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.

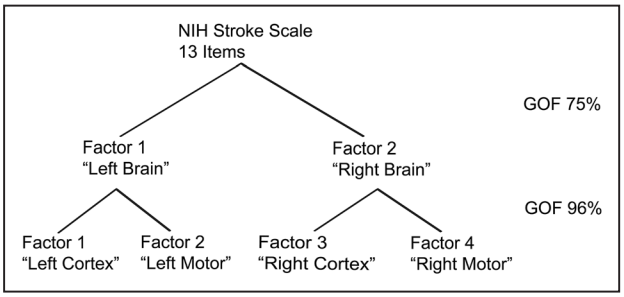


Figure. Factor analysis of the NIHSS. Using principal components factor analysis, there are 2 main factors underlying the NIHSS, corresponding to right and left hemisphere respectively. The 4-factor solution can be viewed as a subset or refinement of the 2-factor solution in which motor function in each 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

Using the National Institutes of Health Stroke Scale: A Cautionary Tale

Patrick Lyden

Stroke. 2017;48:513-519; originally published online January 11, 2017;
doi: 10.1161/STROKEAHA.116.015434
Stroke is published by the American Heart Association, 7272 Greenville Avenue, Dallas, TX 75231

Print ISSN: 0039-2499. Online ISSN: 1524-4628

SUPPLEMENTAL MATERIAL

1. Scoring guidelines for the NIHSS. This is “Chapter 5” from the Manual of Procedures for the original NINDS rt-PA for Acute Stroke Trial.
2. The NIHSS. This form is the only validated version, and includes the instructions for each item printed directly on the scoring sheet. The form has been modified slightly (header items specific to the original trial deleted) to make it more generic and usable.
3. Filming the NIHSS. This document summarizes key facts about the filming of the videos used to train and certify users. These videos are used by all on-line NIHSS certification vendors. During filming, a number of issues were resolved that may be of interest to regulators intending to require NIHSS utilization by non-research health care personnel.

This is the original scoring manual developed for the NINDS Trial of Rt-PA for Acute Stroke. Some changes many have been made since publication (1989).

Chapter 5 - The NIH Stroke Scale

5.1 Overview

The Stroke Scale is a standardized neurological examination intended to describe the neurological deficits found in large groups of stroke patients participating in treatment trials. The instructions contained in this manual reflect primary concern for reproducibility. The goal is to have multiple examiners at different sites rate patients similarly. It is possible to challenge the scale on sub-items, and competent neurologists will disagree over the "best" method for testing some items in individual patients. Nevertheless, our interest in reproducibility among many observers in a large multi-center study is paramount, and to this end, all examiners at all sites must use the scale uniformly. We recognize that for some examiners, this means that some testing may be done one way for the study, and a different way in usual clinical practice. The consolation for this disparity is the knowledge that the reproducibility among examiners using this scale will (hopefully) be extremely high.

There are four general principals underlying the scale in its present form:

1. The most reproducible response is generally the first response. For example, on LOC questions, the patient is asked to state age and the current month. The patient who initially responds incorrectly, but later corrects himself, is scored as having given an incorrect response. This approach is critical, because we have no way of standardizing the myriad verbal and non-verbal cues that might be given to patients to promote a correction of an initially incorrect response.

2. It is not permissible to coach patients on any item unless specified in the instructions. This contradicts neurological teaching, since we are generally interested in a patient's best possible performance. Again, standardization of coaching is not possible, and coaching must be avoided in the interest of reproducibility.

3. Some items are scored only if definitely present. For example, ataxia is scored as absent in the patient with hemiplegia, because it is not definitely present at the time of examination. Although somewhat counter intuitive to some physicians, the item must be scored this way to avoid ambiguity and ensure reproducibility.

4. Most importantly, record what the patient does, not what you think the patient can do even if the findings appear contradictory. Many times a competent examiner forms an impression of the patient's level of function, but this impression must not influence scoring. Scoring should include prior deficits except for the sensory item (see instructions)

The patient's scores should be recorded immediately after the examination, and preferably, each item should be coded as you go through the scale. This is especially necessary at baseline. If baseline results are recorded after the patient has received medication, the examiner may be influenced by the patient's response.

5.2 Certification

Any investigator completing the NIH Stroke Scale for the trial must be certified. Any experienced Clinical Center Personnel (physician or nurse or physician's assistant) may be certified.

5.2.1 Requirements for Certification

Certification requires:

- Review of the NIH Stroke Scale Training Tape
- Completion of NIH Stroke Scales for the five patients shown on the NIH Stroke Scale Certification Tape #1
- Submission of the five completed forms to the Coordinating Center for review
- Approval by the Coordinating Center

5.2.2 Retention of Certification

Retention of certification by a certified investigator requires:

- Completion of the NIH Stroke Scales on the six patients shown on the NIH Stroke Scale Certification Tape #2 (approximately six months after the initial certification)
- Submission of the six completed forms to the Coordinating Center for review.
- Approval by the Coordinating Center

5.3 The NIH Stroke Scale (Form 5)

PURPOSE: This form collects data representing the primary endpoints of the trial (difference from Baseline at 24 hour and 3 month NIH Stroke Scale data). A more complete discussion of purpose is given in the overview (Section 5.1)

WHEN: The Stroke Scale is completed at baseline **PRIOR TO TREATMENT**, at 2 hours $\pm$ 5 minutes post treatment onset, 24 hours $\pm$ 20 minutes post stroke onset, 7 to 10 days and 3 months $\pm$ 2 weeks.

PERSON COMPLETING FORM: The Stroke Scale must be completed by a certified trial investigator.

All stroke scales except for the 2-hour scale must be completed by a certified trial investigator. If possible, the 2-hour scale should also be completed by certified trial personnel. If it is impractical to have a certified investigator perform the 2-hour scale, it may be performed by an uncertified person with telephone supervision by a certified investigator.

The 24-hour NIH Stroke Scale **must** be done by a certified investigator who was not present during treatment. ***The 3-month scale must*** be done by a certified investigator who was not present during treatment. (The two time points are the primary endpoints of the trial.)

INSTRUCTIONS: Extensive instructions are included on the NIH Stroke Scale Form. Additional comments follow:

Three items are used to assess the patient's level of consciousness. It is vital that the items be asked in a standardized manner, as illustrated in the Stroke Scale training tape. Responses must be graded based on what the patient does first. Do not give credit if the patient corrects himself/herself and do not give any clues or coaching.

1a. Level of Consciousness

Instructions:
The investigator must choose a response, even if a full evaluation is prevented by such obstacles as an endotracheal tube, language barrier, orotracheal trauma/bandages. A 3 is scored only if the patient makes no movement (other than reflexive posturing) in response to noxious stimulation.

Comments:
Ask the patient two or three general questions about the circumstances of the admission. Also, prior to beginning the scale, it is assumed that the examiner will have queried the patient informally about the medical history. Based on the answers, score the patient using the 4 point scale on the Stroke Scale form. Remember not to coach. A score of 3 is reserved for the severely impaired patient who makes, at best, reflex posturing movements in response to repeated painful stimuli. If it is difficult to choose between a score of 1 or 2, continue to question the patient about historical items until you feel comfortable in assessing level of consciousness.

