Comparison of 7 Prehospital Stroke Detection Scales

Anne W. Alexandrov, PhD, AGACNP-BC, ANVP-BC, ASC-BC, FAAN

Numerous clinical scales exist to support stroke clinicians' abilities to detect actual stroke patients, especially in settings where imaging technology is not readily available, such as the prehospital care environment. Understanding the overall accuracy of an instrument requires testing of instrument sensitivity, specificity, positive predictive value (PPV) and negative predictive value (NPV) which collectively enable our understanding of how well an instrument does what it is intended to do.

Dekker and colleagues examined the performance of 7 popular prehospital stroke scales used to identify stroke patients among those presenting with stroke-like symptoms.[1] Their work provides important information for emergency service leaders and prehospital providers to consider in selection and use of a stroke scale. To understand the authors' findings, this article provides a review of sensitivity, specificity, PPV, and NPV.

Instruments such as those used to identify stroke patients in the field do not always perform perfectly. All stroke clinicians have encountered patients that a commonly used instrument such as the Face-Arm-Speech-Time (FAST) scale has identified as a stroke, when in reality the patient actually has a stroke mimic diagnosis. Examining sensitivity, specificity, PPV, and NPV across several instruments helps clinicians to understand the overall accuracy of an instrument such as FAST in relation to other prehospital stroke scales.

Imagine using an instrument that produces a "false alarm" diagnosing stroke when a patient hasn't actually had a stroke. Similarly, imagine using an instrument that produces a finding that suggests a patient isn't having a stroke when the patient actually is having a stroke. Sensitivity examines how well the instrument can detect the disease (in this case stroke) among patients that truly had a stroke.[2] High sensitivity is desirable when it is important to accurately identify cases so that a treatment can be provided, such as thrombolysis or thrombectomy. Whenever there would be severe consequences for missing a diagnosis, high sensitivity instruments are desirable, so that no cases that should be treated are missed. However, there are tradeoffs for high sensitivity, and this is false positive cases that can be identified that truly do not have the disease.

When calculating sensitivity, the true positive cases are divided by the sum of true positives and false negatives,[2] or:

$$\text{Sensitivity} = \frac{\text{true strokes}}{(\text{true strokes} + \text{patients incorrectly diagnosed as not having a stroke})}$$

DOI: 10.59236/sc.v2i2.85
ISSN: 2995-7494

On the other hand, specificity shows how well an instrument can detect someone that does NOT have a disease in patients that truly do not have the disease.[2] The tradeoff with high specificity testing is that it increases the risk of identifying false negative cases, for example patients the instrument suggests are not having a stroke that really ARE having a stroke. In the case of emergency diagnoses like stroke, instruments with high specificity may lead to undertreatment of patients with thrombolytics that really should have been treated.

Specificity is the percent of "true negative" cases out of all cases that do not have the condition of interest. When calculating specificity, the true negative cases are divided by the sum of true negatives and false positive cases,[2] or:

Specificity =

$$\frac{\text{true negative strokes}}{\text{(true negative strokes + patients incorrectly diagnosed as having a stroke)}}$$

When an instrument has a sensitivity of 0.80 or 80%, it identifies 80 patients out of 100 as having a disease, but 20% may be missed that actually have the diagnosis (false negative cases). Whereas, when a instrument has a specificity of 0.80 or 80%, it identifies 80 patients out of 100 that do not have the disease, but 20% may be told they have the disease when they actually do not (false positive cases).

It is important to remember that highly sensitive tests are excellent for screening, whereas highly specific tests are good for making definitive diagnoses. For example, high sensitivity prehospital stroke scales could point to stroke so that clinicians don't miss the opportunity to treat a patient with reperfusion therapies, but once at the hospital, more definitive high specificity testing is performed (e.g. imaging) to make the final diagnosis before a thrombolytic agent is given. Sensitivity and specificity should always be examined together to fully understand instruments' strengths and limitations.

PPV refers to the probability that a patient has the disease when the test is positive. PPV is calculated by dividing the true positives by all positive tests (true positives + false positives).[2]

$$\frac{\text{true strokes}}{\text{(true strokes + false positive strokes)}}$$

NPV is the probability that a patient doesn't have the disease when the test is negative. NPV is calculated by dividing the true negative cases by all the negative cases (true negatives + false negatives).[2]

$$\frac{\text{true negative stroke cases}}{\text{(true negative strokes + false negative strokes)}}$$

Similar to sensitivity and specificity, both PPV and NPV are reported as percentages. PPV is directly related to prevalence, or the proportion of a population that has a condition at a given time, whereas NPV is inversely related to prevalence.

