Using the National Institutes of Health Stroke Scale
A Cautionary Tale

Patrick Lyden, MD

The National Institutes of Health Stroke Scale (NIHSS) is the most widely used deficit rating scale in modern neurology: over 500 000 healthcare professionals have been certified to administer it using a web-based platform. Every clinical trial in vascular neurology—prevention, acute treatment, recovery—requires a severity assessment, and the NIHSS became the gold standard for stroke severity rating after the first successful trial in acute stroke therapy, the NINDS r-tPA (National Institute of Neurological Disorders and Stroke recombinant tissue-type plasminogen activator) for Acute Stroke Trial (the Trial).[1] As part of the Trial, detailed and rigorous training/certification procedures were created for the NIHSS that facilitate wider use of the scale outside of research.[2]

Today, payers and regulators demand reportable data on patient outcomes, and such outcomes must be adjusted for baseline severity: the NIHSS has become the de facto metric for regulatory compliance. The Joint Commission, as part of its certification program for Primary Stroke Centers, now requires an NIHSS score within 12 hours of admission for all stroke patients; this assessment is to be done by a certified examiner.[3,4] Federal agencies also require outcomes adjusted for baseline stroke severity—using the NIHSS.[5] Despite widening regulatory requirements, considerable problems may arise in using the NIHSS in clinical practice because the scale was designed for research purposes.[6] Given that the scale was not designed for such widespread—and determinative—application, anyone using (or mandating use of) the NIHSS must understand its development history, clinimetric properties, and its proper bedside administration.

History/Development

During the late 1980s, several stroke-deficit rating scales were in use.[7–10] For use in a National Institutes of Health–sponsored trial of naloxone for acute stroke, investigators combined scales that had been developed at the University of Cincinnati, Canadian neurological scale, the Edinburgh-2 coma scale, and the Oxbury initial severity scale.[11] Greater scores correlated with larger infarctions.[12] This Cincinnati/Naloxone version of the NIHSS served the intended purpose in the Naloxone trial.[13] An intermediate version was used in the Pilot r-tPA for Acute Stroke Trial,[14] but when designing the NINDS r-tPA for Acute Stroke Trial, significant modifications were made to facilitate using the NIHSS in a larger clinical trial.[15] The version used today is this final iteration of the NIHSS, and it differs in important ways from the Cincinnati/Naloxone NIHSS (Table 1). A modified version contains fewer, more reliable items.[16]

The final (r-tPA) version of the NIHSS was validated against infarct volumes.[17] Several scale items require intact language function, so the NIHSS overweights deficits in patients with left versus right brain strokes.[17,18] Thus, left hemisphere strokes score 4 more points than right hemisphere strokes of similar size. The NIHSS is internally consistent, with a reasonable Cronbach's alpha and reproducible across the intended range of users: stroke nurses, vascular neurologists, and ED physicians.[19–21] The scale is reliable when used by non-neurologists who undergo training.[20,21] The total NIHSS score can predict outcome or the presence of large vessel occlusions.[22,23] A reasonable estimate of the NIHSS can be made from chart review.[24]

In 1995, after the publication of the Trial, the NIHSS became the de facto standard for rating clinical deficits in stroke trials. Several contemporary scales were similar[25–28] because there are few ways to put numbers to the neurological examination for the purpose of clinical research or bedside stroke severity measurement. To encourage greater use, such a scale must be short, but to capture all deficits, it must be long; to improve reliability, the scale must be simple, but to measure stroke deficit accurately, the scale must be complex; to capture important neurological findings, the scale must attempt to measure complicated concepts familiar to neurologists (eg, neglect), but during a large-scale clinical trial, non-neurologists must be able to use the scale also. The NIHSS was designed with these principals in mind.

Clinical Trials Versus Clinical Use

When using the NIHSS, it is critical to acknowledge that the scale was not designed to serve as a bedside rating tool for widespread use outside of research trials.[6] Rather, the scale was designed to be used by investigators (MD, RN) in the setting of a clinical trial.[15] The NIHSS design assumes that the

Received September 22, 2016; final revision received November 17, 2016; accepted December 12, 2016.
From the Department of Neurology, Cedars-Sinai Medical Center, Los Angeles, CA.
The online-only Data Supplement is available with this article at http://stroke.ahajournals.org/lookup/suppl/doi:10.1161/STROKEAHA.116.015434/-/DC1.
Correspondence to Patrick Lyden, MD, Department of Neurology, Cedars-Sinai Medical Center, 127 S San Vicente Blvd, Los Angeles, CA 90048. E-mail lydenp@cshs.org
(***Stroke*. 2017;48:513-519. DOI: 10.1161/STROKEAHA.116.015434.**)

