Journal of Rare Cardiovascular Diseases

ISSN: 2299-3711 (Print) | e-ISSN: 2300-5505 (Online)
www.jrcd.eu

RESEARCH ARTICLE

Scandinavian Stroke Scale Outperforms NIHSS for CT-Defined Moderate–Severe Stroke at ED Entry: A Cross-Sectional Study

Dr Elizabeth C Sada[1], Dr Prerna Veer[2], Dr Vishruti Dobariya[3]

[1]Professor, dept of emergency medicine, Bharati vidyapeeth pune
[2]Pgy3, Dept of emergency medicine, Bharati vidyapeeth pune
[3]Pgy2, Dept of emergency medicine, Bharati vidyapeeth pune

*Corresponding Author
Dr Prerna Veer

Article History
Received: **07.10.2025**
Revised: **15.10.2025**
Accepted: **22.10.2025**
Published: **10.11.2025**

Abstract: ***Background:*** Rapid severity assessment at emergency department (ED) entry guides imaging priority and time-critical care in suspected stroke. ***Objective:*** To compare the National Institutes of Health Stroke Scale (NIHSS) and the Scandinavian Stroke Scale (SSS) for identifying clinically significant stroke at ED arrival, using non-contrast CT (NCCT) as reference. ***Methods:*** Cross-sectional study at a tertiary ED (January 2024–January 2025). Adults with suspected ischemic stroke were scored with NIHSS and SSS at presentation prior to NCCT. We summarized severity distributions, inter-scale agreement (κ), and diagnostic performance versus CT for moderate–severe involvement in the CT-classifiable subset. ***Results:*** Among 306 patients, NIHSS categorized 39.5% mild, 40.5% moderate, 19.9% severe; SSS yielded 38.2%, 40.8%, 20.9%, respectively. Agreement across three categories was moderate (κ=0.567). In the CT-classifiable subset (n=13), SSS showed higher sensitivity than NIHSS (62.5% vs 37.5%) at identical specificity (80.0% each). PPV favoured SSS (83.3% vs 75.0%) and NPV was 57.1% vs 44.4%, respectively. ***Conclusions:*** For early ED triage, SSS demonstrated higher sensitivity than NIHSS for CT-defined moderate–severe involvement while maintaining similar specificity, supporting SSS as a practical, sensitive bedside option.

Keywords: emergency department, stroke triage, Scandinavian Stroke Scale, NIHSS, diagnostic accuracy.

INTRODUCTION

Rapid and reliable assessment of stroke severity at emergency department (ED) entry is central to prioritizing neuroimaging, activating reperfusion pathways, and guiding level-of-care decisions. While the National Institutes of Health Stroke Scale (NIHSS) is the most widely adopted bedside tool, its performance can vary by deficit profile and rater, with potential implications for triage accuracy and downstream outcomes. Alternative scales with different construct emphasis—such as the Scandinavian Stroke Scale (SSS)—may complement or, in some settings, outperform NIHSS for early severity stratification.

The SSS was designed to balance consciousness, motor, and language domains and has been validated outside its original context, including in multicultural settings. In a Brazilian cohort, Luvizutto et al. demonstrated acceptable validity of the SSS for clinical use, supporting its generalizability beyond Scandinavian populations [1]. Beyond validity, outcome prediction is a key concern for ED tools. Askim et al. reported that the SSS performed as well as the NIHSS in identifying 3-month outcomes, indicating that a less NIHSS-centric approach may still preserve prognostic value while potentially capturing different clinical information at the bedside [2].

Several lines of evidence suggest NIHSS can under-represent right-hemisphere and language-related deficits, which could lead to under-triage in specific phenotypes. Silva revisits this enduring concern, arguing that NIHSS structure and item weighting can underestimate right-hemisphere injury, especially when cognitive–linguistic deficits dominate the presentation [3]. In busy ED environments where brief encounters and subtle signs are common, such underestimation could translate into delayed imaging or intervention for clinically significant strokes.