In this study, the authors assembled data acquired in the prehospital environment from two large observational cohort studies, namely the Leiden Prehospital Stroke Study (LPSS) and the PREhospital triage of patients with suspected STrOke (PRESTO) study; these data were used as the validation cohort against which different prehospital scales were tested.[1] The cohort consisted of 3317 patients, and all cases in the database were ultimately confirmed at the hospital as being

DOI: 10.59236/sc.v2i2.85
ISSN: 2995-7494

either true stroke cases (n=2240) or stroke mimic diagnoses (n=1077). The authors also had reperfusion treatment data for the patients in the validation cohort which enabled them to explore whether prehospital stroke scales would have accurately or inaccurately identified cases that should have been treated with reperfusion therapy.[1]

An exhaustive review of the literature was performed to identify stroke scales for head-to-head evaluation, resulting in a total of 14 prehospital stroke scales. The data elements to complete each scale were then compared against the available data in the validation cohort, resulting in elimination of 7 stroke scales for which data were unavailable. The remaining 7 scales were examined and compared in their ability to correctly identify stroke patients, including: 1) The Ontario Prehospital Stroke Screen (OPSS); 2) the Cincinnati Prehospital Stroke Scale (CPSS); 3) the shortened NIHSS for Emergency Medical Services (sNIHSS-EMS) scale; 4) the FAST scale; 5) the Medic Prehospital Assessment for Code Stroke scale (MedPACS); 6) the Melbourne Ambulance Stroke Screen (MASS); and 7) the Los Angeles Prehospital Stroke Scale (LAPSS). Since the CPSS and FAST scale utilize the same items for stroke assessment with the only difference being repetition of a sentence in CPSS, data from these two scales were ultimately combined for analysis.[1]

Findings for accuracy (95% confidence interval [CI]) in stroke detection compared to the definitive diagnosis were as follows:

- MedPACS 0.66 (0.65-0.68)
- OPSS 0.66 (0.64-0.68)
- sNIHSS-EMS 0.66 (0.64-0.68)
- MASS 0.65 (0.63-0.67)
- CPSS/FAST 0.65 (0.63-0.66)
- LAPSS 0.60 (0.59-0.62)

LAPSS accuracy was significantly lower than all other scales ($p < 0.01$), and the MedPACS, OPSS, and sNIHSS-EMS were significantly more accurate at diagnosing stroke than MASS and CPSS/FAST ($p = 0.02$).

Scale sensitivity (95% CI) for stroke was as follows:

- MedPACS 84% (82-85)
- OPSS 83% (81-85)
- sNIHSS-EMS 84% (83-86)
- MASS 82% (80-84)
- CPSS/FAST 80% (78-81)
- LAPSS 66% (64-68)

Specificity (95% CI) findings for each scale were as follows:

- MedPACS 30% (27-33)
- OPSS 31% (28-34)
- sNIHSS-EMS 28% (25-31)
- MASS 30% (27-33)
- CPSS/FAST 34% (31-37)
- LAPSS 49% (46-52)

Findings for PPV (95% CI) for each scale were as follows:

- MedPACS 71% (70-73)
- OPSS 71% (70-73)
- sNIHSS-EMS 71% (69-73)
- MASS 71% (69-73)
- CPSS/FAST 72% (70-73)
- LAPSS 73% (71-75)

Lastly, findings for NPV (95% CI) for each scale were as follows:

- MedPACS 47% (43-50)
- OPSS 47% (43-50)
- sNIHSS-EMS 46% (42-50)
- MASS 44% (41-48)
- CPSS/FAST 44% (41-48)
- LAPSS 41% (38-44)

The proportion of missed reperfusion treatments that each scale would have resulted in was 10.3% for MedPACS, 11.2% for OPSS, 10.5% for sNIHSS-EMS, 13.3% for MASS, 14.8% CPSS/FAST, and 25.5% for LAPSS.

Collectively, these findings provide stroke clinicians with a strong recommendation for use of the MedPACS prehospital stroke scale. Not only did MedPACS have the best stroke diagnostic accuracy, but it was the most sensitive and had the best NPV of all the scales tested. MedPACS also resulted in the fewest missed reperfusion treatments. Conversely, the LAPSS performed the worst for diagnostic accuracy; it was the least sensitive to stroke, and it also had the highest specificity. These findings ultimately caused it to have the highest percentage of missed reperfusion cases of all the instruments assessed.

The authors are quick to point out that results from the 3 top performing scales (MedPACS, OPSS, and sNIHSS-EMS) were very similar, and that the relatively complex scoring methods used by the sNIHSS-EMS challenge its use in practice. Ultimately, the authors correctly point out that local emergency and prehospital leaders should carefully evaluate their resources and training mechanisms in relation to the top performing scales to determine the best instrument to support local practice.[1]

Author Affiliations

Dr. Alexandrov is a Professor of Nursing & Neurology at the University of Tennessee Health Science Center, Memphis in the Colleges of Nursing and the College of Medicine's Department of Neurology, and the Program Director for the NET SMART Advanced Practice Fellowship (www.learnstroke.com). She serves as the Editor-in-Chief of *Stroke Clinician*.

References

1. Dekker L, Moudrous W, Daems JD, Buist EF, Venema E, Durieux MD, van Zwet EW, de Schryver EL, Kloos LM, de Laat KF, Aerden LA, Dippel DW, Kerkhoff H, van den Wijngaard IR, Wermer MJ, Roozenbeek B, Kruyt ND. Prehospital stroke detection scales: A head-to-head comparison of 7 scales in patients with suspected stroke. Int J Stroke. 2025 Mar;20(3):268-277. doi: 10.1177/17474930241275123. Epub 2024 Sep 10. PMID: 39127910; PMCID: PMC11874475.
2. Shreffler J, Huecker MR. Diagnostic Testing Accuracy: Sensitivity, Specificity, Predictive Values and Likelihood Ratios. 2023 Mar 6. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2025 Jan–. PMID: 32491423.

DOI: 10.59236/sc.v2i2.85
ISSN: 2995-7494