Table 1. Evolution of the NIHSS

Item	Cincinnati/Naloxone NIHSS[11]			Current NIHSS[15]			Modified NIHSS[16]		
1a	Level of consciousness	Alert	0	Level of consciousness	Alert	0	Level of consciousness		
		Drowsy	1		Not alert, arousable	1			
		Stuporous	2		Not alert, obtunded	2			
		Coma	3		Unresponsive	3			
1b	LOC questions	Answers both correctly	0	LOC questions	Answers both correctly	0	LOC questions	Answers both correctly	0
		Answers one correctly	1		Answers one correctly	1		Answers one correctly	1
		Incorrect	2		Incorrect	2		Incorrect	2
1c	LOC commands	Obeys both correctly	0	LOC commands	Obeys both correctly	0	LOC commands	Obeys both correctly	0
		Obeys one correctly	1		Obeys one correctly	1		Obeys one correctly	1
		Incorrect	2		Incorrect	2		Incorrect	2
2	Pupillary response	Both reactive	0	Gaze	Normal	0	Gaze	Normal	0
		One reactive	1		Partial gaze palsy	1		Partial gaze palsy	1
		Neither reactive	2		Forced deviation	2		Total gaze palsy	2
3	Best **g**aze	Normal	0	Visual fields	No visual loss	0	Visual fields	No visual loss	0
		Partial gaze palsy	1		Partial hemianopsia	1		Partial hemianopsia	1
		Forced deviation	2		Complete hemianopsia	2		Complete hemianopsia	2
					Bilateral hemianopsia	3		Bilateral hemianopsia	3
4	Best visual	No visual loss	0	Facial palsy	Normal	0			
		Partial hemianopia	1		Minor paralysis	1			
		Complete hemianopia	2		Partial paralysis	2			
					Complete paralysis	3			
5	Facial palsy	Normal	0	Motor arm (a) Left (b) Right	No drift	0	Motor arm (a) Left (b) Right	No drift	0
		Minor	1		Drift before 10 s	1		Drift before 10 s	1
		Partial	2		Falls before 10 s	2		Falls before 10 s	2
		Complete	3		No effort against gravity	3		No effort against gravity	3
					No movement	4		No movement	4
6	Best motor arm	No drift	0	Motor leg (a) Left (b) Right	No drift	0	Motor leg (a) Left (b) Right	No drift	0
		Drift	1		Drift before 10 s	1		Drift before 5 s	1
		Cannot resist gravity	2		Falls before 10 s	2		Falls before 5 s	2
		No effort	3		No effort against gravity	3		No effort against gravity	3
					No movement	4		No movement	4
7	Best motor leg	No drift	0	Ataxia	Absent	0			
		Drift	1		One limb	1			
		Cannot resist gravity	2		Two limbs	2			
		No effort	3						
8	Plantar reflex	Normal	0	Sensory	Normal	0	Sensory	Normal	0
		Equivocal	1		Mild loss	1		Abnormal	1
		Extensor	2		Severe loss	2			
		Bilateral extensor	3						
9	Limb ataxia	Absent	0	Language	Normal	0	Language	Normal	0
		Present in upper or lower	1		Mild aphasia	1		Mild aphasia	1
		Present in both	2		Severe aphasia	2		Severe aphasia	2
					Mute or global aphasia	3		Mute or global aphasia	3

(*Continued*)

Table 1. Continued

Item	Cincinnati/Naloxone NIHSS[11]			Current NIHSS[15]			Modified NIHSS[16]		
10	Sensory	Normal	0	Dysarthria	Norma	0			
		Partial loss	1		Mild	1			
		Dense loss	2		Severe	2			
11	Neglect	No neglect	0	Extinction/ inattention	Normal	0	Neglect	Normal	0
		Partial neglect	1		Mild	1		Mild	1
		Complete neglect	2		Severe	2		Severe	2
12	Dysarthria	Normal articulation	0						
		Mild to moderate dysarthria	1						
		Near unintelligible or worse	2						
13	Best language	No aphasia	0						
		Mild to moderate	1						
		Severe aphasia	2						
		Mute	3						
14	Change from previous examination	Same	s						
		Better	b						
		Worse	w						
15	Change from baseline	Same	s						
		Better	b						
		Worse	w						