1b. LOC Questions

Instructions:
The patient is asked the month and his/her age. The answer must be correct - there is no partial credit for being close. Aphasic and stuporous patients who do not comprehend the questions will score 2. Patients unable to speak because of endotracheal intubation, orotracheal trauma, severe dysarthria from any cause, language barrier or any other problem not secondary to aphasia are given a 1. It is important that only the initial answer be graded and that the examiner not "help" the patient with verbal or non-verbal cues.
Comments:
Ask the patient "how old are you now" and wait for a response. Then ask "what month is it now" or "what month are we in now". Count the number of incorrect answers and do not give credit for being "close". Patients who cannot speak are allowed to write. **Do not give a list of**

possible responses from which to choose the correct answer. This may coach the patient. Only the initial answer is graded. This item is never marked "untestable". (Note: On Certification Tape #1 an intubated patient was given a series of responses from which to choose, but the score for this patient would still be 1.) Deeply comatose (1a=3) patients are given a 2.

1c. LOC Commands:

Instructions:
The patient is asked to open and close the eyes and then to grip and release the non-paretic hand. Substitute another one step command if the hands cannot be used. Credit is given if an unequivocal attempt is made but not completed due to weakness. If the patient does not respond to command, the task should be demonstrated to them (pantomime) and score the result (i.e., follows none, one or two commands). Patients with trauma, amputation, or other physical impediments should be given suitable one-step commands. Only the first attempt is scored.

Comments:
Say to the patient "open your eyes...now close your eyes" and then "Make a fist...now open your hand". Use the non-paretic limb. If amputation or other physical impediment prevents the response, use another suitable one step command. The priming phrase is not scored, and these are used only to set the eyes or hand in a testable position. That is, the patient may be asked first to open the eyes if they are closed when you begin the test. Scoring is done on the second phrase "close your eyes". Count the number of incorrect responses and give credit if an unequivocal attempt is made to perform the operative task, but is not completed due to weakness, pain or other obstruction. Only the first attempt is scored and the questions should be asked only once.

Item 2:Best Gaze

Instructions:

Only horizontal eye movements will be tested. Voluntary or reflexive (oculocephalic) eye movements will be scored but caloric testing is not done. If the patient has a conjugate deviation of the eyes that can be overcome by voluntary or reflexive activity, the score will be 1. If a patient has an isolated peripheral nerve paresis (CN III, IV or VI) score a 1. Gaze is testable in all aphasic patients. Patients with ocular trauma, bandages, pre-existing blindness or other disorder of visual acuity or fields should be tested with reflexive movements and a choice made by the investigator. Establishing eye contact and then moving about the patient from side to side will occasionally clarify the presence of a partial gaze palsy.

Comments:
The purpose of this item is to observe and score horizontal eye movements. To this end, use voluntary or reflexive stimuli and record a score of 1 if there is an abnormal finding in one or both eyes. A score of two is reserved for forced eye deviation that cannot be overcome by the oculocephalic maneuver. Do not do caloric testing. In aphasic or confused patients it is helpful to establish eye contact and move about the bed.

This item is an exception to the rules of using the first observable response and not coaching. In the patient who fails voluntary gaze, the oculocephalic maneuver, eye fixation, and tracking with the examiner's face, are used to provide stronger testing stimuli.

Item 3:Visual

Instructions:
Visual fields (**upper and lower quadrants**) are tested by confrontation, using finger counting or visual threat as appropriate. Patient must be encouraged, but if they look at the side of the moving fingers appropriately, this can be scored as normal. If there is unilateral blindness or enucleation, visual fields in the remaining eye are scored. Score 1 only if a clear-cut asymmetry, including quadrantanopia is found. If patient is blind from any cause score 3. Double simultaneous stimulation is performed at this point. If there is extinction patient receives a 1 and the results are also used to answer question 11.

Comments:
Visual fields are tested exactly as demonstrated in the training video. Use finger counting or movement to confrontation and evaluate upper and lower quadrants separately. A score of 3 is reserved for blindness from any cause, including cortical blindness. A score of 2 is reserved for a complete hemianopia, and any partial visual field defect, including quadrantanopia, scores a 1.

Item 4:Facial Palsy

Instructions:
Ask, or use pantomime to encourage the patient to show teeth or raise eyebrows and close eyes. Score symmetry of grimace in response to noxious stimuli in the poorly responsive or non-comprehending patient. If facial trauma/bandages, orotracheal tube, tape or other physical barrier obscures the face, these should be removed to the extent possible.

Comments:
Ask the patient "Show me your teeth...now raise your eyebrows...now close your eyes tightly". Assess the response to noxious stimulation in the aphasic or confused patient. A useful approach to scoring may be as follows: score a 2 for any clear cut upper motor neuron facial palsy. Normal function must be clearly demonstrated to obtain the score of 0. Anything in between, including flattened nasolabial fold, is scored a 1. The severely obtunded or comatose patient; patients with bilateral paresis, patients with unilateral lower motor neuron facial weakness would receive a score of 3.

Items 5 & 6: Motor Arm and Leg

Instructions:
Each limb is tested in turn, beginning with the non-paretic arm, if known. The limb is placed in the appropriate position: extend the arm (palm down) 90 degrees (if sitting) or 45 degrees (if supine) and the leg 30 degrees (always tested supine). Drift is scored if the arm falls before 10 seconds or the leg before 5 seconds. The aphasic patient is encouraged using urgency in the voice and pantomime but not noxious stimulation. Only in the case of amputation or joint fusion at the shoulder or hip may the score be “untestable” and the examiner must clearly write the explanation for scoring as a “untestable”.

Comments:
Perform the test for weakness as illustrated in the video. When testing arms, palms must be down. Count out loud to the patient, until the limb actually hits the bed or other support. The score of 3 is reserved for the patient who exhibits no strength whatsoever, but does minimally move the limb on command when it is resting on the bed. The aphasic patient may understand

what you are testing if you use the non-paretic limb first. Do not test both limbs simultaneously. Be watchful for an initial dip of the limb when released. Only score abnormal if there is a drift after the dip.

Do not coach the patient verbally. Count out loud in strong voice and indicate count using your fingers in full view of the patient. Begin counting the instant you release the limb. (Note that on some of the video illustrated patients, the examiners erroneously delay seconds before beginning to count).

When testing motor leg the patient must be in the supine position to fully standardize the effect of gravity. Note that the examiner is no longer asked to identify the paretic arm or leg. The examiner's assessment of the side of the stroke is given on the Treatment Form (Form 7).

Item 7:Limb Ataxia

Instructions:
This item is aimed at finding evidence of a unilateral cerebellar lesion. Test with eyes open. In case of visual defect, insure testing is done in the intact visual field. The finger-nose-finger and heel-shin tests are performed on both sides, and ataxia is scored only if present out of proportion to weakness. Ataxia is absent in the patient who cannot understand or is paralyzed. Although the use of untestable is discouraged, in the case of amputation, joint fusion or some fractures, the item may be scored “untestable”, and the examiner must clearly write the explanation for not scoring. In case of blindness test by touching nose from extended arm position.

