

## PEER REVIEW HISTORY

BMJ Open publishes all reviews undertaken for accepted manuscripts. Reviewers are asked to complete a checklist review form (<http://bmjopen.bmj.com/site/about/resources/checklist.pdf>) and are provided with free text boxes to elaborate on their assessment. These free text comments are reproduced below.

### ARTICLE DETAILS

<b>TITLE (PROVISIONAL)</b>	Design and validation of a clinical scale for prehospital stroke recognition, severity grading, and prediction of large vessel occlusion – the shortened NIH stroke scale for emergency medical services
<b>AUTHORS</b>	Purrucker, Jan; Härtig, Florian; Richter, Hardy; Engelbrecht, Andreas; Hartmann, Johannes; Auer, Jonas; Hametner, Christian; Popp, Erik; Ringleb, Peter; Nagel, Simon; Poli, Sven

### VERSION 1 - REVIEW

<b>REVIEWER</b>	Mehul D. Patel, Research Assistant Professor of Emergency Medicine University of North Carolina at Chapel Hill, USA
<b>REVIEW RETURNED</b>	07-Apr-2017

<b>GENERAL COMMENTS</b>	<p>The authors developed a shortened version of the NIHSS for use by EMS personnel to screen for stroke, assess severity, and predict LVO in the prehospital setting. Accurate early identification and appropriate triage and transport are critical needs for systems of acute stroke care. The authors contribute yet another prehospital tool, and given previous studies and their data, the incremental value of the sNIHSS-EMS added to previously developed prehospital stroke severity scales is minimal. I feel a much more balanced presentation and interpretation of results is needed (see below).</p> <ol style="list-style-type: none"> <li>1. In the Abstract, the authors suggest their scale is the first fully NIHSS-compatible for prehospital stroke assessment. However, Tirschwell et al's sNIHSS is very similar and also fully compatible with the NIHSS. I think this statement is misleading and should be revised to appropriately reflect prior work in the field. Generally, the authors should thoroughly think through where they state their scale is "the first" anything.</li> <li>2. In the 2nd bullet under Strengths and limitations of the study, the authors claims theirs is "the first scale permitting..." but technically existing scales can also be used for these parallel reasons. I think it'd be more appropriate to say other scales have not been assessed yet for all aspects.</li> <li>3. In Introduction, I question the authors' claim that "all of these scales only focus on single aspects...". Please check this. I recall RACE or CPSSS/C-STAT and perhaps others looked at stroke severity in addition to LVO.</li> <li>4. In Introduction, the sentence, "However, this is time and resource consuming" is more opinion than based on evidence, so it should be</li> </ol>
-------------------------	--

	<p>in the discussion rather than the introduction. Since a only a small proportion of EMS responses and transports are for stroke, I would think a 7 item scale would be more time and resource intensive than say a much shorter screen to rule out non-strokes first. It warrants further discussion and investigation.</p> <p>5. According to Table 1, it appears the sNIHSS-EMS should range from 0-21 (by adding up the max scores for each item) rather than 0-29 as stated. Please check and confirm/revise.</p> <p>6. In Discussion, the authors describe the sNIHSS-EMS as "the first comprehensive stroke scale to provide parallel stroke..." although the other scales could be used in this way. Perhaps it'd be more appropriate to say it's the first scale assessed for all aspects in parallel.</p> <p>7. Further discussion is needed on the lack of grip strength to compare LAMS to the sNIHSS-EMS. It's also important to note that the LAMS is developed from the LAPSS and could be easily and quickly implemented in the prehospital setting to evaluate all aspects of stroke assessment.</p> <p>8. A limitation that only professionals from three European countries were used to develop the scale, and differences in EMS delivery in other countries may limit the generalizability of these findings. Remove or revise the sentence on page 13, "Participation of different EMS systems...around the world."</p>
--	---

<b>REVIEWER</b>	Miguel Soares-Oliveira Cambridge University Hospital NHS Trust UK.
<b>REVIEW RETURNED</b>	19-Apr-2017

<b>GENERAL COMMENTS</b>	<p>Very interesting study and might be very useful. As stressed by the authors, it needs prospective validation. Furthermore, the low specificity of the test for stroke recognition might overload local or regional high-specialized centres. This needs further prospective evaluation. As probably will need the low sensitivity to predict LOV. How many false negatives the test will have in the (real) prehospital population? How much time is it going to add to the treatment time?</p> <p>As a conclusion, it is a very good paper/research that might improve the treatment of patients with acute stroke. Further prospective evaluation to validate those above raised concerns is mandatory.</p>
-------------------------	--

<b>REVIEWER</b>	Meurer, William University of Michigan, USA
<b>REVIEW RETURNED</b>	09-May-2017

<b>GENERAL COMMENTS</b>	<p>This manuscript uses a 2-stage process to evaluate a new method to use NIHSS elements in a prehospital screening for LVO.</p> <p>While I see that your panel rejected some of the items, could you please add the shortened versions from the Tirschwell paper to your analysis? SNIHSS-8 and SNIHSS-5? <a href="https://www.ncbi.nlm.nih.gov/m/pubmed/12468773/">https://www.ncbi.nlm.nih.gov/m/pubmed/12468773/</a></p>
-------------------------	--

	Can you also comment on how well the LVO prediction occurs if comatose patients are excluded? I apologize if you did this and I missed this. It would be useful to note how many patients with LVO actually presented with overt coma in your cohort.
--	---

## VERSION 1 – AUTHOR RESPONSE

Reviewer: 1

The authors developed a shortened version of the NIHSS for use by EMS personnel to screen for stroke, assess severity, and predict LVO in the prehospital setting. Accurate early identification and appropriate triage and transport are critical needs for systems of acute stroke care. The authors contribute yet another prehospital tool, and given previous studies and their data, the incremental value of the sNIHSS-EMS added to previously developed prehospital stroke severity scales is minimal. I feel a much more balanced presentation and interpretation of results is needed (see below).

