enrollments. The mNIHSS should now serve as the primary stroke clinical deficit scale for clinical and research aims. When it comes to the mNIHSS, its time has come!

Keywords

Stroke Scale; NIHSS; modified; mNIHSS; Reliability

Correspondence to: **Brett C. Meyer, M.D.**, Department of Neurosciences, UCSD School of Medicine, Stroke Center (8466), 3rd Floor, OPC, Suite #3, 200 West Arbor Drive, San Diego, CA. 92103-8466, TELE 619-543-7760, FAX 619-543-7771, Email: bcmeyer@ucsd.edu.

There are no other conflicts of interest to disclose.

Introduction

The NIHSS

The NIHSS is a graded neurological examination assessing consciousness, eye movements, visual fields, motor and sensory impairments, ataxia, speech, cognition and inattention. The scale was developed as a communication tool, and has since been used in stroke trials.(1,2) Though other scales have been evaluated, the NIHSS is arguably the most frequently used deficit scale for stroke patient evaluations.(3–5) The NIHSS' reliability, coupled with its ability to predict patient outcome, has helped to foster its use in both clinical and academic arenas. (6) (7) Despite the ubiquity of the NIHSS, the scale contains items that repeatedly prove unreliable; ignoring this may lead to unanticipated consequences.

NIHSS: The Upside

In developing an ideal stroke scale, issues of simplicity, reliability, validity, and generalizability-of-use must be pursued, especially if a scale is to be used by a broad array of practitioners.(8) The merits of the NIHSS include simplicity, minimal time requirements, and numerous assessments of inter- rater reliability in diverse groups including neurologists, non-neurologists, clinical trials' coordinators, community practitioners, and even practitioners in training.(1,2,9–13) Reliability is known to improve with personal and video tape training,(9, 14) and the scale can be abstracted from medical records.(3) Similarly, the NIHSS has shown validity,(15) and ability to predict discharge and 3 month outcomes.(7,16)

NIHSS: The Downside

In spite of its successes, there are problems with the NIHSS. The scale contains items with poor reliability, and has been criticized for its redundancy and complexity.(17) (12) Though the internal structure consists of 2 factors relating to the cerebral hemispheres,(15,18) its emphasis on language and following commands may result in a higher value for dominant hemisphere strokes. For a given NIHSS score, the median volume of right hemisphere strokes is larger than that of left hemisphere strokes.(19)

Another downside is the lessened weighting for posterior circulation strokes, a problem for nearly all neurological deficit scales. Though some items related to the verterbo-basilar system can be scored (e.g. LOC, visual fields, facial palsy, sensory, motor, dysarthria and ataxia), other elements receive no score (e.g. diplopia, dysphagia, gait instability, hearing, and nystagmus). Scales have since been developed to assess vertebro-basilar stroke symptoms.(20)

Reliability Assessments of the NIHSS

The NIHSS has been investigated thoroughly for its reliability since 1989.(1,2) The scale's overall reliability is clear, yet the same items are noted over and over again to show poorer reliability. Table 2 & Table 3 list trials that have assessed both the NIHSS and modified NIHSS (mNIHSS) scales. NIHSS assessments have consistently shown specific items that yield low inter- rater reliability.(Table 2)(14,17) (21) (3,13,22) These items include scoring Level of Consciousness, Gaze, Facial Palsy, Ataxia, and Dysarthria. For these trials, the total number of elements yielding excellent agreement (based on Kappa statistics or Intra-Class Correlations) is 49/90 (54%), while moderate agreement is 30/90 (33%) and poor agreement is 11/90 (12%). This amounts to 41/90 (45%) of items having less than excellent reliability, for a stroke scale that is used nearly universally for clinical deficit evaluations, prognostication, and clinical trial enrollments.

Further reliability assessments have been performed on a great number of participants. The original NIHSS tapes are now > 12 years old, and were imbalanced to patient findings. Recently, the NINDS produced a new DVD for NIHSS training and certification.