Comments:
Ataxia must be clearly present out of proportion to any weakness. Using the finger-nose-finger and the heel-test, count the number of ataxic limbs, up to a maximum of two. The aphasic patient will often perform the test normally if first the limb is passively moved by the examiner. Otherwise, the item is scored 0 for absent ataxia. If the weak patient suffers mild ataxia, and you cannot be certain that it is out of proportion to the weakness, give a score of 0. Remember this is scored positive only when ataxia is present.

Item 8:Sensory

Instructions:
Sensation or grimace to pin prick when tested, or withdrawal from noxious stimulus in the obtunded or aphasic patient. Only sensory loss attributed to stroke is scored as abnormal and the examiner should test as many body areas [arms (not hands), legs, trunk, face] as needed to accurately check for hemisensory loss. A score of 2, "severe or total," should only be given when a severe or total loss of sensation can be clearly demonstrated. Stuporous and aphasic patients will therefore probably score 1 or 0. The patient with brain stem stroke who has bilateral loss of sensation is scored 2. If the patient does not respond and is quadriplegic score 2. Patients in coma (item 1a=3) are arbitrarily given a 2 on this item.

Comments:
Do not test limb extremities, i.e., hands and feet when testing sensation because an unrelated neuropathy may be present. Do not test through clothing.

Item 9: Best Language

Instructions:
A great deal of information about comprehension will be obtained during the preceding sections of the examination. The patient is asked to describe what is happening in the attached picture, to name the items on the attached naming sheet, and to read from the attached list of sentences. Be complete. Have the patient name all items on the naming sheet and read all phrases on the two reading sheets. Comprehension is judged from responses here as well as to all of the commands in the preceding general neurological exam. If visual loss interferes with the tests, ask the patient to identify objects placed in the hand, repeat, and produce speech. The intubated patient should be asked to write. The patient in coma (question 1a=3) will arbitrarily score 3 on this item. The examiner must choose a score in the patient with stupor or limited cooperation but a score of 3 should be used only if the patient is mute and follows no one step commands.

Comments:
It is anticipated that most examiners will be ready to score this item based on information obtained during the history taking and the 8 prior items. The attached picture and naming sheet therefore should be used to confirm your impression. It is common to find unexpected difficulties when the formal testing is done, and therefore every patient must be tested with the picture, naming sheet, and sentences. The score of 3 is reserved for the globally mute or comatose patient. Mild aphasia would score a 1. To choose between a score of 1 or 2 use all the provided materials; it is anticipated that a patient who missed more than two thirds of the naming objects and sentences or who followed only very few and simple one step commands would score a 2.

This item is an exception to the rule that the first response is used, since several different tools are used to assess language. The stroke scale form contains lengthy examples of the defects associated with each score because of the great potential for variability in answering this question.**Item 10:Dysarthria**

Instructions:

If the patient is thought to be normal, an adequate sample of speech must be obtained by asking patient to read or repeat words from the attached list. If the patient has severe aphasia, the clarity of articulation of spontaneous speech can be rated. Only if the patient is intubated or has other physical barrier to producing speech, may the item be scored “untestable”, and the examiner must clearly write an explanation for not scoring. Do not tell the patient why he/she is being tested.

Comments:

Use the attached word list in all patients and do not tell the patient that you are testing clarity of speech. It is **common** to find slurring of one or more words in patients one might otherwise score as normal. The score of 0 is reserved for patients who read all words without any slurring. Aphasic patients and patients who do not read may be scored based on listening to the speech that they do produce or by asking them to repeat the words after you read them out loud. The score of 2 is reserved for the patient who cannot be understood in any meaningful way, or who is mute.

On this question, normal speech must be identified to score a 0, so the unresponsive patient receives the score of 2.

Item 11:Extinction and inattention

Instructions:

Sufficient information to identify neglect may be obtained during the prior testing. If the patient has a severe visual loss preventing visual double simultaneous stimulation, and the cutaneous stimuli are normal, the score is normal. If the patient has aphasia but does appear to attend to both sides, the score is normal. The presence of visual spatial neglect or anosagnosia may also be taken as evidence of abnormality. Since the abnormality is scored only if present, the item is never untestable.

Comments:

This item is open to significant variation among examiners, and all neurologists have slightly different methods of assessing neglect. Therefore, to the extent possible, test only double simultaneous stimulation to visual and tactile stimuli and score 2 if one side extinguishes to both modalities, a 1 if only to one modality. If the patient does not extinguish, but does show other well developed evidence of neglect, score a 1.

5.4 Coma

A patient with a 3 on Item 1a (Level of Consciousness) is considered to be in a coma. A patient suspected to be in coma should be stimulated by rubbing on the chest or by using a painful stimulus. A 3 is scored 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 less than 3 must be tested on all items.**

For patients scoring a 3 on Item 1a, the remaining items should be scored as follows:

Item 1b (LOC Questions) - Score 2.

Item 1c (LOC Commands) - Score 2.

Item 2 (Best Gaze) - Patient can be in coma and have gaze palsy that can be overcome by moving the head. Thus the occulocephalic maneuver must be done and the patient scored.

Item 3 (Visual) - Test using bilateral threat.

Item 4 (Facial Palsy) - Score 3.

Items 5 and 6 (Motor Arm and Leg) - This item is interpreted as the voluntary ability to attain a posture. Score 4 for both arm and leg.

Item 7 (Limb Ataxia) - Scored only if present, out of proportion to weakness. Score 0.

Item 8 (Sensory) - Score 2 (arbitrary).

Item 9 (Best Language) - Score 3.

Item 10 (Dysarthria) - Score 2.

Item 11 (Extinction and inattention) - Coma implies loss of all cognitive abilities. Score 2.

5.5 Persons Who Refuse to Cooperate

In the event that a patient refuses to perform the tasks in the course of the examination resulting in an item untested, a detailed explanation must be clearly written on the form. All untested items will be reviewed by the medical monitor and discussed with the examiner if necessary.

5.6 Calculating a Score

In computing a score, the following items should **not** be added to the total:

- For Item 7 (Limb Ataxia) codes for affected sides (right and/or left arm and leg;

  1 = yes, 2 = no, 9 = untestable).

■ Distal Motor Function.

■ Any 9's.

5.7 Outliers

There are questions in the certification Tapes 1 and 2 that do not have a single answer. Thus the distribution of responses from those who have completed the certification is used. A response given by 12% or fewer examiners is considered an outlier. Any examiner having 10 or more outliers for Tape 1 or 12 or more outliers for Tape 2 is not certified to do stroke scales for the trial. An examiner who is not certified must redo the certification before they can perform stroke scale evaluations for the trial. They should carefully review the training tape before repeating the certification. Examiners having 6 to 9 outliers for Tape 1 or 7 - 11 outliers for Tape 2 are required to repeat the certification but can continue to do stroke scales for the trial in the interim.