The original (Cincinnati/Naloxone) and the current (r-tPA) NIH Stroke Scales are shown to highlight the differences. An intermediate version (not shown) was used in the Pilot r-tPA for Acute Stroke Trial.[14] The instructions for the original version provided a stroke scale glossary. The current version uses a form for recording the data that contains detailed instructions for the use of the scale; the scale is not valid without the instructions physically attached to the scoring sheet, and simple summary sheets are likely not valid. The original r-tPA version of the NIHSS form and instructions are provided in the online-only Data Supplement after removing trial-specific data elements, as well as the original scoring manual. NIH indicates National Institutes of Health; NIHSS, National Institutes of Health Stroke Scale; and r-tPA, recombinant tissue-type plasminogen activator.

user will cooperate with extensive training prior to attempting certification. The scale is intended to be used with training to assure reproducibility: when the scale is used across clinical trial sites by users of differing skill levels, the results must be reproducible.[29] The accuracy of the scale—whether it captures each individual patient's deficit accurately—is secondary. Thus, the scale does not accurately reflect a patient's coordination; gait impairment; cortical sensory function; distal motor function; memory; or cognition. This lack of accuracy was designed intentionally as a sacrifice to gain reproducibility. If one wanted to accurately capture the deficits in each individual patient, one would do a standard neurological examination and write a detailed narrative.[30] The accuracy of such a narrative depends heavily on the training, skill, and interest of the examiner, so the results cannot be quantified and cannot be reproduced by untrained examiners of varying skill levels. In contrast, a simplified examination scale used in serial examinations of groups of patients shows excellent characterization of the group behavior over time.[31]

To gain reproducibility and to allow non-neurologists (emergency physicians and nurses) to participate in the Trial, scoring rules were designed to facilitate reproducibility (Table 2). For example, the cardinal rule in using the NIHSS is "score what you see, not what you think." In other words, a skilled neurologist would not down-score a patient with aphasia for failing to answer 2 questions about orientation—the neurologist would know that the aphasia prevented valid testing of orientation (item 1b; Table 1). Clinical trial designers could not assure that the non-neurologist MD, or the non-neuro-specialist RN, would do similarly in all circumstances. Therefore, the scoring rules were written to force the user to score a 1- or 2-point deficit, even in the face of obvious aphasia. This

Table 2. Selected Scoring Rules for the NIHSS

Item	Rule
All	Score what you see, not what you think
All	Score the first response, not the best response, except item 9 best language
All	Do not coach
1a	May be assessed casually while taking history
2	Only assess horizontal gaze
5 and 6	Count out loud and use your fingers to show the patient your count

A few selected scoring rules from the original NIHSS training manual are presented for illustration purposes. The first column refers to the NIHSS item to which the rule applies. Some of the rules are counterintuitive and are needed to assure reproducibility across multiple skill levels. The remainder of all rules and instructions are provided in the online-only Data Supplement. NIHSS indicates National Institutes of Health Stroke Scale.

scoring rule makes little sense neurologically—the aphasia is the problem and the patient does not have stupor or delirium—but the answers to item 1b will be reproducible.[32,33] To enhance reproducibility further, several more scoring rules were written that typically strike the skilled neurologist as counterintuitive (Table 2; online-only Data Supplement).

Certification and Training

Because the NIHSS instructions include counterintuitive scoring rules, training and certification would be critical, including actual demonstration of the scoring rules on live patients. At the time (late 1980s), video technology was emerging, and training videos were being produced for ongoing clinical trials.[34] Prior to beginning the Trial, a training videotape and 2 certification videotapes were produced (detailed methods provided in the online-only Data Supplement). All participants in the Trial were required to view the training tape and one certification tape and score each certification patient; only after passing central review and approval were investigators certified to enroll patients. To overcome bias introduced by the video technique, we designed a scoring system that accounted for the artificial limitations of the video viewing process.[15] Users who failed certification were asked to rewatch the training video and try again. After 6 months, all users were asked to view and score another certification tape to assure continued proficiency.

One intended consequence of the tapes was that new investigators could be added easily to the trial, an innovation at the time. However, an unintended consequence of this scoring system is that there were >1 correct responses to many of the case scenarios, creating an impression of leniency in the scoring.[35] Nevertheless, the scoring system does allow easy certification of online viewers.