Representative patients, exhibiting the full range of scores, and balanced for lesion side, findings, and total score were recruited. Inter- rater reliability was assessed in 112 raters including stroke nurses (26%), ED/ other physicians (34%) and neurologists (40%). The scale maintained a high intra-class correlation coefficient of 0.94. Reliability was higher for previously NIHSS certified examiners. Of the 15 NIHSS items, 2 items showed excellent agreement, 11 showed moderate agreement, and 2 showed poor agreement.(23) As with the patients on the tapes, items with poor NIHSS reliability included scoring Facial Palsy (kappa=0.38) and Ataxia (kappa=0.21). Level of Consciousness (kappa=0.46) and Dysarthria (kappa=0.56) showed only moderate agreement. These data suggest that the poor reliability of these items is not an artifact of the videotape technique itself.

Since posting this certification method on the Internet, thousands of practitioners have since been certified through the American Heart Association (www.professionaleducationcenter.americanheart.org). Among over 15,000 individuals who have taken this online certification, 2 NIHSS items showed excellent agreement, 11 showed moderate agreement, and 2 showed poor agreement. The items with poorer NIHSS reliability included Facial Palsy (kappa=0.25), Ataxia (kappa=0.15), Level of Consciousness (kappa=0.43), and Dysarthria (kappa=0.46), and Gaze (kappa=0.44).

The National Stroke Association also invited assessment of 7,405 unique raters who completed videotape certification between 1998 and 2004.(24) In this diverse sample of practitioners, Aphasia (modified kappa=0.60) and Facial Palsy (modified kappa=0.65) showed poorer reliability and contributed most to NIHSS variance. Eight NIHSS items showed excellent agreement, 4 showed moderate agreement, and 3 showed poor agreement. Again, as expected, the worst scoring element was Ataxia (kappa=0.25).

Given the global nature of stroke, it is not surprising that the NIHSS has been assessed in multiple languages including Spanish, Italian, and Chinese. In 98 patients, an assessment of inter-rater reliability, in extensively trained investigators using a Spanish language NIHSS, showed kappas ranging from 0.77 to 0.99 with the lowest kappa being for Facial Palsy.(25) Other versions of the Spanish NIHSS have also been developed.(26)

In a study of 48 patients, the Chinese version was reliable, with 13/15 items reaching an “acceptable” kappa sore >0.40. The lowest scores were consistent for 3 groups assessed: Facial Palsy (kappa=0.57–0.78 for physicians, kappa=0.72–0.78 for nurses, and kappa=0.28–0.69 for combined physicians & nurses), and Ataxia (kappa=0.47–0.88 for physicians, kappa=0.52–0.70 for nurses, and kappa=0.23–0.47 for combined physicians & nurses).(27)

Finally, an Italian version of the NIHSS was also developed (personal communication, Pezzella et al). In a study of 1,556 physicians and nurses, the Italian version was reliable overall. Again, the lowest NIHSS kappa scores recorded were for gaze (kappa=0.29), LOC Questions (kappa0.05), and ataxia (kappa=0.38).

In summary, the total NIHSS score and many of the items are reliable in multiple publications, in various settings, with variable levels of practitioner training, and in multiple languages. Unfortunately, the same individual items are noted over and over again to show poorer reliability. These items generally include Loss of Consciousness, Facial Palsy, Ataxia and Dysarthria. These elements may contribute to difficulties in practitioner communication, incorrect hospital care patterns that are based on the NIHSS (e.g. decisions to give thrombolytics or specific inpatient care protocols), variable trial enrollments, and even possible difficulties with assessing patient outcomes.

The mNIHSS

The modified National Institutes of Health Stroke Scale (mNIHSS) is also a graded neurological examination, developed using formal clinimetric analyses.(17) (Table 1) First, investigators conducted investigations of NIHSS' reliability, validity and internal structure. (14) (15) A simplified version was proposed that maintained similar internal structure.(15) Level of Consciousness was redundant, and was therefore dropped from the new scale. Ataxia showed poor reliability, so was excluded. Facial Palsy and Dysarthria showed poor reliability, and were redundant, so they were eliminated. The Sensory item was simplified due to poor reliability. With fewer items and simpler grading, the mNIHSS was intended to be simpler and easier to administer. The sample picture, list of words, and sample sentences from the original NIHSS were not changed, and could continue to be used as necessary to help assess relevant scale elements.