NIH STROKE SCALE

Patient Identification. ___ _____ ___ _____ ___ ___

Date of Exam ___ ___/___ ___/___ ___

Time: ___ ___:___ ___ [ ]am [ ]pm

Person Administering Scale _____

Administer stroke scale items in the order listed. Record performance in each category after each subscale exam. Do not go back and change scores. Follow directions provided for each exam technique. Scores should reflect what the patient does, not what the clinician thinks the patient can do. The clinician should record answers while administering the exam and work quickly. Except where indicated, the patient should not be coached (i.e., repeated requests to patient to make a special effort).

Instructions	Scale Definition	Score
1a. Level of Consciousness: The investigator must choose a response if a full evaluation is prevented by such obstacles as an endotracheal tube, language barrier, orotracheal trauma/bandages. A 3 is scored only if the patient makes no movement (other than reflexive posturing) in response to noxious stimulation.	0 = **Alert;** keenly responsive. 1 = **Not alert**; but arousable by minor stimulation to obey, answer, or respond. 2 = **Not alert**; requires repeated stimulation to attend, or is obtunded and requires strong or painful stimulation to make movements (not stereotyped). 3 = Responds only with reflex motor or autonomic effects or totally unresponsive, flaccid, and areflexic.	_____
1b. LOC Questions: The patient is asked the month and his/her age. The answer must be correct - there is no partial credit for being close. Aphasic and stuporous patients who do not comprehend the questions will score 2. Patients unable to speak because of endotracheal intubation, orotracheal trauma, severe dysarthria from any cause, language barrier, or any other problem not secondary to aphasia are given a 1. It is important that only the initial answer be graded and that the examiner not "help" the patient with verbal or non-verbal cues.	0 = **Answers** both questions correctly. 1 = **Answers** one question correctly. 2 = **Answers** neither question correctly.	_____
1c. LOC Commands: The patient is asked to open and close the eyes and then to grip and release the non-paretic hand. Substitute another one step command if the hands cannot be used. Credit is given if an unequivocal attempt is made but not completed due to weakness. If the patient does not respond to command, the task should be demonstrated to him or her (pantomime), and the result scored (i.e., follows none, one or two commands). Patients with trauma, amputation, or other physical impediments should be given suitable one-step commands. Only the first attempt is scored.	0 = **Performs** both tasks correctly. 1 = **Performs** one task correctly. 2 = **Performs** neither task correctly.	_____

NIH STROKE SCALE

Patient Identification. ___ _____ ___ _____ ___ ___

Date of Exam ___ ___/___ ___/___ ___

2. Best Gaze: Only horizontal eye movements will be tested. Voluntary or reflexive (oculocephalic) eye movements will be scored, but caloric testing is not done. If the patient has a conjugate deviation of the eyes that can be overcome by voluntary or reflexive activity, the score will be 1. If a patient has an isolated peripheral nerve paresis (CN III, IV or VI), score a 1. Gaze is testable in all aphasic patients. Patients with ocular trauma, bandages, pre-existing blindness, or other disorder of visual acuity or fields should be tested with reflexive movements, and a choice made by the investigator. Establishing eye contact and then moving about the patient from side to side will occasionally clarify the presence of a partial gaze palsy.	0 = **Normal.** 1 = **Partial gaze palsy;** gaze is abnormal in one or both eyes, but forced deviation or total gaze paresis is not present. 2 = **Forced deviation,** or total gaze paresis not overcome by the oculocephalic maneuver.	_____
3. Visual: Visual fields (upper and lower quadrants) are tested by confrontation, using finger counting or visual threat, as appropriate. Patients may be encouraged, but if they look at the side of the moving fingers appropriately, this can be scored as normal. If there is unilateral blindness or enucleation, visual fields in the remaining eye are scored. Score 1 only if a clear-cut asymmetry, including quadrantanopia, is found. If patient is blind from any cause, score 3. Double simultaneous stimulation is performed at this point. If there is extinction, patient receives a 1, and the results are used to respond to item 11.	0 = **No visual loss.** 1 = **Partial hemianopia.** 2 = **Complete hemianopia.** 3 = **Bilateral hemianopia** (blind including cortical blindness).	_____
4. Facial Palsy: Ask – or use pantomime to encourage – the patient to show teeth or raise eyebrows and close eyes. Score symmetry of grimace in response to noxious stimuli in the poorly responsive or non-comprehending patient. If facial trauma/bandages, orotracheal tube, tape or other physical barriers obscure the face, these should be removed to the extent possible.	0 = **Normal** symmetrical movements. 1 = **Minor paralysis** (flattened nasolabial fold, asymmetry on smiling). 2 = **Partial paralysis** (total or near-total paralysis of lower face). 3 = **Complete paralysis** of one or both sides (absence of facial movement in the upper and lower face).	_____

NIH STROKE SCALE

Patient Identification. ___ _____ ___ _____ ___ ___

Date of Exam ___ ___/___ ___/___ ___

5. Motor Arm: The limb is placed in the appropriate position: extend the arms (palms down) 90 degrees (if sitting) or 45 degrees (if supine). Drift is scored if the arm falls before 10 seconds. The aphasic patient is encouraged using urgency in the voice and pantomime, but not noxious stimulation. Each limb is tested in turn, beginning with the non-paretic arm. Only in the case of amputation or joint fusion at the shoulder, the examiner should record the score as untestable (UN), and clearly write the explanation for this choice.	0 = **No drift;** limb holds 90 (or 45) degrees for full 10 seconds. 1 = **Drift;** limb holds 90 (or 45) degrees, but drifts down before full 10 seconds; does not hit bed or other support. 2 = **Some effort against gravity;** limb cannot get to or maintain (if cued) 90 (or 45) degrees, drifts down to bed, but has some effort against gravity. 3 = **No effort against gravity;** limb falls. 4 = **No movement.** UN = **Amputation** or joint fusion, explain: _____ **5a. Left Arm** **5b. Right Arm**	_____ _____
6. Motor Leg: The limb is placed in the appropriate position: hold the leg at 30 degrees (always tested supine). Drift is scored if the leg falls before 5 seconds. The aphasic patient is encouraged using urgency in the voice and pantomime, but not noxious stimulation. Each limb is tested in turn, beginning with the non-paretic leg. Only in the case of amputation or joint fusion at the hip, the examiner should record the score as untestable (UN), and clearly write the explanation for this choice.	0 = **No drift;** leg holds 30-degree position for full 5 seconds. 1 = **Drift;** leg falls by the end of the 5-second period but does not hit bed. 2 = **Some effort against gravity;** leg falls to bed by 5 seconds, but has some effort against gravity. 3 = **No effort against gravity;** leg falls to bed immediately. 4 = **No movement.** UN = **Amputation** or joint fusion, explain: _____ **6a. Left Leg** **6b. Right Leg**	_____ _____