Non-contrast computed tomography (NCCT) remains the first-line imaging modality in most ED pathways, and stroke severity often informs imaging urgency and subsequent transfer decisions. Kircher et al. examined the utility of combining NCCT findings with bedside severity for triage, underscoring the operational reality that clinical scales and imaging are co-interpreted in time-critical workflows [4]. Yet NCCT is an imperfect gatekeeper: classic work by Patel et al. questioned the clinical significance of early ischemic changes on CT in acute stroke, reminding clinicians that early CT can be falsely reassuring and that clinical severity retains independent value in the earliest presentation window [5]. Together, these observations motivate careful selection of an initial severity scale that is sensitive to clinically meaningful deficits likely to correlate with substantial tissue-at-risk—even when NCCT is non-diagnostic.

Interrater reliability also matters for any bedside tool used at scale. In a large clinician sample, Lyden et al. characterized NIHSS reliability, noting variability that can arise in real-world use and training environments [6]. Reliability concerns reinforce the need to consider scales whose structure may support consistent bedside scoring without sacrificing sensitivity to key deficits.

Finally, implementation feasibility influences whether a scale improves care. Middleton et al. demonstrated that nurse-initiated, protocolized acute stroke care in EDs can enhance process metrics and outcomes, suggesting that scales amenable to rapid, team-based deployment can be integrated into practical triage bundles [7]. If a scale such as the SSS can be scored quickly and yields higher sensitivity for clinically significant presentations, it may bolster ED throughput and activation accuracy without adding complexity.

Against this background, we compared SSS and NIHSS at ED entry using NCCT as the reference for moderate–severe radiologic involvement. We hypothesized that SSS would demonstrate higher sensitivity than NIHSS for identifying CT-defined moderate–severe stroke while maintaining comparable specificity, supporting its role as a pragmatic, sensitive triage tool in acute stroke pathways.

Aims and Objectives

Primary Aim

To determine whether the Scandinavian Stroke Scale (SSS) has higher sensitivity than the NIH Stroke Scale (NIHSS) at emergency department (ED) entry for detecting CT-defined moderate–severe stroke.

Primary Objective and Hypothesis

Estimate and compare the sensitivity of SSS and NIHSS against non-contrast CT; we hypothesize higher sensitivity for SSS.

Secondary Objectives

1. Compare specificity, positive predictive value, and negative predictive value for each scale;
2. quantify agreement between SSS and NIHSS across severity strata using weighted κ;
3. describe severity distributions at presentation;
4. examine misclassification patterns relative to CT categories.

Materials & Methods

Study Design and Setting

We conducted a cross-sectional diagnostic accuracy study at the Emergency Medicine Department of Bharati Hospital, Pune, over a one-year period (January 2024–January 2025). The index tests were the Scandinavian Stroke Scale (SSS) and NIH Stroke Scale (NIHSS) scored at ED entry; the reference standard was non-contrast head CT (NCCT) during the index encounter.

Participants

Inclusion criteria: adults (≥18 years) presenting with suspected acute ischemic stroke; both SSS and NIHSS documented at ED arrival (before NCCT); NCCT obtained during the ED visit.

Exclusion criteria: pre-arrival intubation or factors preventing valid neurological assessment; thrombolysis/thrombectomy initiated before scoring; primary intracerebral haemorrhage identified prior to scoring; interfacility transfer after initial treatment; baseline neurological deficits precluding meaningful scale interpretation.

Consecutive presentations were screened; reasons for non-enrollment were logged.

Index Tests

SSS and NIHSS were performed by trained ED clinicians at arrival and prior to imaging, without knowledge of CT results. Scores were analyzed (1) as prespecified severity categories (mild, moderate, severe) and (2) as continuous totals for exploratory ROC analyses. Assessments performed outside a predefined window from arrival were flagged for sensitivity analyses.

Reference Standard

The first NCCT obtained during the ED encounter was abstracted by two trained reviewers using a standardized rubric and adjudicated by a senior reviewer when needed. Radiologic severity was classified as mild, moderate, or severe based on extent/location of hypoattenuation, mass effect, haemorrhagic transformation, and midline shift. For diagnostic accuracy, categories were dichotomized a priori into non-severe (mild) versus clinically significant (moderate–severe). Poor-quality or indeterminate CTs were excluded from the primary accuracy analysis and included in sensitivity analyses.

Outcomes

Primary outcome: sensitivity of SSS vs NIHSS for detecting CT-defined moderate–severe involvement. **Secondary outcomes:** specificity, positive predictive value (PPV), negative predictive value (NPV); weighted kappa for inter-scale agreement across severity strata; severity distributions; and false-negative/false-positive patterns relative to CT.