Reliability was demonstrated with certification data from the NINDS rt-PA stroke trials used previously for testing of the NIHSS.(17) When combining certification tapes, the mNIHSS showed fewer items with poor kappa scores (decreasing from 20% to 14%). The mNIHSS also demonstrated validity using factor and coefficient analysis. The resulting mNIHSS was simple, reliable and valid. A subsequent study prospectively assessed the mNIHSS in 45 patients with stable deficits.(21) Whereas total NIHSS scores did not differ between examiners by > 4 points, total mNIHSS scores did not differ by > 2 points. The NIHSS showed good reliability with 10/15 (67%) excellent items, 4/15 (27%) good items, and 1/15 (7%) poorly reliable items. The mNIHSS scored even better with 10/11 (91%) excellent items and only 1/11 (9%) good items. Zero (0%) items showed poor mNIHSS reliability. The mNIHSS was also reliably abstracted from medical records with no poorly reliable mNIHSS items found.(28)

Assessments of the mNIHSS have consistently shown lesser amounts of items with low inter-rater reliability than when using the NIHSS.(Table 3)(13,14,17,21,22,28) Summarizing the trials listed, the total number of items yielding excellent agreement is 47/66 (71%), compared to the 49/90 (54%) for the NIHSS. The number of mNIHSS items with moderate agreement is 16/66 (24%) compared to 30/90 (33%) for the NIHSS. The number of mNIHSS items showing only poor agreement is 3/66 (5%), compared to 11/90 (12%) for the NIHSS. Overall, this amounts to only 19/66 (29%) of items having less than excellent reliability, compared to the much higher 41/90 (45%) of NIHSS items.

What We Gain by Using the mNIHSS

The improved reliability benefits alone may be enough to justify the transition to the mNIHSS. However, there are other significant benefits. Given the unreliability of some of the NIHSS items, patients may score high on the NIHSS when they actually have mild strokes but questionable other findings. There should be a significant difference between a score of 4 and a score of 8. Patient #1 may poorly follow commands, and have mild right sensory/ motor deficits, but may also have a forced right gaze preference and neglect to 2 modalities (scoring a worrisome 8 on the NIHSS). Patient #2 may score the same for command, sensory and motor, but may have questionable incoordination of arm and leg, questionable slurring, and questionable facial abnormality (also scoring an 8). Though Patient #2 is certainly of lesser severity, this patient may score 4 points higher simply due to the unreliability of "soft findings".

Alternatively, patients may score as mild even if they have more severe deficits, since unreliability may result in certain items being unscored. Failing to treat mild stroke patients with rt-PA may result in bad outcomes. Roughly 1/3 of patients excluded from treatment because of being too mild, were dead or disabled at discharge.(29) This finding was replicated in patients deemed "too good to treat".(30) Perhaps one of the reasons that mild patients should be treated is that they could truly be less mild than they score on the NIHSS. If they score a 3,

instead of their true 7, one may be more likely to withhold rt-PA therapy. This intrinsic variability may result in different treatment behaviors, and different patient outcomes.

In the NIHSS, 7 of 42 points are related to language function, while only 2 of 42 points are attributed to neglect functions. Redundant items noted in the NIHSS have been deleted from the mNIHSS, resulting in a more balanced clinical scale. Therefore, lateralization bias may be minimized.

Since the mNIHSS variance is lower than that of the NIHSS, power is greater and results in the potential for smaller clinical trial sample sizes.(17) This benefit is important, especially given the need for large sample sizes in research trials. Further, one of the difficulties with trial enrollments is the necessity of enrolling patients with sufficient deficit that could translate into a measurable improvement at outcome. In the past, this has necessitated choosing a minimum NIHSS inclusion criteria of 5–7, to perhaps account for some “soft findings” being scored. Perhaps to circumvent this inclusion criteria issue, some investigators have chosen combinations of more reliable clinical deficit findings (even with low NIHSS scores) for enrollment.(NCT00252239) Using these combinations makes generalization of trial results to the overall population much more difficult. Using the mNIHSS in lieu of the NIHSS may help to minimize various inclusion combinations, and thus improve the generalizability of a trial result.

Some trials have chosen to set a minimum NIHSS for trial enrollment. In a prospective assessment of the scales, the median NIHSS was 5 while the median mNIHSS was 3.(21) Similarly, in a prospective telemedicine trial, the median NIHSS was 7 while the median mNIHSS was 5 (unpublished data (NCT00283868)). The median 2 point difference should support setting this enrollment minimum at 3–5 for the mNIHSS instead of the NIHSS 5–7.