NIH STROKE SCALE

Patient Identification. ___ _____ ___ _____ ___ ___

Date of Exam ___ ___/___ ___/___ ___

7. Limb Ataxia: This item is aimed at finding evidence of a unilateral cerebellar lesion. Test with eyes open. In case of visual defect, ensure testing is done in intact visual field. The finger-nose-finger and heel-shin tests are performed on both sides, and ataxia is scored only if present out of proportion to weakness. Ataxia is absent in the patient who cannot understand or is paralyzed. Only in the case of amputation or joint fusion, the examiner should record the score as untestable (UN), and clearly write the explanation for this choice. In case of blindness, test by having the patient touch nose from extended arm position.	0 = **Absent.** 1 = **Present in one limb.** 2 = **Present in two limbs.** UN = **Amputation** or joint fusion, explain: _____	_____
8. Sensory: Sensation or grimace to pinprick when tested, or withdrawal from noxious stimulus in the obtunded or aphasic patient. Only sensory loss attributed to stroke is scored as abnormal and the examiner should test as many body areas (arms [not hands], legs, trunk, face) as needed to accurately check for hemisensory loss. A score of 2, "severe or total sensory loss," should only be given when a severe or total loss of sensation can be clearly demonstrated. Stuporous and aphasic patients will, therefore, probably score 1 or 0. The patient with brainstem stroke who has bilateral loss of sensation is scored 2. If the patient does not respond and is quadriplegic, score 2. Patients in a coma (item 1a=3) are automatically given a 2 on this item.	0 = **Normal;** no sensory loss. 1 = **Mild-to-moderate sensory loss;** patient feels pinprick is less sharp or is dull on the affected side; or there is a loss of superficial pain with pinprick, but patient is aware of being touched. 2 = **Severe to total sensory loss;** patient is not aware of being touched in the face, arm, and leg.	_____
9. Best Language: A great deal of information about comprehension will be obtained during the preceding sections of the examination. For this scale item, the patient is asked to describe what is happening in the attached picture, to name the items on the attached naming sheet and to read from the attached list of sentences. Comprehension is judged from responses here, as well as to all of the commands in the preceding general neurological exam. If visual loss interferes with the tests, ask the patient to identify objects placed in the hand, repeat, and produce speech. The intubated patient should be asked to write. The patient in a coma (item 1a=3) will automatically score 3 on this item. The examiner must choose a score for the patient with stupor or limited cooperation, but a score of 3 should be used only if the patient is mute and follows no one-step commands.	0 = **No aphasia;** normal. 1 = **Mild-to-moderate aphasia;** some obvious loss of fluency or facility of comprehension, without significant limitation on ideas expressed or form of expression. Reduction of speech and/or comprehension, however, makes conversation about provided materials difficult or impossible. For example, in conversation about provided materials, examiner can identify picture or naming card content from patient's response. 2 = **Severe aphasia;** all communication is through fragmentary expression; great need for inference, questioning, and guessing by the listener. Range of information that can be exchanged is limited; listener carries burden of communication. Examiner cannot identify materials provided from patient response. 3 = **Mute, global aphasia;** no usable speech or auditory comprehension.	_____