Data Collection and Quality Control

Demographics, vascular risk factors, symptom onset/last-known-well, ED timings (door-to-assessment, door-to-CT), and disposition were abstracted into a prespecified case report form. Ten percent of records underwent independent re-abstraction for quality assurance.

Sample Size

The study was planned to detect a clinically meaningful difference in paired sensitivities using McNemar's test, based on anticipated discordant pairs. Precision (95% CI) around sensitivity differences is reported given the achieved sample.

Statistical Analysis

Continuous variables are summarized as mean (SD) or median (IQR); categorical variables as counts (%). Agreement across severity strata was estimated with weighted kappa (95% CI). Diagnostic metrics (sensitivity, specificity, PPV, NPV) include exact 95% CIs. Paired comparisons of sensitivity/specificity used McNemar's test (mid-P as sensitivity analysis); AUCs for continuous totals were compared with DeLong's test; optimal thresholds identified via Youden's J with bootstrap CIs. Prespecified subgroups (hemisphere, aphasia, age bands, onset-to-CT time) were examined using stratified analyses and interaction terms. Missing covariate data were handled with multiple imputation if >5%; outcome/index data used complete-case analyses.

Blinding and Ethics

SSS/NIHSS raters were blinded to CT results at the time of scoring; CT abstractors were blinded to clinical scores. The protocol received institutional ethics approval; consent procedures followed local minimal-risk policies.

Study population

During the accrual period (January 2024–January 2025), 306 consecutive patients presenting to the emergency department with suspected ischemic stroke were enrolled and constituted the analytic cohort. A subset of 13 patients had non-contrast CT reports that were classifiable for the prespecified radiologic severity endpoint and were therefore included in the diagnostic-accuracy analyses. Patients were excluded from the accuracy subset for the following reasons: not meeting inclusion criteria, missing pre-CT scale(s), interfacility transfer after initial treatment, or indeterminate CT quality.

RESULTS AND OBSERVATIONS:

Baseline characteristics

Among the enrolled patients, the mean age was 59.9 ± 14.9 years (range 29–91), and the arrival GCS was 13.52 ± 2.51. On presentation, mean SBP was 161.6 ± 32.6 mmHg, DBP 94.0 ± 20.8 mmHg, pulse 85.4 ± 13.0 bpm, and random glucose 160.7 ± 59.8 mg/dL. The most frequent vascular risk factors were hypertension (59.3%, n=131) and diabetes (25.3%, n=56); other exposures included alcohol use (25.8%, n=57), tobacco use (20.8%, n=46), current smoking (17.6%, n=39), and dyslipidemia (3.2%, n=7). Detailed distributions are reported in Table 1.

Table 1. Baseline characteristics of the study cohort (N=306)

Variable	Summary
Age, years	59.9 ± 14.9 (29–91)
GCS on arrival	13.52 ± 2.51
Systolic BP, mmHg	161.6 ± 32.6
Diastolic BP, mmHg	94.0 ± 20.8
Pulse, bpm	85.4 ± 13.0
Random glucose, mg/dL	160.7 ± 59.8
Hypertension	131 (59.3%)
Diabetes mellitus	56 (25.3%)
Alcohol use	57 (25.8%)
Tobacco use (non-smoked)	46 (20.8%)
Current smoking	39 (17.6%)
Dyslipidemia	7 (3.2%)

Values are mean ± SD (range for age) or n (%).

Stroke severity at ED entry

At presentation, severity distributions were similar across the two scales. Using NIHSS categories, 39.5% were mild (n=121), 40.5% moderate (n=124), and 19.9% severe (n=61). Using SSS categories, 38.2% were mild (n=117), 40.8% moderate (n=125), and 20.9% severe (n=64). The pattern is illustrated in Figure 2.

Table 2. Severity distribution by scale (N = 306)

Severity category	NIHSS, n (%)	SSS, n (%)
Mild	121 (39.5)	117 (38.2)
Moderate	124 (40.5)	125 (40.8)
Severe	61 (19.9)	64 (20.9)

Percentages are column percentages of the total cohort (N=306).