Data abstraction is a reliable method for determining the NIHSS from medical records.(3,31, 32) This means of data collection has been used to evaluate deficit when an NIHSS was not specifically performed at admission. Since the mNIHSS can be more reliably abstracted from records, a more accurate and easier to obtain record of initial patient presentation is now available.(33)

Another advantage of transitioning to the use of the mNIHSS is its improved consistency in assessing daily stroke patient changes/ improvements. Severely affected acute stroke patients may not be able to receive NIHSS scores for Ataxia or Dysarthria because their arousal state may preclude testing these items. Since these items are not scored abnormal unless patients produce testable behaviors, these patients may be “too sick to score” on these items. Though the patients may clinically improve, their NIHSS scores may artificially worsen since now items such as Ataxia and Dysarthria can receive the scores that were previously unscored. Since these items have been removed from the mNIHSS, this difficulty can be avoided, or at least lessened.

Multiple investigations have assessed the reliability of performing the NIHSS via telemedicine. One experience noted a good interrater correlation coefficient (0.97, $p<0.001$).(34) Not surprisingly though, Ataxia showed poor inter-rater reliability (kappa= −0.07). Dysarthria (kappa= 0.55), Sensory (kappa= 0.48) and Facial Palsy (kappa= 0.40) were each only moderately reliable. Two further prospective assessments of NIHSS telemedicine reliability include the STRokE DOC Telemedicine trial (NCT00283868)(22,35) and an assessment of untrained telemedicine practitioners (NCT00390286).(13) (Table 2) With trained investigators, telemedicine NIHSS reliability was poor for Ataxia (kappa= 0.34) and Facial Palsy (kappa= 0.22), and only moderate for Dysarthria (kappa= 0.61).(22) With untrained investigators in a separate cohort, Neglect (kappa= 0.72), Ataxia (kappa= 0.65), Facial Palsy (kappa= 0.62), Dysarthria (kappa= 0.06) and Gaze (kappa= 0.60) scored as only moderately

reliable. These 2 telemedicine reliability assessments also tested the mNIHSS.(13,22) (Table 3) In these assessments, the mNIHSS showed improved reliability over the NIHSS, with a combined 86% of items showing excellent reliability, and 0% items showing poor reliability. If the mNIHSS were used in telemedicine evaluations, practitioners could be more certain that assessed telemedicine deficits represent actual patient findings.

A trial assessing obtaining NIHSS and mNIHSS elements via telephone as compared to telemedicine reported 9% greater data collection in telemedicine than telephone. In the telephone arm, NIHSS questions showing > 15% missing data were consistent with items with known poorer NIHSS reliability, including Gaze, Visual Fields, Limb Ataxia, Sensory, and Neglect. (unpublished data (NCT00283868)) Poorly reliable NIHSS items may be less likely to be performed via telephone, as they may be complicated, confusing, unreliable, or the practitioner may not feel comfortable with their scoring structure. This again suggests the increased value of using the mNIHSS.

Disadvantages & Limitations

Neither the NIHSS nor the mNIHSS are the ideal stroke scale. Both fail to accurately or reliably detect patients with posterior circulation findings. With the removal of the ataxia item, there may be concern that the mNIHSS would be even less able to assess brainstem strokes. However, since ataxia is a poorly reliable NIHSS item anyway, the benefit of using a scale that inconsistently/ variably assesses the posterior circulation, may not outweigh the consistency of the mNIHSS. Furthermore, many posterior circulation events are captured by other mNIHSS items (e.g. LOC, visual fields, sensory, and motor). Finally and most importantly, many clinical trials routinely include only anterior circulation strokes, so there is less need to measure posterior circulation deficits for these purposes.

One question that still remains is that of predictive validity. In previous longitudinal analyses, the NIHSS predicted outcomes at discharge and 3 months.(7,16) The mNIHSS has been assessed for validity as well. Retrospectively, the mNIHSS has shown to be predictive of patient outcomes using the NINDS data set.(17) Prospectively, the mNIHSS has shown both construct and concurrent validity.(21) The question of prospective mNIHSS predictive ability is currently being assessed in the STRokE DOC telemedicine trial, and using data from combined databases. The best way to fully assess the mNIHSS for predictive ability is for investigators to begin using this scale in research trials.