Use in Trials

After the publication of the Trial in 1995, and regulatory approval of r-tPA for acute stroke in 1996, clinical trialists expressed interest in using the NIHSS for their clinical trials. Hundreds of the videotapes were produced and shipped; centralized scoring was done at Henry Ford Hospital. After a few years, the videotapes were replaced with training/certification digital videodiscs that demonstrated each NIHSS scale item and its scoring rules in detail (see detailed methods in the online-only Data Supplement). The NINDS took over responsibility for distributing the digital videodiscs to interested groups who were organizing large clinical trials, and the author provided central review and grading services using the scoring algorithm developed during the Trial.[36]

Today, most NIHSS training and certification is performed by 3 online services (Table 3). For scoring, all services use the published NINDS algorithm, as verified by the author.[36] None of the vendors require the student user to view the training video, despite evidence that such training is necessary.[20,37,38] Nevertheless, online certification has been validated and is ongoing.[2] As of February 2016, one of the sites had certified over 500 000 different student users, most of them multiple times. Recertification is generally required annually, although some clinical trial sponsors allow longer intervals; there is no data that supports any particular recertification schedule. The Training and Certification videos have been translated into multiple languages.[39–45] Generally, the testing materials (word list, sentences, and naming card) were translated literally, but in some cases, a more rigorous process was used.[46] For use in China, entirely new video was recorded using Chinese patients and investigators.[43] Although ideal, reshooting the video in each country would be prohibitively expensive, so in some countries, the English video was dubbed using actors.[44]

Modified Versions

There have been a few attempts to improve the accuracy of the NIHSS by removing items that lack sufficient reproducibility. The most validated modified NIHSS (Table 1) collapses items 3 and 4 into only normal/abnormal responses and eliminates the Ataxia item altogether.[16,47] The modified NIHSS is particularly well suited to applications in telemedicine.[48] In other modifications, an attempt has been made to shorten or simplify the scale or focus on a few, easy-to-teach items; although none of these shorter versions have been subjected to the same rigorous validation as the original scale, they may be useful in situations that do not require the rigor of a clinical trial.

Table 3. Certification and Training Products

Era	Technology	Type
NINDS Trial	Videotape[15]	Mandatory training tape Certification tape 1 (n=5) Certification tape 2 (n=6)
Post NINDS	Digital Videodisc[36]	Mandatory training DVD
Pre-Internet		Certification DVDs: Group A (n=6) Group B (n=6) Group C (n=6)
Internet	Web-based video streaming[2]: http://www.nihstrokescale.org/ https://learn.heart.org/nihss.aspx http://apexinnovations.com/NIHStrokeScale.html	Optional training video Certification video New group every year or 2 y
Future	Web-based	Mandatory training video Re-certify based on scoring pattern

There are multiple venues available for training and certification on the NIHSS using different technology. The first column, Era, refers to the time frame during which the technology was (or is) most relevant. Originally, during the NINDS r-tPA for Acute Stroke Trial era, videotapes were developed for training the investigators. After the Trial was published, and after wider adoption of the NIHSS, new videos were recorded on DVDs to provide a better training system. With the advent of internet video-streaming technology, there are 3 online services that provide the video and a certificate for successful training. In a future era, hopefully, recertification videos will be selected so as to optimize the learning experience for the user. DVD indicates digital videodiscs; NIHSS, National Institutes of Health Stroke Scale; NINDS, National Institute of Neurological Disorders and Stroke; and r-tPA, recombinant tissue-type plasminogen activator.

The NIHSS contains 4 factors in formal factor analysis.[16,32] These factors represent, as intended, the 2 cerebral hemispheres. Of particular significance, each hemisphere factor resolves into cortical and subcortical factors (Figure). This result, which has been replicated, suggests that the NIHSS serves its intended purpose: numbers are generated that quantify the function of the key brain areas above the tentorium. An alternative method to report NIHSS scores would be to generate factor scores. Such an approach would overcome the known predilection of the trial toward higher scores in left-hemisphere stroke.[17,49] Factor scores do not carry clinically intuitive meaning, however, and would not be accepted easily. A frequently cited weakness of the NIHSS is the failure to capture or quantify brain stem function, although this aspect of the scale was designed intentionally: most clinical trials exclude brain stem strokes because of their infrequency and possible differing natural history.

Use of the NIHSS Outside of Clinical Trials

In addition to certifying examiners involved in clinical trials, the NIHSS has been used in demographic and epidemiological studies. A reasonably accurate NIHSS can be reconstructed from well-documented neurological examinations recorded in medical records.[24] Extracted NIHSS scores may not be comparable to scores recorded by certified users working in the context of clinical trials, however.