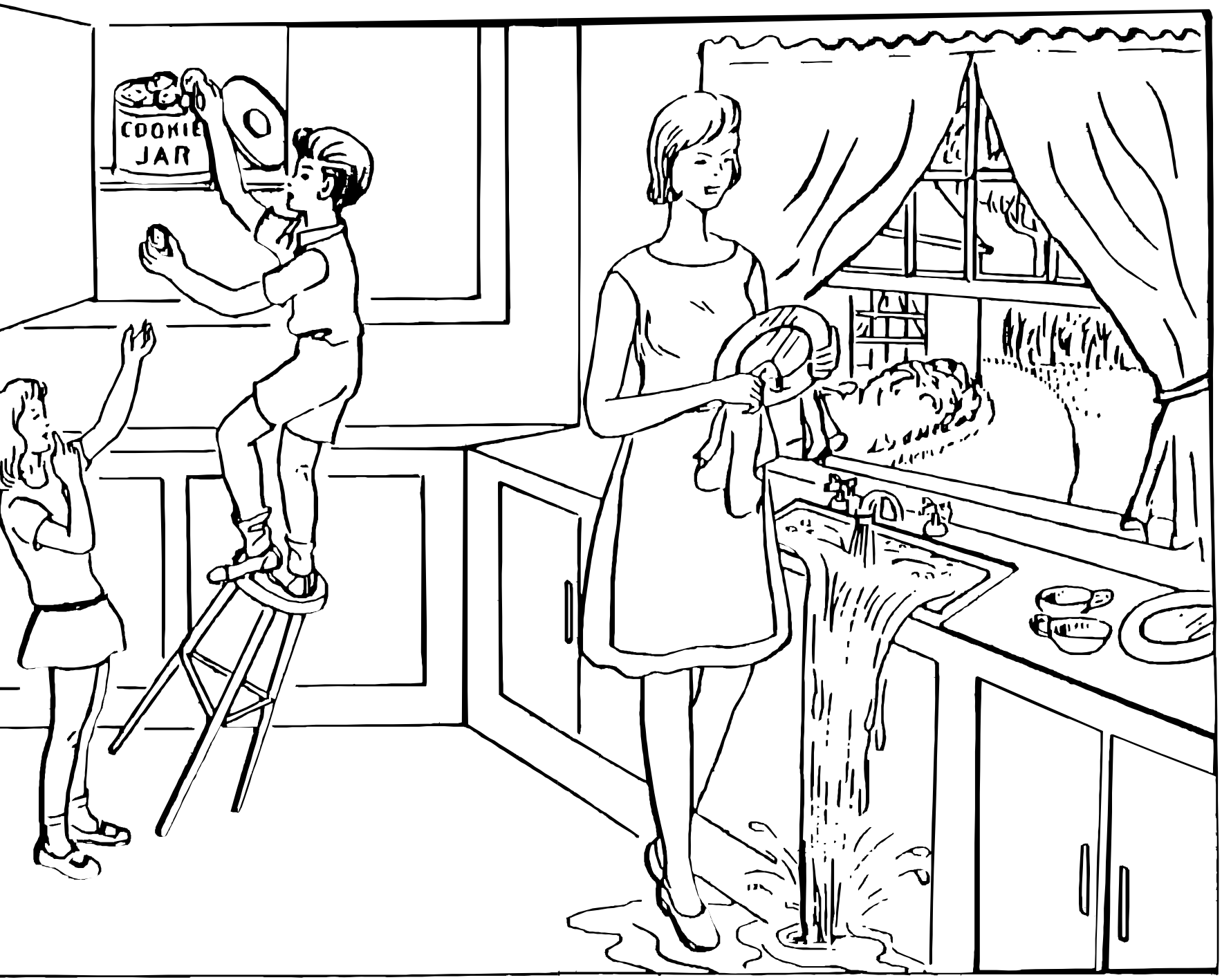
NIH STROKE SCALE

Patient Identification. ___ _____ ___ _____ ___ ___

Date of Exam ___ ___/___ ___/___ ___

10. Dysarthria: If patient is thought to be normal, an adequate sample of speech must be obtained by asking patient to read or repeat words from the attached list. If the patient has severe aphasia, the clarity of articulation of spontaneous speech can be rated. Only if the patient is intubated or has other physical barriers to producing speech, the examiner should record the score as untestable (UN), and clearly write an explanation for this choice. Do not tell the patient why he or she is being tested.	0 = **Normal.** 1 = **Mild-to-moderate dysarthria;** patient slurs at least some words and, at worst, can be understood with some difficulty. 2 = **Severe dysarthria;** patient's speech is so slurred as to be unintelligible in the absence of or out of proportion to any dysphasia, or is mute/anarthric. UN = **Intubated** or other physical barrier, explain:_____	_____
11. Extinction and Inattention (formerly Neglect): Sufficient information to identify neglect may be obtained during the prior testing. If the patient has a severe visual loss preventing visual double simultaneous stimulation, and the cutaneous stimuli are normal, the score is normal. If the patient has aphasia but does appear to attend to both sides, the score is normal. The presence of visual spatial neglect or anosagnosia may also be taken as evidence of abnormality. Since the abnormality is scored only if present, the item is never untestable.	0 = **No abnormality.** 1 = **Visual, tactile, auditory, spatial, or personal inattention** or extinction to bilateral simultaneous stimulation in one of the sensory modalities. 2 = **Profound hemi-inattention or extinction to more than one modality;** does not recognize own hand or orients to only one side of space.	_____

TOTAL _____


COOKIE
JAR

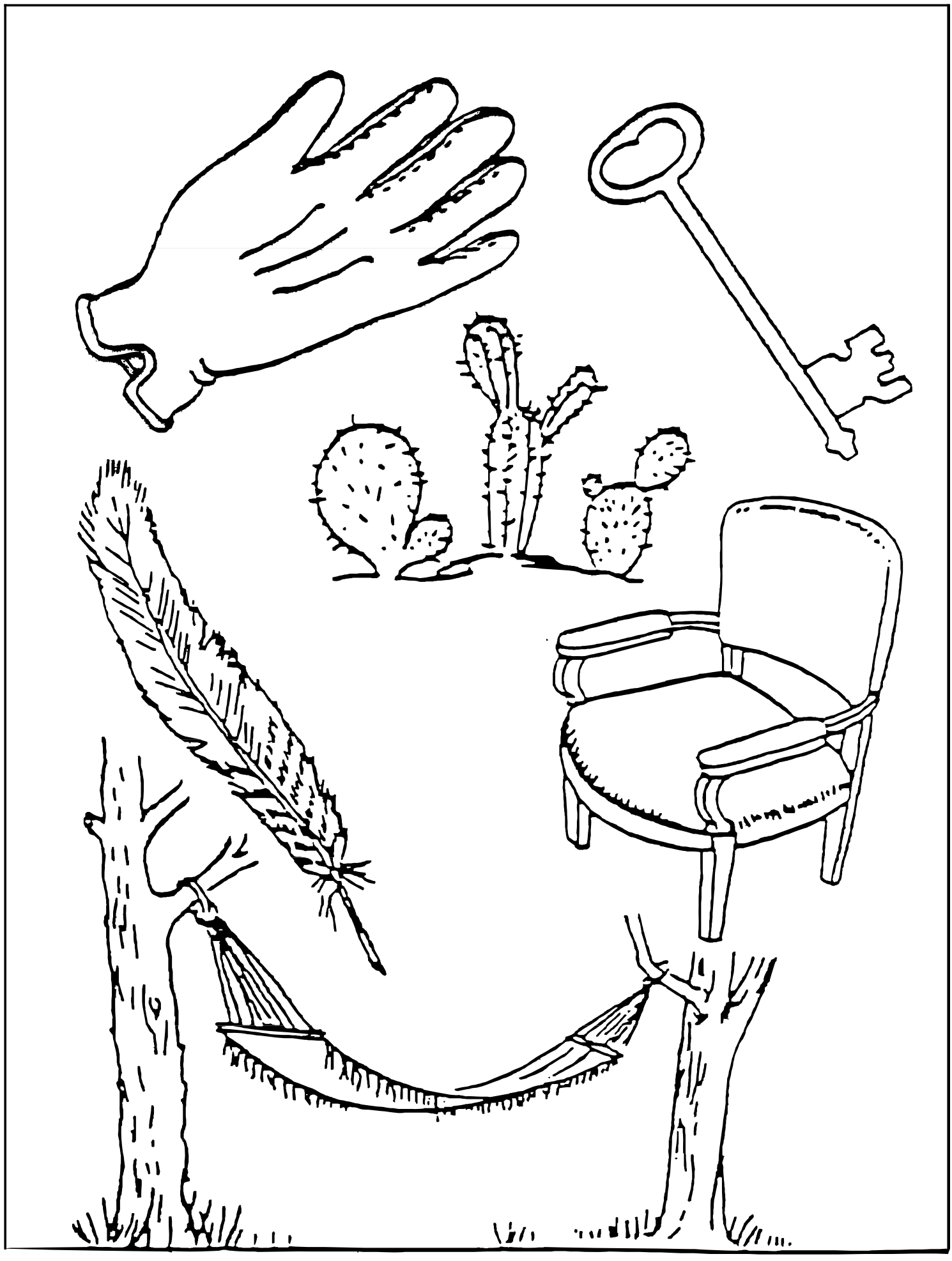
You know how.

Down to earth.

I got home from work.

Near the table in the dining room.

They heard him speak on the radio last night.



MAMA

TIP – TOP

FIFTY – FIFTY

THANKS

HUCKLEBERRY

BASEBALL PLAYER

CATERPILLAR

Filming the NIHSS

This document summarizes the process used in filming the NIHSS training and certification videos for use in the NINDS rt-PA for Acute Stroke Trial. These videos are now used widely for training and certification, and thus are of interest to those seeking to understand the strengths and limitations of NIHSS video training and certification.

The first step in validating a stroke deficit rating scale is to train users. Prior to the video era, medicine was taught at the bedside. Clinical trial investigator meetings included break-out sessions to train investigators to properly use the outcome assessment tools. Although unimaginable to today's students, creating a training video involved specialized, expensive cameras and editing equipment; engineering a "master" videotape that was used to then produce videotape reproductions; and physical-mail distribution of VHS videotapes[1]. This creative process involved collaborations with video production teams who brought with them extensive experience, including a sensibility and training that was counterproductive to our purpose. For example, the camera operator with no knowledge of "facial paresis" would light and frame the video shot in a pleasing manner that tended to obscure the subtle flattening of the nasolabial fold that was of greater interest to a neurologist. Thus, in our first attempt to produce an accurate training video, we found it essential to have one of the neurologists behind the camera, directing the videographer to tighten the frame, alter the lighting, or change focus to highlight each clinical deficit. For eye movements and facial weakness, the "close-up" required an ultra-tight shot that violated the film-school rules these professionals knew.