Future Challenges

Besides using the mNIHSS in future trials to assess outcomes, the mNIHSS can be used in reassessments of prior trials as well. NIHSS variability may have consequences for single or multicenter trial outcomes. Though some clinical trials, limited to those with blinded treatment arms, have attempted to mandate the same investigator do deficit assessments at both enrollment and 90 days, this is usually impractical and often impossible. Even if a single practitioner is used, the inherent unreliability of some NIHSS items may result in up to a 6 point difference in deficit evaluations. One question that can be assessed is whether the patient's clinical outcomes, and therefore the success of the trial itself, can be reanalyzed due to the improved reliability of the mNIHSS. Further, prior trials can be reanalyzed using an extrapolated mNIHSS (from the originally reported NIHSS). Future trials can be designed using this mNIHSS, and can capitalize of its improved reliability.

Conclusion

The mNIHSS: It's Time Has Come

The mNIHSS is not the ideal stroke scoring scale. It is, however, a significant improvement over the NIHSS. The NIHSS has items with poor reliability, which may translate into problems ranging from miscommunication between practitioners, to variable reporting of clinical trial outcomes. The mNIHSS is more reliable, allowing for improved practitioner communication. The mNIHSS preserves the validity shown for the NIHSS, with improved hemisphere balance. It has shown reliability when abstracted from medical records, when used over a telemedicine link, and when used in research trials. The mNIHSS can now be used in the care of the stroke patient, both for acute management and future clinical trials. Though the NIHSS is a good clinical deficit scale, when it comes to the mNIHSS, its time has come.

Acknowledgements and Funding

This paper was supported in part by the NINDS P50 NS044148 and by the Department of Veterans' Affairs, Research Division.

References

1. 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(7):864–870. [PubMed: 2749846]
2. Goldstein LB, Bartels C, Davis JN. Interrater reliability of the NIH stroke scale. Archives of Neurology 1989;46:660–662. [PubMed: 2730378]
3. 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;30:1534–1537. [PubMed: 10436096]
4. Cote R, Battista RN, Wolfson C, Boucher J, Adam J, Hachinski V. The Canadian Neurological Scale: Validation and reliability assessment. Neurology 1989;39:638. [PubMed: 2710353]
5. 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. [PubMed: 8342193]
6. Wityk R, Pessin MS, Kaplan RF, Caplan LR. Serial assessment of acute stroke using the NIH stroke scale. Stroke 1994;25:362. [PubMed: 8303746]
7. Adams HP Jr, Davis PH, Leira EC, Chang KC, Bendixen BH, Clarke WR, 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;53(1):126–131. [PubMed: 10408548]
8. Asplund K. Clinimetrics in stroke research. Stroke 1987;18:528. [PubMed: 3564114]
9. Schmulling S, Grond M, Rudolf J, Kiencke P. Training as a Prerequisite for Reliable Use of NIH Stroke Scale. Stroke 1998;29:1258–1259. [PubMed: 9626306]
10. Albanese MA, Clarke WR, Adams HP Jr, Woolson RF. Ensuring reliability of outcome measures on 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. [PubMed: 8073453]
11. Goldstein L, Samsa G. Reliability of the National Institutes of Health Stroke Scale. Stroke 1997;28 (2):307–310. [PubMed: 9040680]
12. 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. Cerebrovascular Diseases 1999;9(6):323–327. [PubMed: 10545689]
13. Meyer BC, Raman R, Chacon MR, Jensen M, Werner JD. Reliability of site-independent telemedicine when assessed by telemedicine-naive stroke practitioners. J Stroke Cerebrovasc Dis 2008;17(4):181–186. [PubMed: 18589337]
14. Lyden P, Brott T, Tilley B, Welch KMA, Mascha EJ, Levine S, et al. Improved reliability of the NIH stroke scale using video training. Stroke 1994;25:2220–2226. [PubMed: 7974549]

15. Lyden P, Lu M, Jackson C, Marler J, Kothari R, Brott T, et al. Underlying Structure of the National Institutes of Health Stroke Scale: Results of a Factor Analysis. Stroke 1999;30:2347. [PubMed: 10548669]

