

The Pitfalls of NIHSS: Time for a New Clinical Acute Stroke Severity Scoring System in the Emergency?

Archita Makharia, Ayush Agarwal, Divyani Garg, Venugopalan Y. Vishnu, M. V. Padma Srivastava

Department of Neurology, All India Institute of Medical Sciences, New Delhi, India

Abstract

Stroke is the second leading cause of death and a major cause of disability worldwide. Stroke severity scales serve as reliable means to track a patient's neurological deficit, predict outcome, and guide treatment decisions in clinical practice. The National Institute of Health Stroke Scale (NIHSS) was introduced over 30 years ago, marking a significant milestone in the field of stroke. Over the years, there have been notable advancements in acute stroke care. Despite several modifications made to NIHSS, none has yet succeeded in effectively capturing all the complex effects of a stroke. This review focuses on the pitfalls of NIHSS and emphasizes the need for a quick and comprehensive clinical and upgraded version of the stroke severity rating scale.

Keywords: Acute, ischemic, mRS, NIHSS, stroke

INTRODUCTION

Stroke is the second leading cause of death and a major cause of disability worldwide. Precise prediction models of stroke severity and comprehensive assessment tools are, therefore, necessary for the evaluation of a patient with stroke. These scales serve as important tools to track a patient's neurological deficit, predict outcome, and guide treatment decisions.

During the 1980s, several stroke deficit rating scales were developed to assess the severity of strokes. The first version of National Institute of Health Stroke Scale (NIHSS) – Cincinnati/Naloxone NIHSS – was created by combining parameters from various sources including the University of Cincinnati scale, Canadian Neurological Scale, Edinburgh-2 coma scale, and Oxbury initial severity scale.^[1] As part of the pilot recombinant-tissue plasminogen activator (r-tPA) for acute stroke trial, certain components such as plantar reflex and pupillary response were omitted and an intermediate version of NIHSS was formulated.^[2] However, it was not until the NINDS r-tPA trial in 1995 that the current version of NIHSS became established as a standard stroke deficit rating scale.^[3]

It is important to note that NIHSS was designed for use in clinical trials and not as a bedside rating scale. As a result, its design assumes that the user will cooperate with extensive training before attempting certification. It was deliberately designed to sacrifice accuracy for reproducibility, giving the cardinal rule of scoring what you see, not what you think or feel. NIHSS is the de facto standard for acute stroke neurological deficit rating.^[4] It has many advantages: it is reliable, valid, reproducible, and consistent, it predicts the stroke size, the infarct volume, and ultimately the stroke outcome. In addition, it serves as a manual for thrombolysis and decisions regarding secondary prevention. Nevertheless, no single stroke scale can effectively capture all the complex effects of a stroke. Despite

the many advantages, there are few challenges and pitfalls of NIHSS that require attention.

Zero NIHSS- stroke

It is well recognized that even with an NIHSS of zero, stroke patients may exhibit signs and symptoms of impairment. A score of zero on NIHSS does not imply that a stroke has not occurred. In a comprehensive study involving 4000 patients with stroke, of which 2618 had ischemic stroke, Martin-Schild *et al.*^[5] reported that there were 20 individuals who obtained an NIHSS score of zero. The common symptoms experienced by individuals in the NIHSS zero population include headache, vertigo, and nausea. A retrospective study conducted between 2003 and 2013 on 108 NIHSS-zero individuals revealed that these patients had experienced an increased number of lacunar and infratentorial strokes. Twenty-five percent of these patients had some residual impairment even at 3 months. Out of these 108 patients, seven died within the first year.^[6] Therefore, it is crucial that even patients with an NIHSS score of zero receive intensive management and adequate follow-up to effectively monitor their neurological status and ensure optimal outcomes.

Address for correspondence: Dr. Ayush Agarwal,
Assistant Professor, Department of Neurology, AIIMS, New Delhi, India.
E-mail: ayushthetaurian@gmail.com

Submitted: 21-Sep-2023 **Revised:** 02-Nov-2023 **Accepted:** 24-Nov-2023

Published: 18-Jan-2024

This is an open access journal, and articles are distributed under the terms of the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 License, which allows others to remix, tweak, and build upon the work non-commercially, as long as appropriate credit is given and the new creations are licensed under the identical terms.