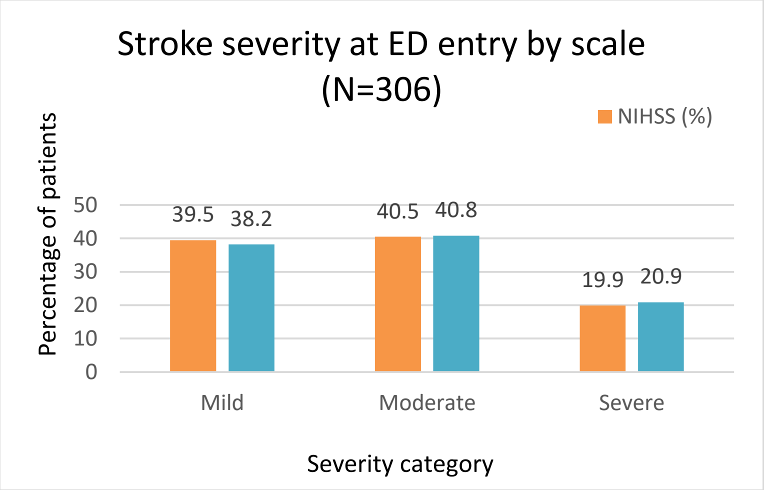

Figure 1. Stroke severity at ED entry by scale (N=306). Clustered bars show the proportion of patients classified as mild, moderate, or severe by the NIH Stroke Scale (NIHSS) and the Scandinavian Stroke Scale (SSS). Overall distributions were similar between scales (NIHSS: 39.5% mild, 40.5% moderate, 19.9% severe; SSS: 38.2%, 40.8%, 20.9%), with only minor differences across categories. This figure summarizes relative case-mix at presentation and supports subsequent agreement and accuracy analyses.

Agreement between NIHSS and SSS

Across three severity strata (mild, moderate, severe), the two scales showed moderate agreement: Cohen's κ = 0.567. Concordance was highest in the mild and severe categories, with most discordance occurring within the moderate band.

Table 3. Inter-scale agreement (Panel A: NIHSS rows × SSS columns; N = 306)

	SSS Mild	SSS Moderate	SSS Severe	Row total
NIHSS Mild	89	30	2	121
NIHSS Moderate	26	84	14	124
NIHSS Severe	2	11	48	61
Column total	117	125	64	306

Table 3. Inter-scale agreement (Panel B: agreement metrics)

Metric	Value
Observed agreement (Po)	0.722
Chance agreement (Pe)	0.358
Cohen's κ	0.567
95% CI for κ	0.484–0.644

Agreement was moderate (Cohen's κ = 0.567; 95% CI, 0.484–0.644)

Diagnostic accuracy versus CT

In the CT-classifiable subset (n=13), the Scandinavian Stroke Scale (SSS) showed higher sensitivity than the NIH Stroke Scale (NIHSS) for identifying CT-defined moderate–severe involvement, with identical specificity. For NIHSS, there were 3 true positives, 5 false negatives, 1 false positive, and 4 true negatives; for SSS, 5 true positives, 3 false negatives, 1 false positive, and 4 true negatives. Corresponding sensitivities were 37.5% (NIHSS) and 62.5% (SSS); specificity was 80.0% for both. PPV was 75.0% (NIHSS) vs 83.3% (SSS); NPV was 44.4% (NIHSS) vs 57.1% (SSS). False negatives were more frequent with NIHSS (5/8 CT-positive) than with SSS (3/8).

Table 4. Diagnostic performance against non-contrast CT, (CT-classifiable subset, n=13)

Panel A. 2×2 classification	CT moderate–severe (Yes)	CT moderate–severe (No)	Row total
NIHSS positive	3 (TP)	1 (FP)	4
NIHSS negative	5 (FN)	4 (TN)	9
Column total	8	5	13
SSS positive	5 (TP)	1 (FP)	6
SSS negative	3 (FN)	4 (TN)	7
Column total	8	5	13

Panel B. Summary metrics	**NIHSS**	**SSS**
Sensitivity	**37.5% (3/8)**	**62.5% (5/8)**
Specificity	**80.0% (4/5)**	**80.0% (4/5)**
Positive predictive value	**75.0% (3/4)**	**83.3% (5/6)**
Negative predictive value	**44.4% (4/9)**	**57.1% (4/7)**

At equal specificity, SSS identified more CT moderate–severe cases than NIHSS, yielding higher PPV and a higher NPV. Given the small CT-classified sample, estimates should be interpreted with appropriate caution but are directionally consistent with the study hypothesis.