16. Schlegel D, Kolb SJ, Luciano JM, Tovar JM, Cucchiara BL, Liebeskind DS, et al. Utility of the NIH Stroke Scale as a predictor of hospital disposition. Stroke 2003;34(1):134–137. [PubMed: 12511764]

17. Lyden PD, Lu M, Levine S, Brott TG, Broderick J. A Modified National Institutes of Health Stroke Scale for Use in Stroke Clinical Trials. Preliminary Reliability and Validity. Stroke 2001;32:1310–1317. [PubMed: 11387492]

18. Millis SR, Straube D, Iramaneerat C, Smith EV Jr, Lyden P. Measurement properties of the National Institutes of Health Stroke Scale for people with right- and left-hemisphere lesions: further analysis of the clomethiazole for acute stroke study-ischemic (class-I) trial. Arch Phys Med Rehabil 2007;88(3):302–308. [PubMed: 17321821]

19. Woo D, Broderick J, Kothari R, Lu M, Brott T, Marler J, et al. Does the National Institutes of Health Stroke Scale Favor Left Hemisphere Strokes. Stroke 1999;30:2355. [PubMed: 10548670]

20. Gur AY, Lampl Y, Gross B, Royter V, Shopin L, Bornstein NM. A new scale for assessing patients with vertebrobasilar stroke-the Israeli Vertebrobasilar Stroke Scale (IVBSS): inter-rater reliability and concurrent validity. Clin Neurol Neurosurg 2007;109(4):317–322. [PubMed: 17254701]

21. Meyer BC, Hemmen TM, Jackson C, Lyden PD. Modified National Institutes of Health Stroke Scale for Use in Stroke Clinical Trials. Stroke 2002;33:1261–1266. [PubMed: 11988601]

22. Meyer BC, Lyden PD, Al-Khoury L, Cheng Y, Raman R, Fellman R, et al. Prospective reliability of the STRokE DOC wireless/site independent telemedicine system. Neurology 2005;64(6):1058–1060. [PubMed: 15781827]

23. Lyden P, Raman R, Liu L, Grotta J, Broderick J, Olson S, et al. NIHSS training and certification using a new digital video disk is reliable. Stroke 2005;36(11):2446–2449. [PubMed: 16224093]

24. Josephson SA, Hills NK, Johnston SC. NIH Stroke Scale reliability in ratings from a large sample of clinicians. Cerebrovasc Dis 2006;22(5–6):389–395. [PubMed: 16888381]

25. Dominguez R, Vila JF, Augustovski F, Irazola V, Castillo PR, Rotta Escalante R, et al. Spanish cross-cultural adaptation and validation of the National Institutes of Health Stroke Scale. Mayo Clin Proc 2006;81(4):476–480. [PubMed: 16610567]

26. Montaner J, Alvarez-Sabin J. [NIH stroke scale and its adaptation to Spanish]. Neurologia 2006;21(4):192–202. [PubMed: 16832774]

27. 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(4):457–463. [PubMed: 16146632]

28. 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(2):568–570. [PubMed: 12574577]

29. Barber PA, Zhang J, Demchuk A, Hill MD, Buchan AM. Why are stroke patients excluded from TPA therapy? Neurology 2001;56:1015–1020. [PubMed: 11320171]

30. Smith EE, Abdullah AR, Petkovska I, Rosenthal E, Koroshetz WJ, Schwamm LH. Poor outcomes in patients who do not receive intravenous tissue plasminogen activator because of mild or improving ischemic stroke. Stroke 2005;36(11):2497–2499. [PubMed: 16210552]

31. Williams LS, Yilmaz E, Lopez-Yunez AM. Retrospective Assessment of Initial Stroke Severity With the NIH Stroke Scale. Stroke 2000;31:858–862. [PubMed: 10753988]

32. Bushnell CD, Johnston DC, Goldstein L. Retrospective Assessment of Initial Stroke Severity. Comparison of the NIH Stroek Scale and the Canadian Neurological Scale. Stroke 2001;32(3):656–660. [PubMed: 11239183]

33. Chalela JA, Kasner SE, Jauch EC, Pancioli AM. Safety of air medical transportation after tissue plasminogen activator administration in acute ischemic stroke. Stroke 1999;30(11):2366–2368. [PubMed: 10548672]