For reprints contact: WKHLRPMedknow_reprints@wolterskluwer.com

DOI: 10.4103/aian.aian_842_23

Anterior versus posterior circulation stroke

It is evident that this scoring system is disproportionately biased toward deficits caused by lesions in the anterior circulation, such as motor functions and cortical signs. In contrast, posterior circulation deficits, characterized by cranial nerve abnormalities and ataxia, receive fewer points or may even be excluded from scoring due to the coexistence of motor deficits.

The observations from the TOAST trial further highlight the disparities of the NIHSS system in the evaluation of scores assigned to those with anterior or posterior circulation strokes. The initial median NIHSS was 3 (interquartile range [IQR] 1–6) in anterior circulation stroke, while it was slightly lower at 2.5 (IQR 1–5) in posterior circulation stroke.^[7] The Oxfordshire Community Stroke Project also showed a higher median NIHSS score assigned to anterior circulation stroke than to posterior circulation stroke (NIHSS score 16 vs. 10).^[8] Another study involving 1569 patients with ischemic stroke revealed that the median NIHSS score upon admission was consistently lower by 5 points in patients with posterior circulation strokes compared to their counterparts with anterior circulation strokes. Interestingly, more than 75% of individuals diagnosed with posterior circulation stroke had a baseline NIHSS ranging from 0 to 5. The cut-off to achieve >80% sensitivity for poor outcome was >4 in the anterior circulation stroke and >2 in the posterior circulation stroke.^[9]

For better evaluation of various characteristics of a posterior circulation stroke, many investigators have introduced an expanded version of NIHSS score. The revised scoring system has incorporated additional criteria such as nystagmus, Horner's syndrome, cranial nerve palsies (specifically the glossopharyngeal and vagus nerves), imbalance in Romberg's position, truncal ataxia, and retropulsion or lateropulsion.^[10] Furthermore, a separate scoring system, POST-NIHSS, has been developed to provide more comprehensive assessment of patients having mild–moderate symptoms of stroke (NIHSS <10). POST-NIHSS was developed using random forest classification algorithm and constrained optimization in a derivation cohort of 202 patients, which was further validated in a separate prospective cohort of 65 patients. Through this process, several predictors such as age, NIHSS score, abnormal cough, dysphagia, and gait/truncal ataxia were identified as significant factors influencing the functional outcome. To quantify for these factors accurately, POST-NIHSS was computed by adding 5 points for abnormal cough, 4 points for dysphagia, and 3 points for gait/truncal ataxia to the original NIHSS score.^[11]

Right- versus left-sided stroke

Another notable limitation of NIHSS is its inherent bias toward left-sided stroke in comparison to right-sided stroke. The reasons behind this discrepancy are multifaceted.

Firstly, many studies have consistently shown that right-sided strokes often present less frequently and at a later stage to the hospital. In trials like NINDS, CLASS-1, Beth Israel Deaconess Medical Centre registry, and rt-PA registry

of the Helsinki University Central hospital, the ratio of left-:right-sided infarcts at presentation ranged from 1.08 to 1.2.^[12,13] The delayed presentation results from disturbed perception of the left hemibody (hemineglect), resulting in decreased awareness of stroke.^[14] Findings from the Asymptomatic Carotid Atherosclerosis Study (ACAS) study also revealed clinically silent ischemic lesions were more commonly observed in right-sided strokes.^[15]

Secondly, the discrepancy between points awarded by NIHSS for language function (up to 7 points) versus neglect (up to 2 points) adds to this bias. Consequently, patients with right-sided strokes are 45% less likely to be thrombolysed compared to those with a left-sided stroke.^[16]

Furthermore, patients having similar NIHSS score between 0 and 5 have been found to have larger lesions (on diffusion weighted imaging- MRI (DWI-MRI) sequences) on the right side compared to those having left-sided stroke (mean volume, 8.8 vs. 3.2 cm³, *P*-0.04). However, this disparity in the infarct volume of the right side and left side was not present in those having higher NIHSS scores.^[17]

Interrater reliability

The reliability of NIHSS is a crucial factor that determines its effectiveness in yielding consistent results, regardless of the administrator. This is measured as interrater/interobserver reliability and measured with κ coefficient. The degree of interrater agreement based on the kappa statistic may be interpreted with the following scale: <0, poor agreement; 0–0.20, slight agreement; 0.21–0.40, fair agreement; 0.41–0.60, moderate agreement; 0.61–0.80, substantial agreement; and 0.81–1.00, almost perfect agreement.^[18]