In recent years, regulatory and payer agencies have required severity descriptors in stroke patients. Baseline severity score correlates with discharge disposition, mortality, and other outcomes.[50,51] To comply, large numbers of bedside clinicians are accessing NIHSS certification online (Table 3). These users do not view the training video, so it is unclear whether or how they come to understand the scoring rules; these users may be uninterested in the scoring rules and are unlikely to be involved in clinical research.[38] Thus, using the NIHSS for severity scoring by such individuals seems potentially fraught with error. At a minimum, regulators should be aware that NIHSS scores generated by casual, bedside users should not be compared with scores generated in clinical trials. On the other hand, bedside users could acquire sufficient proficiency using the NIHSS to communicate with each other. A patient's total NIHSS score portrays a vaguely accurate description of the patient equivalent to the descriptors mild, moderate, or severe stroke. In this context, it may not matter that the user does not understand the proper method to perform the scale or to use the scoring rules. A few critical steps in using the scale at the bedside are summarized in Table 2.

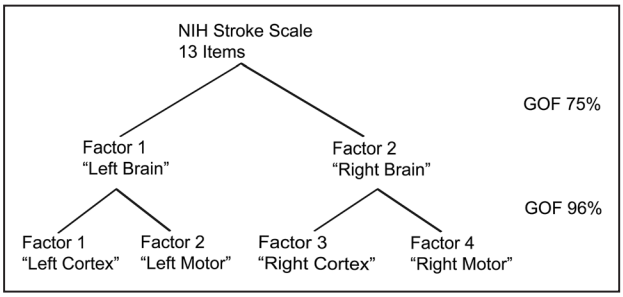

Figure. Factor analysis of the NIHSS. Using principal components factor analysis, there are 2 main factors underlying the NIHSS, corresponding to right and left hemisphere respectively. The 4-factor solution can be viewed as a subset or refinement of the 2-factor solution in which motor function in each hemisphere separates from other functions. GOF indicates goodness of fit; and NIHSS, National Institutes of Health Stroke Scale. Reprinted from Lyden et al[33] with permission of the publisher. Copyright ©1999, American Heart Association, Inc.

Because of the scoring rules, certain scores are impossible to obtain, especially at the higher end. For example, the score contains 42 possible points were a patient to score the worst on all items, but this cannot happen. In a coma patient, certain scores default to 0, for example, item 7, ataxia (Table 4), and the maximum score in a comatose patient is 39.

After the publication of major neurothrombectomy trials, there is renewed interest in using the NIHSS to select patients most likely eligible for thrombectomy.[52] Although there is a good correlation between NIHSS and likelihood of finding an eligible large vessel occlusion, no specific cut point of the NIHSS seems optimal for field use.[22,53] Many agencies seek to use a field assessment for triaging patients to a comprehensive stroke center; at this time, neither the full scale nor any derived scale has sufficient sensitivity and specificity to be used in this way.[54] Nevertheless, a baseline NIHSS is useful in identifying patients more likely to have an eligible lesion, even though it may not be good enough for field triage in which some patients may be diverted away from appropriate resources.

Table 4. Scoring the NIHSS for a Patient in Coma

Item	Score
1a	3 (defines coma)
1b	2
1c	2
2	0, 1, or 2
3	0, 1, or 2
4	3
5a	4
5b	4
6a	4
6b	4
7	0
8	2
9	3
10	2
11	2
Total	35–39

A patient who scores 3 on item 1a (level of consciousness) is considered to be in a coma. A patient in coma should be stimulated by rubbing on the chest or by using a painful stimulus. A 3 is scored for item 1a only if the patient makes no movement (other than reflexive posturing) in response to the noxious stimulation. Patients who appear to be in coma and who score <3 must be tested on all scale items. (Excerpt from NINDS Manual of Procedures, "The NIH Stroke Scale," provided in the online-only Data Supplement).[1] Once the patient is clearly found to be in coma, the prespecified (and for some items arbitrary) values are used for each item. NIH indicates National Institutes of Health; NIHSS, National Institutes of Health Stroke Scale; and NINDS, National Institute of Neurological Disorders and Stroke.

Future Studies

No data exist to determine whether widespread use of the NIHSS at the bedside yields scores that are reproducible or whether users certifying without training use the scale correctly. If typical bedside use of the NIHSS today is unreliable, considerable effort will be needed to design an effective training strategy. It may be necessary to alter the online web-based training sites so that training is required before users can certify. Also, research is needed to determine how often users should be required to recertify. At the moment, annual or biannual recertification seems most common, but regulators should be aware that there is no data to support such timelines: recertification may be best if it occurs more or less often or on a progressive timeframe based on past performance. Over a longer period of time, say after 3 or 4 recertifications, perhaps it should be mandatory to rereview the training materials. Further studies are sorely needed to determine whether certified users make more errors over time, as the interval from training lengthens.