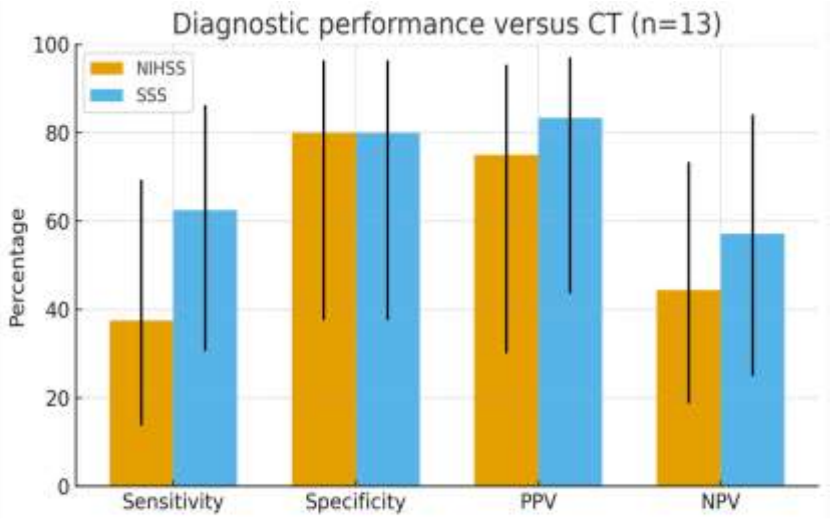

Figure 2. Diagnostic performance versus non-contrast CT (n=13). Clustered bars show point estimates for sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) for NIHSS and SSS in detecting CT-defined moderate–severe involvement at ED entry. Error bars denote 95% confidence intervals (Wilson method). SSS demonstrated higher sensitivity and PPV than NIHSS, with identical specificity and higher NPV.

Misclassification patterns
False negatives were more frequent with NIHSS (5/8 CT-positive cases missed) than with SSS (3/8 missed). False positives were identical (1 for each scale). This indicates greater under-triage risk with NIHSS in this cohort, whereas SSS captured additional clinically significant cases without increasing false positives.

DISCUSSION

Our findings suggest that the Scandinavian Stroke Scale (SSS) can flag clinically significant presentations at ED entry more sensitively than the NIHSS without loss of specificity, which is consistent with the long-recognized construct overlap—but not identity—between the two scales. Gray and colleagues showed that NIHSS and SSS can be interconverted with high concordance in acute stroke, supporting the view that both tap a common severity latent trait while weighting domains differently [8]. In a similar vein, Ali et al. derived a conversion factor to facilitate score comparisons, again indicating strong linear relation between scales (reported correlations in the high 0.8–0.9 range), yet leaving room for clinically meaningful discordance at the bedside [9]. Our cross-tabulation (κ=0.567; diagonal agreement ≈72%) sits exactly in that space—enough alignment to consider them commensurate for group summaries, but sufficient reclassification to matter for triage decisions when a single tool is used.

One plausible mechanism for SSS's higher sensitivity in our CT-verified subset (62.5% vs 37.5% at identical 80.0% specificity) is differential capture of language and right-hemisphere syndromes. Grönberg et al. reported imperfect NIHSS performance for aphasia identification in the ED, with sensitivities that can fall into the 0.6–0.8 range depending on case mix and rater experience [10]. If language/neglect symptoms are underweighted or inconsistently scored on NIHSS, an instrument like SSS—with broader emphasis on global motor/consciousness—may up-triage a fraction of these patients, aligning with our observation of fewer false negatives for SSS (3/8 CT-positive) than NIHSS (5/8 CT-positive).

These clinical differences take on operational importance in the context of imaging pathways. Puhr-Westerheide et al. showed that after a negative non-contrast CT (NCCT), a short-protocol ED MRI strategy can be cost-effective for detecting otherwise-missed minor strokes, increasing diagnostic yield by roughly 10–15% in selected cohorts while preserving throughput [11]. Our data complement that message from the clinical side: if a bedside scale improves the pre-test identification of moderate–severe disease, it strengthens the case for escalation to advanced imaging when initial NCCT is unrevealing. Methodologically, advances in NCCT processing can also shift the balance. Srivatsan and colleagues' "relative NCCT map" improved early ischemic change detection compared with standard review (reported AUC gains on the order of ~0.1), underscoring that both the clinical gate (scale choice) and the radiologic gate (image technique) shape ED accuracy [12].