Several studies have confirmed the overall reliability of NIHSS in this context, including a study comparing neurologists and research nurses.^[19] However, the interrater reliability of different items of the score is a subject of concern, as some items exhibit inconsistent agreement among different raters. Goldstein *et al.* conducted a study which revealed that although components such as language, motor, neglect, and loss of consciousness parts had substantial interrater agreement, the same was not true for extraocular movements, dysarthria, and facial palsy.^[20] In a separate study involving 7405 raters, with 38,148 individual NIHSS item responses, the agreement on the limb ataxia item was extremely low, whereas the three items assessing gaze, aphasia, and facial weakness were low using an unweighted κ statistic. Repeated NIHSS certification also did not lead to improved agreement among raters in these domains.^[21] In another study measuring the interrater reliability of the 15 items, two showed poor agreement, 11 showed moderate agreement, and two showed excellent agreement based on κ scores.^[22] These inconsistencies in scoring could impact clinical decision-making and trial results.

Redundancy

Some items of NIHSS have been found to be redundant in certain situations. For example, testing ataxia in a patient

with gross motor dysfunction may not be feasible. Similarly, testing dysarthria in a patient with aphasia and vice versa is useless. In addition, it is difficult to make an accurate assessment of NIHSS in altered mental status. To address these redundancies, a modified version of NIHSS (mNIHSS) was developed, which removed the items related to facial palsy, dysarthria, and level of consciousness 1a. The sensory item (item 8) was simplified from three to two choices, making 31 as the total score of mNIHSS.^[23] By reducing the number of items and simplifying the grading criteria, it was intended that administering mNIHSS would become simpler and more user-friendly. In a validation of mNIHSS against NIHSS, the number of elements with excellent agreement increased from 54% to 71%, while the number of elements with poor agreement decreased from 12% to 5%. Overall, 45% of NIHSS items had less than excellent reliability versus only 29% for mNIHSS.^[23] However, with the removal of the ataxia item, there may be concern that mNIHSS would be even less able to assess brainstem strokes.

Time-consuming

The administration of NIHSS can be quite time-consuming. In a study by Shafqat *et al.*,^[24] it was established that the bedside assessment of NIHSS required a mean time of 6.55 min (range 4–12 min) and the remote assessment via telemedicine required a mean time of 9.70 min (range 6–18 min).

Given that the decisions regarding revascularization are guided by NIHSS score, there is a need for a tool that is faster and simpler; thus, several shortened versions such as Slim NIHSS and Shortened NIHSS (sNIHSS-4, sNIHSS-5, sNIHSS-8, sNIHSS EMS) have been developed. Of all the characters, eight items related to right leg, left leg, gaze, visual fields, language, level of consciousness, facial palsy, and dysarthria have been observed to be the most predictive of good outcome at 3 months after a stroke. sNIHSS-8 has been devised to include all the abovementioned eight items. A shorter sNIHSS-5 has also been devised that includes the first five items mentioned above. The accuracies of sNIHSS-8 and sNIHSS-5 as suggested by the area under curve on receiver operator characteristics (ROC) in the validation models are 0.77 and 0.76, respectively.^[25] These ROC values denote that these shortened versions are not adequate. They face validity issues and may overlook valuable information, particularly in minor strokes.

Language

The cultural appropriateness in the language part of NIHSS may vary across different settings. While it was initially developed in English for the Western population, some images such as hammocks may not be easily recognizable to individuals from other ethnic backgrounds. To address this issue, many countries and cultures have modified and validated their own versions of NIHSS that are better tailored to their linguistic and cultural contexts, for example, the Hindi version of NIHSS (HV-NIHSS) was created by substituting English sentences and words with Hindi equivalents, while also making

culturally appropriate adaptations.^[26] NIHSS has similarly been translated and validated into other languages such as Cantonese for Hong Kong, Chinese, German, Spanish, Estonian, Hungarian, Italian, Marathi, Portuguese, and Telugu. This localized approach ensures greater inclusivity and accuracy when administering the test within diverse communities.

The cookie theft picture commonly utilized for language assessment in stroke patients presents certain inherent shortcomings. Its outdated appearance features a woman clad in an apron and washing dishes. Consequently, when individuals are prompted to describe the image, they often remark on its perpetuation of stereotypical gender roles. To address this limitation, an updated version of the cookie theft picture has been developed that includes additional objects to describe, displays greater diversity, is colorful, and overcomes the stereotypical gender roles.^[27]

Ceiling effect

A phenomenon known as the ceiling effect arises when a significant proportion of participants achieve the maximum score on the scale. NIHSS score may have a ceiling effect as certain items cannot be tested in patients with very severe strokes and will be given the highest score.^[28] Consequently, when majority of the participants are clustered near the highest possible score, the measurement becomes less meaningful in its assessment.