In a live demonstration, the student can observe both the examiner and the patient simultaneously. In a recorded video, the only way to show both the examiner and the patient is to widen the frame, but then the details of the response cannot be seen well. In consultation with our video production colleagues, we developed a 2-camera approach to the production of the video[1]. The video editor could later insert a close-up of the patient response into the frame, allowing the viewer to appreciate both the examiner's technique, and simultaneously a detailed view of the response (Figure). Prior to beginning the NINDS rt-PA for Acute Stroke Trial (the Trial), we produced a training video and 2 certification videos at Henry Ford Hospital. Actual patient volunteers were recruited from the large stroke ward managed at that time by Dr. K.M.A. Welch. Several of the NINDS investigators volunteered to perform aspects of the NIHSS in front of the camera, while others watched from behind the camera to assure accuracy. The collaborating professionals from the video production industry reluctantly responded to our repeated requests to "forget what you learned in film school". Hundreds of hours of tape—including multiple "takes" of each shot needed to illustrate each examination finding—were then edited over a 2-week period in Detroit to produce the needed master videotape. To assure the greatest neurological accuracy across all patient vignettes, one of us (PL) sat with the editor and selected each shot for each patient demonstration. The plan was to create one "training" tape and two "certification" tapes that were to be

viewed after training. The 2 different certification groups were intended to comprehensively test all possible responses to all items on the NIHSS; however, since the videotaping was completed in 2 separate 1-week sessions, and since the only available subjects were whatever patients happened to be admitted during those two weeks, it was impossible to cover all the possible responses, a significant shortcoming of the videotapes[1]. To allow the user time to write a score on the paper answer sheet, pauses were inserted by the video editor during which the screen displayed the graphic message "Record Your Score". After mastering and reproduction, the tapes were distributed to the study sites; all participants in the Trial were required to view the training tape and then one certification tape and score each certification patient. On Certification Tape 1 there were 5 patients and on Tape 2 there were 6 patients, and each patient was shown doing 15 tasks, for a total of 165 tasks followed by "Record Your Score" moments. Once the test was completed, scores were reviewed centrally for "grading", and the successful student-viewer was sent a coffee mug imprinted with the phrase "Record Your Score". Only after passing central review and approval was any investigator allowed to begin enrolling patients into the Trial.

Certification and training

Video training and certification introduces bias into the learning and testing process. The student-viewer needs to be able to see the finding accurately, apply the scoring rule correctly, and then derive the correct score. Unfortunately, even our innovative 2-camera shooting paradigm could not accurately depict every finding on every patient. Subtle findings on sensory and ataxia items were especially problematic[2, 3]. We thus could not create a certification scoring system in which we compared the student-viewer scores to the actual, known deficits in that patient. Therefore, we designed a scoring system that accounted for the artificial limitations of the video viewing process[1].

First, we asked 3 highly experienced and interested investigators from the Trial Steering committee to view and score all 11 certification patients; reproducibility was excellent. As expected, however, even expert users did not see every finding accurately, even though they agreed with each other. Therefore, we created a scoring algorithm: after all (n=162) investigators at all study sites returned their scoring sheets, the mode response was determined for each test item for each certification patient. We required that the mode be clearly identifiable and where more than one response qualified as a "mode" then both, or in some cases 3, responses were allowed as "correct". Once the mode responses were known, each response sheet was scored: any responses more than 1 response from the mode response was scored "outlier". We used the rule that any student-viewer could score one outlier per patient. Thus, no more than 5 outliers were allowed for Certification Tape 1, and no more than 6 outliers were allowed for Certification Tape 2. Users who scored less than the allowed number of outliers on Tape 1 were "certified" and allowed to begin entering patients into the trial. Users who scored more than the allowed number of outliers were asked to re-watch the training video and try again. After 6 months all users were asked to view and score Certification Tape 2 to assure continued proficiency with the scale. Exactly as for Tape 1, mode responses were tabulated and outliers identified.

One intended consequence of the Tapes was that new investigators could be added easily to the trial at any time; this design solved a chronic problem for trials run in the 1980s and 1990s: rather than delay certification of new investigators to the next investigator's training workshop, new team members at the trial study sites were asked to view the Training Tape and then the Certification Tape 1. New users were scored for outlier responses and certified as was done with the original investigators. During the course of the Trial, dozens of new investigators, including MDs and RNs, were added to the trial using these tapes, an innovation for clinical trials at the time. However, an unintended consequence of this scoring system was that there were a number of "correct" responses to many of the case scenarios, creating an impression of leniency in the scoring[4].

Use in trials

Following the publication of the NINDS rt-PA for Acute Stroke Trial in 1995, and regulatory approval of the drug in 1996, clinical trialists expressed interest in using the NIHSS for their clinical trials. Soon there were requests for the Training and Certification Videotapes. Given the culture of the 1990s, and the fact that the tapes had been produced under NIH auspices, all of these requests were granted. Both NIH and industry-sponsored studies were allowed to purchase copies of the Tapes at cost from Henry Ford Hospital—hundreds of videotapes were produced and shipped. Scoring of the user test results were all done by the statisticians at Henry Ford Hospital; the "answer sheet" has never been released or published. Published trials at the time stated in their methods sections that their investigators were trained and certified by NINDS using the Tapes.

After a few years, it became obvious that limitations of the original videotapes were unacceptable. The public communications team at NINDS (Marion Emr and Margo Warren) commissioned a re-shoot of the videos using professionals from the video production industry. The re-shoot was planned to correct many of the deficiencies in the first round of Tapes. Rather than attempt shooting in a clinic or hospital room as was done for the first set of Tapes, a professional set was built in the television studio at the University of California, San Diego, to allow more accurate lighting and careful camera positioning. Again the 2-camera arrangement was used to allow simultaneous recording of the examiner and the patient-response (Figure). The set included several features to enhance demonstration of the patient findings and overcome limitations discovered in the first production. For example, to optimize the visualization of limb drift (Items 5 and 6, Table One of the main manuscript) we placed horizontal blinds over the faux-window on the set: the drifting limb could easily be observed against the horizontal blinds in the background. Special 'fill-in' lighting was designed to optimize visualization of the eye movements and the facial asymmetry. Again, video production professionals were asked to ignore their rules of good photography craft, and instead frame the patient overly tightly or focus too close or too far away. Outpatients from the UCSD Stroke Center and inpatients at several area hospitals were selected so as to assure that every single response on every NIHSS Item was illustrated in both the Training and the Certification videos. By the time of the re-shoot, videotape technology had been replaced by digital video disks (DVD), which facilitated the certification process: since the student-viewer could select each patient item in turn, there was no need for the repetitive pause screen "Record Your Score".

The re-shoot again included many investigators from NINDS-sponsored trials who appeared on camera to illustrate the correct performance of each NIHSS Item. Video recording was completed over 2 weeks in February 2003. Again, one of us (PL) sat with the editing team to select the shots that best illustrated each finding for each item on all patients; editing was completed over several weeks in Washington D.C. A Training videodisk was created that presented each NIHSS scale Item and its scoring rules in detail using the video demonstrations. Then, the recordings of Certification patients were divided into 3 groups carefully created such that each group contained a balance of mild/severe and right/left hemisphere strokes. Over all 3 groups, each response on each Item was shown at least once. Today, Groups A, B, and C are presented sequentially, although in some implementations a group is picked at random for certification. After digital mastering, the Training Program and the Certification Patients were copied onto DVDs. The NINDS took over responsibility for distributing the DVDs to interested groups who were organizing a number of large clinical trials and UCSD provided review and grading services, either directly or via a website[5-8].