34. Shafqat S, Kvedar JC, Guanci MM, Chang Y, Schwamm LH. Role for telemedicine in acute stroke. Feasibility and reliability of remote administration of the NIH stroke scale. Stroke 1999;30(10):2141–2145. [PubMed: 10512919]

35. Meyer BC, Raman R, Rao R, Fellman RD, Beer J, Werner J, et al. The “Stroke Team Remote Evaluation Using a Digital Observation Camera (STRokE DOC)” Telemedicine Clinical Trial Technique: “Video Clip, Drip and/ or Ship”. Int. J. Stroke 2007;2:4:281–287. [PubMed: 18705930]

Table 1

The mNIHSS and Scoring Guide

The table includes each of 11 items on the mNIHSS deficit scale including item number,item name, score guide and instructions and total score (out of 31). The item numbers correspond to the original scale.

Item	Item Name	Scoring Guide	Score
1b	LOC Questions	0 = Answers both correctly. 1 = Answers one correctly. 2 = Answers neither correctly.	
1c	LOC Commands	0 = Performs both tasks correctly. 1 = Performs one task correctly. 2 = Performs neither task.	
2	Gaze	0 = Normal. 1 = Partial gaze palsy. 2 = Total gaze palsy.	
3	Visual Fields	0 = No visual loss. 1 = Partial hemianopia. 2 = Complete hemianopia. 3 = Bilateral hemianopia.	
5a	Left Arm Motor	0 = No drift 1 = Drift before 10 seconds 2 = Falls before 10 seconds 3 = No effort against gravity 4 = No movement UN = Amputation or joint fusion, explain:	
5b	Right Arm Motor	0 = No drift 1 = Drift before 10 seconds 2 = Falls before 10 seconds 3 = No effort against gravity 4 = No movement UN = Amputation or joint fusion, explain:	
6a	Left Leg Motor	0 = No drift 1 = Drift before 5 seconds 2 = Falls before 5 seconds 3 = No effort against gravity 4 = No movement UN = Amputation or joint fusion, explain:	
6b	Right Leg Motor	0 = No drift	

Item	Item Name	Scoring Guide	Score
		1 = Drift before 5 seconds 2 = Falls before 5 seconds 3 = No effort against gravity 4 = No movement UN = Amputation or joint fusion, explain:	
8	Sensory	0 = Normal 1 = Abnormal	
9	Language	0 = Normal 1 = Mild aphasia 2 = Severe aphasia 3 = Mute or global aphasia	
11	Neglect	0 = Normal 1 = Mild 2 = Severe	
			Total Score (out of 31):

* Scoring from original scale

Table 2

Inter- Rater Reliability Comparisons in Trials (NIHSS)

This table includes a comparison of the 6 prior evaluations which have assessed both the NIHSS and mNIHSS, including the reliability comparisons between these trials for the NIHSS. Listed are the trials, the kappa sores for individual items, the study specific threshold for excellent agreement, and percentage of items with excellent, moderate, or poor agreement (color coded in green, yellow, and red) for each trial. An overall NIHSS summary for all trials is also included. The denominator of 90 comes from the total number of items across 6 trials (15 items in each trial × 6 trials = 90 items).