Severity adjustment of outcomes is essential in modern health care. Publicly reported outcomes (mortality, 30-day readmission) must be understood in context of stroke severity.[5,50,51] Repeatedly shown, the primary drivers of long-term outcome after stroke are initial severity—almost always quantified with the NIHSS—age, and a few comorbidities, such as diabetes mellitus. Given the profound impact of baseline stroke severity on outcome, it would seem essential that casual bedside users of the NIHSS understand the design limitations, proper technique, and scoring rules. Although the NIHSS was designed for use in clinical trials, severity scoring has grown far beyond the rigorous boundaries required of stroke research teams. More serious consideration must be given to selecting the best professionals for recording the baseline severity score at hospital admission because such scores will powerfully influence that hospital's outcomes—many of which are publically reported.

Conclusions

The NIHSS in current use evolved from an earlier version that is no longer used. The scale now used (Table 1) was designed to be reproducible when used by physicians and nurses seeking to participate in clinical trials, and may be useful in clinical practice with appropriate training and certification. Scores for left hemisphere stroke exceed right hemisphere by four points, so severity scoring must include the side of the infarct. Online video training and certification systems are available and widely used. Use of the NIHSS by casual (nonresearch) bedside users has not been extensively validated, however, and the NIHSS should be used with caution outside of a research trial for rating stroke patients' severity. Regulators seeking to add a severity adjustment to administrative data should approach the NIHSS with a full understanding of its limitations.

Acknowledgments

I am grateful to Karen Rapp, RN, for a critical review of the article.

Sources of Funding

This study was supported by the Carmen and Louis Warschaw Foundation.

Disclosures

None.

References

1. The National Institute of Neurological Disorders and Stroke rt-PA Stroke Study Group. Tissue plasminogen activator for acute ischemic stroke. *N Engl J Med*. 1995;333:1581–1587.
2. Lyden P, Raman R, Liu L, Emr M, Warren M, Marler J. National Institutes of Health Stroke Scale certification is reliable across multiple venues. *Stroke*. 2009;40:2507–2511. doi: 10.1161/STROKEAHA.108.532069.
3. Commission TJ. Comprehensive Stroke Certification: Standardized Performance Measures. 2016. https://www.jointcommission.org/performance_measures_for_comprehensive_stroke_centers/. Accessed September 15, 2016.
4. Leifer D, Bravata DM, Connors JJ 3rd, Hinchey JA, Jauch EC, Johnston SC, et al; American Heart Association Special Writing Group of the Stroke Council; Atherosclerotic Peripheral Vascular Disease Working Group; Council on Cardiovascular Surgery and Anesthesia; Council on Cardiovascular Nursing. Metrics for measuring quality of care in comprehensive stroke centers: detailed follow-up to Brain Attack Coalition comprehensive stroke center recommendations: a statement for healthcare professionals from the American Heart Association/American Stroke Association. *Stroke*. 2011;42:849–877. doi: 10.1161/STR.0b013e318208eb99.
5. Fonarow GC, Pan W, Saver JL, Smith EE, Reeves MJ, Broderick JP, et al. Comparison of 30-day mortality models for profiling hospital performance in acute ischemic stroke with vs without adjustment for stroke severity. *JAMA*. 2012;308:257–264. doi: 10.1001/jama.2012.7870.
6. Marsh EB, Lawrence E, Gottesman RF, Llinas RH. The NIH Stroke Scale has limited utility in accurate daily monitoring of neurologic status. *Neurohospitalist*. 2016;6:97–101. doi: 10.1177/1941874415619964.
7. Lyden PD, Lau GT. A critical appraisal of stroke evaluation and rating scales. *Stroke*. 1991;22:1345–1352.
8. Muir KW, Weir CJ, Murray GD, Povey C, Lees KR. Comparison of neurological scales and scoring systems for acute stroke prognosis. *Stroke*. 1996;27:1817–1820.
9. Gilroy J, Barnhart MI, Meyer JS. Treatment of acute stroke with dextran 40. *JAMA*. 1969;210:293–298.
10. Norris JW. Steroid therapy in acute cerebral infarction. *Arch Neurol*. 1976;33:69–71.
11. 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;20:864–870.
12. Brott T, Marler JR, Olinger CP, Adams HP Jr, Tomsick T, Barsan WG, et al. Measurements of acute cerebral infarction: lesion size by computed tomography. *Stroke*. 1989;20:871–875.
13. Olinger CP, Adams HP Jr, Brott TG, Biller J, Barsan WG, Toffol GJ, et al. High-dose intravenous naloxone for the treatment of acute ischemic stroke. *Stroke*. 1990;21:721–725.
14. Brott TG, Haley EC Jr, Levy DE, Barsan W, Broderick J, Sheppard GL, et al. Urgent therapy for stroke. Part I. Pilot study of tissue plasminogen activator administered within 90 minutes. *Stroke*. 1992;23:632–640.
15. 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;25:2220–2226.
16. 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.
17. Woo D, Broderick JP, Kothari RU, Lu M, Brott T, Lyden PD, et al. Does the National Institutes of Health Stroke Scale favor left hemisphere strokes? NINDS t-PA Stroke Study Group. *Stroke*. 1999;30:2355–2359.
18. Fink JN, Selim MH, Kumar S, Silver B, Linfante I, Caplan LR, et al. Is the association of National Institutes of Health Stroke Scale scores and acute magnetic resonance imaging stroke volume equal for patients with right- and left-hemisphere ischemic stroke? *Stroke*. 2002;33:954–958.
19. 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. *Cerebrovasc Dis*. 1999;9:323–327. doi: 10.1159/000016006.