Once again, the scoring algorithm developed during the rt-PA for Acute Stroke Trial was used to certify users. Since there was no pool of trial investigators, however, it was necessary to validate the scoring rules in stages[9]. First, the scoring system was "seeded" with correct responses based on 51 responses from expert users at 3 leading stroke centers (UCSD, UT Houston, and University of Cincinnati). The first 50 student-viewers were scored using the 'outlier' method (no more than one outlier per patient) but in place of modal responses the seeded responses were used. After 50 users were certified, we tabulated the modal responses and created a new set of accepted scores by finding the mode response; as before, on some questions there were 2 or even 3 modal responses. Certifications then proceeded as before by scoring outlier responses and allowing no more than one outlier per patient in the certification group. Occasionally over the first 2 years of certification the modal responses were re-summarized, and occasional adjustments were made to the scoring algorithm. As with the Tapes, no "answer sheet" has ever been released or published: the online scoring vendors use the modal responses generated and reviewed by one investigator (PL) and further adjustments may be made from time to time to assure consistency over time.

After a year or 2, depending on regulatory requirements, users return to the DVD or a website, and re-certify by watching the next Certification Group in sequence. Once users have re-certified on all 3 groups, additional re-certifications pick one of the 3 groups to use over again. Some vendors have implemented a 'random' selection process for the 4th and subsequent certification group; others simply start over at Group A. Thus, Group "D" will be one of the original 3 certification groups. It is not possible to create new groupings of Certification patients without disrupting the careful balance of mild/severe and left/right hemisphere strokes among all 3 groups. In the future, a better implementation plan would be that each time a user certifies, a unique constellation of cases would be selected, based on that user's prior test-taking history and seeking to balance mild/moderate patients during each certification experience.

Citations

1. 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
2. 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
3. Lyden PD, Lu M, Levine SR, 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
4. Hills NK, Josephson SA, Lyden PD, Johnston SC. Is the NIHSS certification process too lenient? *Cerebrovasc Dis*. 2009;27:426-432
5. Lees KR, Zivin JA, Ashwood T, Davalos A, Davis SM, Diener HC, et al. NXY-059 for acute ischemic stroke. *N Engl J Med*. 2006;354:588-600
6. Lyden P, Shuaib A, Ng K, Levin K, Atkinson RP, Rajput A, et al. Clomethiazole Acute Stroke Study in ischemic stroke (CLASS-I): final results. *Stroke*. 2002;33:122-128
7. Johnston KC, Li JY, Lyden PD, Hanson SK, Feasby TE, Adams RJ, et al. Medical and neurological complications of ischemic stroke: experience from the RANTTAS trial. *Stroke*. 1998;29:447-453
8. Lees KR, Asplund K, Carolei A, Davis SM, Diener HC, Kaste M, et al. Glycine antagonist (gavestinel) in neuroprotection (GAIN International) in patients with acute stroke: a randomised controlled trial. GAIN International Investigators. *Lancet*. 2000;355:1949-1954
9. 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

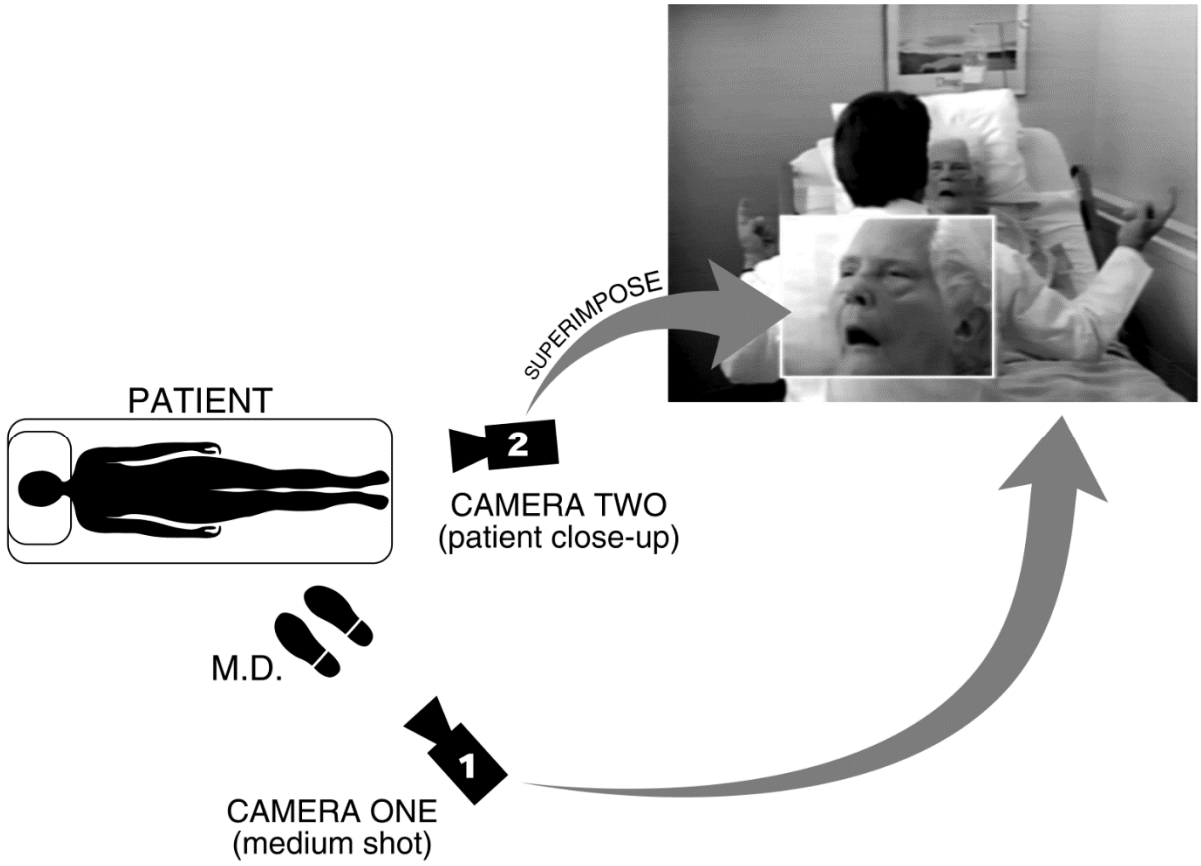


Figure. Framing the test examiner and the subject. Neurological responses can be subtle, but the examiner's stimulus may be large. To show both the larger body movements and commands of the examiner, but also show the possibly subtle responses of the patient, a 2-camera solution was devised. One camera records the examiner showing correct technique, while the second camera records the subtlety of the response, usually in extreme close-up. In post-production, the two images can be merged by the editor to make a single view that demonstrates clearly the method and the response. (Figure reprinted with permission[1]).