Measures impairment, not disability

It is important to note that NIHSS measures only impairment and does not assess disability. It is crucial to understand that two individuals with the same impairment may experience different levels of disability based on factors such as their lifestyle and job requirements. For instance, a professional musician might exhibit weakness of only the small muscles of the hand and dexterity, yet score a zero on NIHSS. However, this loss of dexterity holds a greater significance for them. This highlights the need for a comprehensive evaluation of an individual's overall functioning and capabilities.

Does not replace a neurological examination

While NIHSS may be a valuable tool, it should not be seen as a replacement for thorough neurological examination. It was originally developed for use in clinical trials, and its application at the bedside has limitations.

Large vessel occlusion prediction

Randomized controlled trials and their meta-analysis have proved that endovascular thrombectomy (EVT) is superior to best medical management in patients with acute ischemic stroke having a large vessel occlusion (LVO). Early suspicion of LVO is important for immediate preparedness of thrombectomy suites or referral to thrombectomy centers (if the treating center does not have that expertise). The best cut-off value above which an LVO should be suspected on NIHSS was found to be 7 (sensitivity- 81%, positive predictive value- 84%).^[29] However, up to 30% patients with LVO have lower NIHSS scores on initial admission.^[30] Therefore, a score which has superior LVO predictability might be better in the present EVT era. Rapid Arterial occlusion Evaluation (RACE)

scale is one such scale which was designed based on NIHSS items with a higher LVO prediction value.^[31] It has five components: facial, arm and leg weakness (0–2 for each item), gaze (0–1), and aphasia or agnosia (0–2 based on the side of the infarct: left vs. right). A RACE score ≥ 5 was found to be predictive of an LVO with a sensitivity of 85%^[31,32] and is significantly less time-consuming compared to NIHSS.

CONCLUSION

NIHSS was introduced over 30 years ago, marking a significant milestone in the field of stroke. Over the years, there have been notable advancements in acute stroke care. Despite several modifications made to NIHSS, such as the expanded NIHSS, and POST-NIHSS, none has yet succeeded in accurately capturing all the intricate effects of stroke. In addition, attempts to simplify and expedite decision-making through slim and shortened versions of the scale have proven to be inefficient. Further investigation is needed to determine whether additional training, modification of examination elements, or clearer definitions could improve the scoring system. The time has now come to develop an upgraded version of stroke severity rating scale that prioritizes efficiency without compromising comprehensiveness – one that addresses clinical nuances for enhanced accuracy and precision. Till that time, despite its fallacies, NIHSS remains arguably the best scale that we have.

Financial support and sponsorship

Nil.

Conflicts of interest

There are no conflicts of interest.