20. Schmülling S, Grond M, Rudolf J, Kiencke P. Training as a prerequisite for reliable use of NIH Stroke Scale. *Stroke*. 1998;29:1258–1259.
21. 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.
22. Heldner MR, Zubler C, Mattle HP, Schroth G, Weck A, Mono ML, et al. National Institutes of Health stroke scale score and vessel occlusion in 2152 patients with acute ischemic stroke. *Stroke*. 2013;44:1153–1157. doi: 10.1161/STROKEAHA.111.000604.
23. Kharitonova T, Mikulik R, Roine RO, Soinne L, Ahmed N, Wahlgren N; Safe Implementation of Thrombolysis in Stroke Investigators. Association of early National Institutes of Health Stroke Scale improvement with vessel recanalization and functional outcome after intravenous thrombolysis in ischemic stroke. *Stroke*. 2011;42:1638–1643. doi: 10.1161/STROKEAHA.110.606194.
24. 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.
25. Adams RJ, Meador KJ, Sethi KD, Grotta JC, Thomson DS. Graded neurologic scale for use in acute hemispheric stroke treatment protocols. *Stroke*. 1987;18:665–669.
26. Lindenstrøm E, Boysen G, Waage Christiansen L, Würtzen Nielsen P. Reliability of Scandinavian neurological stroke scale. *Cerebrovasc Dis*. 1991;1:103–107.
27. Orgogozo JM, Asplund K, Boysen G. A unified form for neurological scoring of hemispheric stroke with motor impairment. *Stroke*. 1992;23:1678–1679.
28. Côté R, Battista RN, Wolfson C, Boucher J, Adam J, Hachinski V. The Canadian Neurological Scale: validation and reliability assessment. *Neurology*. 1989;39:638–643.
29. Feinstein AR. *Clinimetrics*. New Haven: Yale University Press; 1987.
30. Caplan L. Localization in clinical neurology. *JAMA*. 1986;255:413–415.
31. Wityk RJ, Pessin MS, Kaplan RF, Caplan LR. Serial assessment of acute stroke using the NIH Stroke Scale. *Stroke*. 1994;25:362–365.
32. 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. *Arch Neurol*. 2004;61:1677–1680. doi: 10.1001/archneur.61.11.1677.
33. 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. NINDS tPA Stroke Trial Investigators. *Stroke*. 1999;30:2347–2354.
34. Albanese MA, Clarke WR, Adams HP Jr, Woolson RF. Ensuring reliability of outcome measures in 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.
35. Hills NK, Josephson SA, Lyden PD, Johnston SC. Is the NIHSS certification process too lenient? *Cerebrovasc Dis*. 2009;27:426–432. doi: 10.1159/000209237.
36. 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;36:2446–2449. doi: 10.1161/01.STR.0000185725.42768.92.
37. Josephson SA, Hills NK, Johnston SC. NIH Stroke Scale reliability in ratings from a large sample of clinicians. *Cerebrovasc Dis*. 2006;22:389–395. doi: 10.1159/000094857.
38. André C. The NIH Stroke Scale is unreliable in untrained hands. *J Stroke Cerebrovasc Dis*. 2002;11:43–46. doi: 10.1053/jscd.2002.123974.
39. Prasad K, Dash D, Kumar A. Validation of the Hindi version of National Institute of Health Stroke Scale. *Neurol India*. 2012;60:40–44. doi: 10.4103/0028-3886.93587.
40. Domínguez R, Vila JF, Augustovski F, Irazola V, Castillo PR, Escalante RR, et al. Spanish cross-cultural adaptation and validation of the National Institutes of Health Stroke Scale. *Mayo Clinic Proc*. 2006;81:476–480.