		Video Tape Assessments [14,18]	Video Tape Assessments [14,18]	mNIHSS-Prospective [22]	STRokE DOC-Aim 1 [23]	TACTIC-Untrained [13]	Medical Record Abstraction [3]
#	Item Name	NIHSS Tape1	NIHSS Tape2	NIHSS	NIHSS	NIHSS	NIHSS
1a	LOC	0.62	0.42	0.46	100% Agree	0.87	0.74
1b	LOC Questions	0.68	0.90	0.94	0.93	0.96	0.71
1c	LOC Commands	0.00	0.93	0.94	100% Agree	100% Agree	0.78
2	Gaze	0.02	0.51	0.66	100% Agree	0.60	0.74
3	Visual Fields	0.94	0.81	0.88	0.93	0.78	0.76
4	Facial Palsy	0.38	0.20	0.74	0.22	0.62	0.27
5a	Left Arm Motor	0.79	0.92	0.97	0.88	0.94	0.91
5b	Right Arm Motor	0.79	0.94	0.96	0.82	0.97	0.90
6a	Left Leg Motor	0.80	0.95	0.95	0.74	0.95	0.90
6b	Right Leg Motor	0.71	0.66	0.98	0.80	0.89	0.89
7	Limb Ataxia	0.23	0.56	0.69	0.34	0.65	0.70
8	Sensory	0.94	0.81	0.89	0.80	100% Agree	0.63
9	Language	0.39	0.57	0.84	0.73	0.89	0.92
10	Dysarthria	0.72	0.42	0.29	0.61	0.60	0.36
11	Neglect	0.54	0.53	0.89	0.80	0.72	0.59
Study Specific Kappa Scoring		>0.75: Excellent	>0.75: Excellent	≥0.75: Excellent	≥0.75: Excellent	≥0.75: Excellent	>0.75: Excellent
% Excellent		5/15 (33%)	7/15 (47%)	10/15 (67%)	10/15 (67%)	10/15 (67%)	7/15 (47%)
% Moderate		5/15 (33%)	7/15 (47%)	4/15 (27%)	3/15 (20%)	5/15 (33%)	6/15 (40%)
% Poor		5/15 (33%)	1/15 (7%)	1/15 (7%)	2/15 (13%)	N/A	2/15 (13%)

Overall NIHSS: Excellent (49/90=54%) Moderate (30/90=33%) Poor (11/90=12%)

Int J Stroke. Author manuscript; available in PMC 2010 August 1.

Table 3

Inter- Rater Reliability Comparisons in Trials (mNIHSS)

This table includes a comparison of the 6 prior evaluations which have assessed both the NIHSS and mNIHSS, showing the reliability comparisons between these trials for the mNIHSS.Listed are the trials, the kappa sores for individual items, the study specific threshold for excellent agreement, and percentage of items with excellent, moderate, or poor agreement (color coded in green, yellow, and red) for each trial. An overall mNIHSS summary for all trials is also included. The denominator of 66 comes from the total number of items across 6 trials (11 items in each trial × 6 trials = 66 items).

		Video Tape Assessments [14,18]	Video Tape Assessments [14,18]	mNIHSS-Prospective [22]	STRokE DOC-Aim 1 [23]	TACTIC-Untrained [13]	Medical Record Abstraction [29]
#	Item Name	mNIHSS Tape1	mNIHSS Tape2	mNIHSS	mNIHSS	mNIHSS	mNIHSS
1a	LOC	N/A	N/A	N/A	N/A	N/A	N/A
1b	LOC Questions	0.68	0.90	0.94	0.92	0.96	0.71
1c	LOC Commands	0.00	0.93	0.94	100% Agree	100% Agree	0.78
2	Gaze	0.02	0.51	0.66	100% Agree	0.60	0.74
3	Visual Fields	0.94	0.81	0.88	0.86	0.78	0.76
4	Facial Palsy	N/A	N/A	N/A	N/A	N/A	N/A
5a	Left Arm Motor	0.79	0.92	0.97	0.84	0.95	0.91
5b	Right Arm Motor	0.79	0.94	0.96	0.82	0.97	0.90
6a	Left Leg Motor	0.80	0.95	0.95	0.74	0.95	0.90
6b	Right Leg Motor	0.71	0.66	0.98	0.80	0.89	0.89
7	Limb Ataxia	N/A	N/A	N/A	N/A	N/A	N/A
8	Sensory	0.94	0.81	0.91	0.83	100% Agree	0.58
9	Language	0.39	0.57	0.84	0.69	0.89	0.92
10	Dysarthria	N/A	N/A	N/A	N/A	N/A	N/A
11	Neglect	0.54	0.53	0.89	0.80	0.72	0.59
Study Specific Kappa Scoring		>0.75: Excellent	>0.75: Excellent	≥0.75: Excellent	≥0.75: Excellent	≥0.75: Excellent	>0.75: Excellent
% Excellent		5/11 (45%)	7/11 (64%)	10/11 (91%)	9/11 (82%)	9/11 (82%)	7/11 (64%)
% Moderate		3/11 (27%)	4/11 (36%)	1/11 (8%)	2/11 (18%)	2/11 (18%)	4/11 (36%)
% Poor		3/11 (27%)	N/A	N/A	N/A	N/A	N/A

Overall mNIHSS: Excellent (47/66=71%) Moderate (16/66=24%) Poor (3/66=5%)