REFERENCES

- Olinger CP, Adams HPJ, Brott TG, Biller J, Barsan WG, Toffol GJ, *et al.* High-dose intravenous naloxone for the treatment of acute ischemic stroke. *Stroke* 1990;21:721–5.
- Brott TG, Haley EC, 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;23:632–40.
- 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;25:2220–6.
- Lyden P. Using the National Institutes of Health Stroke Scale: A cautionary tale. *Stroke* 2017;48:513–9.
- Martin-Schild S, Albright KC, Tanksley J, Pandav V, Jones EB, Grotta JC, *et al.* Zero on the NIHSS does not equal the absence of stroke. *Ann Emerg Med* 2011;57:42–5.
- Eskioglu E, Huchmandzadeh Millotte M, Amiguet M, Michel P. National Institutes of Health Stroke Scale Zero Strokes. *Stroke* 2018;49:3057–9.
- Chung JW, Park SH, Kim N, Kim WJ, Park JH, Ko Y, *et al.* Trial of ORG 10172 in Acute Stroke Treatment (TOAST) classification and vascular territory of ischemic stroke lesions diagnosed by diffusion-weighted imaging. *J Am Heart Assoc* 2014;3:e001119.
- Yang Y, Wang A, Zhao X, Wang C, Liu L, Zheng H, *et al.* The Oxfordshire Community Stroke Project classification system predicts clinical outcomes following intravenous thrombolysis: A prospective cohort study. *Ther Clin Risk Manag* 2016 Jun 29;12:1049–56.
- Inoa V, Aron AW, Staff I, Fortunato G, Sansing LH. Lower NIH stroke scale scores are required to accurately predict a good prognosis in posterior circulation stroke. *Cerebrovasc Dis* 2014;37:251–5.
- Olivato S, Nizzoli S, Cavazzuti M, Casoni F, Nichelli PF, Zini A. e-NIHSS: An Expanded National Institutes of Health Stroke Scale Weighted for Anterior and Posterior Circulation Strokes. *J Stroke Cerebrovasc Dis* 2016;25:2953–7.
- Alemseged F, Rocco A, Arba F, Schwabova JP, Wu T, Cavicchia L, *et al.* Posterior National Institutes of Health Stroke Scale Improves Prognostic Accuracy in Posterior Circulation Stroke. *Stroke* 2022;53:1247–55.
- National Institute of Neurological Disorders and Stroke rt-PA Stroke Study Group. Tissue plasminogen activator for acute ischemic stroke. *N Engl J Med* 1995;333:1581–7.
- Wahlgren NG, Ranasinha KW, Rosolacci T, Franke CL, van Erven PM, Ashwood T, *et al.* Clomethiazole acute stroke study (CLASS): Results of a randomized, controlled trial of clomethiazole versus placebo in 1360 acute stroke patients. *Stroke* 1999;30:21–8.
- Mesulam MM. A cortical network for directed attention and unilateral neglect. *Ann Neurol* 1981;10:309–25.
- Enderectomy for asymptomatic carotid artery stenosis. Executive Committee for the Asymptomatic Carotid Atherosclerosis Study. *JAMA* 1995;273:1421–8.
- Di Legge S, Fang J, Saposnik G, Hachinski V. The impact of lesion side on acute stroke treatment. *Neurology* 2005;65:81–6.
- 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;33:954–8.
- Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977;33:159–74.
- 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;28:307–10.
- Goldstein LB, Bertels C, Davis JN. Interrater reliability of the NIH stroke scale. *Arch Neurol* 1989;46:660–2.
- Josephson SA, Hills NK, Johnston SC. NIH Stroke Scale reliability in ratings from a large sample of clinicians. *Cerebrovasc Dis* 2006;22:389–95.
- 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;36:2446–9.
- Meyer BC, Lyden PD. The modified National Institutes of Health Stroke Scale: Its time has come. *Int J Stroke* 2009;4:267–73.
- Shafqat S, Kvedar JC, Guanci MM, Chang Y, Schwamm LH. Role for telemedicine in acute stroke. Feasibility and reliability of remote administration of the NIH stroke scale. *Stroke* 1999;30:2141–5.
- Tirschwell DL, Longstreth WT, Becker KJ, Gammans RE, Sabounjian LA, Hamilton S, *et al.* Shortening the NIH Stroke scale for use in the prehospital setting. *Stroke* 2002;33:2801–6.
- Prasad K, Dash D, Kumar A. Validation of the Hindi version of National Institute of Health Stroke Scale. *Neuro India* 2012;60:40–4.
- Berube S, Nonnemacher J, Densky C, Glenn S, Saxena S, Wright A, *et al.* Stealing cookies in the twenty-first century: Measures of spoken narrative in healthy versus speakers with aphasia. *Am J Speech Lang Pathol* 2019;28 (1 Suppl):321–9.
- Muir KW, Weir CJ, Murray GD, Povey C, Lees KR. Comparison of neurological scales and scoring systems for acute stroke prognosis. *Stroke* 1996;27:1817–20.
- Heldner MR, Hsieh K, Broeg-Morvay A, Mordasini P, Buhlmann M, Jung S, *et al.* Clinical prediction of large vessel occlusion in anterior circulation stroke: Mission impossible? *J Neurol* 2016;263:1633–40.
- Volbers B, Groger R, Engelhorn T, Marsch A, Macha K, Schwab S, *et al.* Acute stroke with large vessel occlusion and minor clinical deficits: Prognostic factors and therapeutic implications. *Front Neurol* 2021;12:736795.
- de la Ossa NP, Carrera D, Gorchs M, Querol M, Millan M, Gomis M, *et al.* Design and validation of a prehospital stroke scale to predict large arterial occlusion. The rapid arterial occlusion evaluation scale. *Stroke* 2014;45:87–91.
- Carrera D, Gorchs M, Querol M, Abilleira S, Ribo M, Millan M, *et al.* Revalidation of the RACE scale after its regional implementation in Catalonia: A triage tool for large vessel occlusion. *J Neurointerv Surg* 2019;11:751–6.