41. Oh MS, Yu KH, Lee JH, Jung S, Ko IS, Shin JH, et al. Validity and reliability of a korean version of the national institutes of health stroke scale. *J Clin Neurol*. 2012;8:177–183. doi: 10.3988/jcn.2012.8.3.177.
42. Berger K, Weltermann B, Kolominsky-Rabas P, Meves S, Heuschmann P, Böhner J, et al. [The reliability of stroke scales. The german version of NIHSS, ESS and Rankin scales]. *Fortschr Neurol Psychiatr*. 1999;67:81–93. doi: 10.1055/s-2007-993985.
43. Cheung RT, Lyden PD, Tsoi TH, Huang Y, Liu M, Hon SF, et al. Production and validation of Putonghua- and Cantonese-Chinese language National Institutes of Health Stroke Scale training and certification videos. *Int J Stroke*. 2010;5:74–79. doi: 10.1111/j.1747-4949.2010.00411.x.
44. Pezzella FR, Picconi O, De Luca A, Lyden PD, Fiorelli M. Development of the Italian version of the National Institutes of Health Stroke Scale: It-NIHSS. *Stroke*. 2009;40:2557–2559. doi: 10.1161/STROKEAHA.108.534495.
45. Hussein HM, Abdel Moneim A, Emara T, Abd-Elhamid YA, Salem HH, Abd-Allah F, et al. Arabic cross cultural adaptation and validation of the National Institutes of Health Stroke Scale. *J Neurol Sci*. 2015;357:152–156. doi: 10.1016/j.jns.2015.07.022.
46. Rendu E, Cooke C, Lyden P, Acquadro C, Conway K. Translation of the National Institutes of Health Stroke Scale (NIHSS) List of Words: a challenging task. *Qual Life Res*. 2012;20:31–32.
47. 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.
48. 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;64:1058–1060. doi: 10.1212/01.WNL.0000154601.26653.E7.
49. Fink JN, Frampton CM, Lyden P, Lees KR; Virtual International Stroke Trials Archive Investigators. Does hemispheric lateralization influence functional and cardiovascular outcomes after stroke?: an analysis of placebo-treated patients from prospective acute stroke trials. *Stroke*. 2008;39:3335–3340. doi: 10.1161/STROKEAHA.108.523365.
50. Schlegel DJ, Tanne D, Demchuk AM, Levine SR, Kasner SE; Multicenter rt-PA Stroke Survey Group. Prediction of hospital disposition after thrombolysis for acute ischemic stroke using the National Institutes of Health Stroke Scale. *Arch Neurol*. 2004;61:1061–1064. doi: 10.1001/archneur.61.7.1061.
51. 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;34:134–137.
52. Goyal M, Menon BK, van Zwam WH, Dippel DW, Mitchell PJ, Demchuk AM, et al; HERMES Collaborators. Endovascular thrombectomy after large-vessel ischaemic stroke: a meta-analysis of individual patient data from five randomised trials. *Lancet*. 2016;387:1723–1731. doi: 10.1016/S0140-6736(16)00163-X.
53. Heldner MR, Hsieh K, Broeg-Morvay A, Mordasini P, Bühlmann M, Jung S, et al. Clinical prediction of large vessel occlusion in anterior circulation stroke: mission impossible? *J Neurol*. 2016;263:1633–1640. doi: 10.1007/s00415-016-8180-6.
54. Hastrup S, Damgaard D, Johnsen SP, Andersen G. Prehospital acute stroke severity scale to predict large artery occlusion: design and comparison with other scales. *Stroke*. 2016;47:1772–1776. doi: 10.1161/STROKEAHA.115.012482.

KEY WORDS: clinimetrics ■ ischemic ■ National Institutes of Health ■ outcome ■ stroke ■ stroke management