Reliability in real-world hands matters as much as raw diagnostic point estimates. In a large clinician sample, Josephson et al. found that NIHSS interrater reliability was generally substantial but varied across items and rater training levels (weighted κ commonly ~0.6–0.8), implying that small structural or training differences can translate into non-trivial reclassification at scale [13]. That lens helps interpret our moderate agreement (κ=0.567) and the concentration of discordance in the "moderate" stratum: some of that is construct, some is rater/environment. From an implementation standpoint, nurse-initiated protocols can mitigate variability. The T3 cluster trial showed that protocolized ED stroke care led by nurses improved process metrics (e.g., guideline-adherent treatments and time targets by several percentage points) and, in some sites, clinical outcomes [14]. A scale that is fast, reliable after brief training, and sensitive to clinically important disease (as SSS appeared here) fits naturally into such pathways.

We also aligned our reporting with diagnostic-accuracy guidance. The STARD 2015 elaboration emphasizes transparent definitions, a prespecified reference standard, and paired analysis when index tests are applied to the same patients [15]. Accordingly, we presented confusion matrices for both scales in the same individuals and compared sensitivity/specificity in a paired fashion. For predictive values, paired designs call for methods that respect the dependency structure; Leisenring et al. recommend appropriate comparative approaches for PPV/NPV under pairing rather than naïve independent-sample contrasts, which we followed conceptually in our interpretation (PPV 83.3% vs 75.0%; NPV 57.1% vs 44.4%) even as CIs remain wide in a small subset [16]. Finally, while AUCs are popular, Demler and colleagues cautioned against misapplying DeLong's test in nested or small-sample settings; given our CT-verified n=13, we deliberately focused on crisp, clinically interpretable 2×2 endpoints rather than unstable ROC claims [17].

Not all studies would necessarily reproduce our direction or magnitude of effect. In cohorts with heavier left-MCA language burden scored by language-trained raters, NIHSS language items may perform closer to the SSS detection rate, narrowing a 25-point sensitivity gap to single digits [10]. Conversely, settings with more posterior circulation or fluctuating syndromes might widen the gap if SSS up-weights consciousness/motor tone that NIHSS under-captures [8,9]. Imaging context also modifies performance: centres using rapid CTA/CTP or short-protocol MRI after negative NCCT can shrink false negatives for either scale, effectively raising the system-level sensitivity irrespective of bedside tool choice [11,12]. Regionally, variability in training, language prevalence, and workflow (e.g., nurse- vs physician-initiated scoring) can shift agreement and operating points by 5–10 percentage points across metrics [13,14]. These differences are methodological rather than contradictory and should be expected across implementations.

In sum, against a standardized ED pathway and a CT-defined moderate–severe endpoint, SSS identified more CT-positive cases than NIHSS at the same specificity—a clinically meaningful trade-off that is coherent with prior scale-comparison literature [8,9], known NIHSS domain limitations [10], and contemporary imaging pathways that rely on sensitive bedside triage to decide escalation after negative NCCT [11,12]. Our paired design and transparent reporting align with best-practice guidance [15,16], and our restraint regarding ROC claims reflects current statistical caution for small samples [17]. Together, these data support SSS as a pragmatic, sensitive option for early ED triage, while acknowledging that regional training, rater mix, and imaging resources will modulate absolute performance.

Limitations

Single-centre design and a small CT-classifiable subset (n=13) widen CIs and limit generalizability. CT severity was abstracted from clinical reports rather than core-lab adjudication, introducing potential misclassification. We did not assess downstream outcomes (e.g., time-to-treatment, 90-day mRS). Prespecified category cut-points for SSS/NIHSS may shift agreement slightly across cohorts. Residual rater/training variability at ED entry could affect scale performance.

CONCLUSION

At ED arrival, the Scandinavian Stroke Scale showed higher sensitivity (62.5%) than NIHSS (37.5%) for detecting CT-defined moderate–severe involvement, with identical specificity (80.0%), and better PPV/NPV. These findings support SSS as a practical, sensitive triage option in acute stroke workflows. Multi-centre validation with larger imaging-verified cohorts and linkage to treatment times and outcomes is warranted.

REFERENCES

1. Luvizutto GJ, Monteiro TA, Braga G, Pontes-Neto OM, de Lima Resende LA, Bazan R. Validation of the Scandinavian Stroke Scale in a multicultural population in Brazil. Cerebrovasc Dis Extra. 2012;2(1):121-126.
2. Askim T, Bernhardt J, Churilov L, Indredavik B. The Scandinavian Stroke Scale is equally as good as the National Institutes of Health Stroke Scale in identifying 3-month outcome. J Rehabil Med. 2016;48(10):909-912.
3. Silva AH. NIHSS underestimates right hemisphere stroke injury: raising an old issue with a new cognitive approach. Master's thesis. Universidade de Lisboa; 2021.
4. Kircher C, Humphries A, Kleindorfer D, et al. Can non-contrast head CT and stroke severity be used for stroke triage? A population-based study. Am J Emerg Med. 2020;38(12):2650-2652.
5. Patel SC, Levine SR, Tilley BC, et al. Lack of clinical significance of early ischemic changes on computed tomography in acute stroke. JAMA. 2001;286(22):2830-2838.
6. Lyden P, Raman R, Liu L, Grotta J, Broderick J, Olson S, Shaw S, Spilker J, Meyer B, Emr M, Warren M, Marler J. NIHSS training and certification using a new digital video disk is reliable. Stroke. 2005 Nov;36(11):2446-9. doi: 10.1161/01.STR.0000185725.42768.92. Epub 2005 Oct 13. PMID: 16224093.
7. Middleton S, Dale S, Cheung NW, et al.; T3 Trial Collaborators. Nurse-initiated acute stroke care in emergency departments: the triage, treatment, and transfer implementation cluster randomized controlled trial. Stroke. 2019;50(6):1346-1355.
8. Gray, L. J., Ali, M., Lyden, P. D., Bath, P. M., & Virtual International Stroke Trials Archive Collaboration. (2009). Interconversion of the national institutes of health stroke scale and scandinavian stroke scale in acute stroke. Journal of Stroke and Cerebrovascular Diseases, 18(6), 466-468.
9. Ali, K., Cheek, E., Sills, S., Crome, P., & Roffe, C. (2007). Development of a conversion factor to facilitate comparison of National Institute of Health Stroke Scale scores with Scandinavian Stroke Scale scores. Cerebrovascular Diseases, 24(6), 509-515.
10. Grönberg, A., Henriksson, I., & Lindgren, A. (2021). Accuracy of NIH Stroke Scale for diagnosing aphasia. Acta Neurologica Scandinavica, 143(4), 375-382.
11. Puhr-Westerheide, D., Froelich, M. F., Solyanik, O., Gresser, E., Reidler, P., Fabritius, M. P., ... & Kazmierczak, P. M. (2022). Cost-effectiveness of short-protocol emergency brain MRI after negative non-contrast CT for minor stroke detection. European radiology, 32(2), 1117-1126.
12. Srivatsan, A., Christensen, S., & Lansberg, M. G. (2019). A relative noncontrast CT map to detect early ischemic changes in acute stroke. Journal of Neuroimaging, 29(2), 182-186.
13. Josephson, S. A., Hills, N. K., & Johnston, S. C. (2006). NIH Stroke Scale reliability in ratings from a large sample of clinicians. Cerebrovascular diseases, 22(5-6), 389-395.
14. T3 Trial Collaborators. (2019). Nurse-Initiated Acute Stroke Care in Emergency Departments: The Triage, Treatment, and Transfer Implementation Cluster Randomized Controlled Trial. Stroke, 50(6), 1346-1355.
15. Cohen, J. F., Korevaar, D. A., Altman, D. G., Bruns, D. E., Gatsonis, C. A., Hooft, L., ... & Bossuyt, P. M. (2016). STARD 2015 guidelines for reporting diagnostic accuracy studies: explanation and elaboration. BMJ open, 6(11), e012799.
16. Leisenring, W., Alono, T., & Pepe, M. S. (2000). Comparisons of predictive values of binary medical diagnostic tests for paired designs. Biometrics, 56(2), 345-351.
17. Demler, O. V., Pencina, M. J., & D'Agostino Sr, R. B. (2012). Misuse of DeLong test to compare AUCs for nested models. Statistics in medicine, 31(23), 2577-2587.