1. In the Abstract, the authors suggest their scale is the first fully NIHSS-compatible for prehospital stroke assessment. However, Tirschwell et al's sNIHSS is very similar and also fully compatible with the NIHSS. I think this statement is misleading and should be revised to appropriately reflect prior work in the field. Generally, the authors should thoroughly think through where they state their scale is "the first" anything.

Response: We agree with the Reviewer that the scales derived by Tirschwell et al. follow the same concept of using an NIHSS-compatible items assessment and we therefore modified the objectives in the abstract as follows.

"To develop a NIH Stroke Scale (NIHSS)-compatible, all-in-one scale for rapid and comprehensive prehospital stroke assessment including stroke recognition, severity grading and progression monitoring as well as prediction of large vessel occlusion (LVO)."

2. In the 2nd bullet under Strengths and limitations of the study, the authors claims theirs is "the first scale permitting..." but technically existing scales can also be used for these parallel reasons. I think it'd be more appropriate to say other scales have not been assessed yet for all aspects.

Response: We agree, that some existing severity scales might be used for parallel LVO prediction as well as stroke recognition. However, vice versa, recognition scales (such as the FAST) cannot be used for stroke severity assessment and/or LVO prediction. To better reflect this, we modified the manuscript to:

"Strengths and limitations of this study (2nd bullet point):

- The new clinical scale (sNIHSS-EMS), developed and validated in this study, is the first scale assessed for parallel stroke recognition, severity grading, and LVO prediction.

Discussion, first sentence:

"The sNIHSS-EMS is the first comprehensive stroke scale assessed for parallel stroke recognition, severity grading, and LVO prediction."

3. In Introduction, I question the authors' claim that "all of these scales only focus on single aspects...". Please check this. I recall RACE or CPSSS/C-STAT and perhaps others looked at stroke severity in addition to LVO.

Response: We agree that some LVO prediction scales offer a score that correlates with the NIHSS, the in-hospital gold-standard, and thus can be used to describe stroke severity. An exception is e.g. the VAN-score (Teleb MS et al., J Neurointerv Surg; 2017;9:122–126.) which offers only a dichotomous response (VAN + or VAN -).

We modified the Introduction, which now reads:

However, most of these scales only focus on single aspects of acute stroke care, i.e. either stroke recognition,<sup>1</sup> 2 early prediction of outcome,<sup>3</sup> prediction of thrombolysis,<sup>4</sup> 5 or severity grading and large vessel occlusion (LVO).<sup>3</sup>, 6-18

4. In Introduction, the sentence, "However, this is time and resource consuming" is more opinion than based on evidence, so it should be in the discussion rather than the introduction. Since a only a small proportion of EMS responses and transports are for stroke, I would think a 7 item scale would be more time and resource intensive than say a much shorter screen to rule out non-strokes first. It warrants further discussion and investigation.

Response: We agree with the Reviewer and removed the following sentences from the introduction:

removed: "However, this is time and resource consuming. Additionally, communication with receiving hospitals might be complicated by the use of multiple scales."

5. According to Table 1, it appears the sNIHSS-EMS should range from 0-21 (by adding up the max scores for each item) rather than 0-29 as stated. Please check and confirm/revise.

Response: We acknowledge the reviewers comment. To avoid any misunderstanding, we modified Table 1 and split the (R+L) single row for motor items into two separate rows for left and right side. The maximum score is 29.

6. In Discussion, the authors describe the sNIHSS-EMS as "the first comprehensive stroke scale to provide parallel stroke..." although the other scales could be used in this way. Perhaps it'd be more appropriate to say it's the first scale assessed for all aspects in parallel.

Response: We agree and thank the reviewer for his suggestion. Accordingly, we modified the first sentence of the discussion, which now reads:

"The sNIHSS-EMS is the first comprehensive stroke scale assessed for parallel stroke recognition, severity grading, and LVO prediction."

7. Further discussion is needed on the lack of grip strength to compare LAMS to the sNIHSS-EMS. It's also important to note that the LAMS is developed from the LAPSS and could be easily and

quickly implemented in the prehospital setting to evaluate all aspects of stroke assessment.

Response: To acknowledge the limitation of the lack of the 'grip strength' item and thus impossibility of external validation of the LAMS, we added the following sentence to the limitations in the discussion (page 12):

"We were not able to assess LVO prediction of the LAMS because the item 'grip strength' is not part of the NIHSS and thus, was not routinely documented in our cohort. According to a retrospective validation study in anterior circulation stroke, the sensitivity of the LAMS for LVO prediction was reported as 81% (at a threshold of 4).<sup>16</sup>"

8. A limitation that only professionals from three European countries were used to develop the scale, and differences in EMS delivery in other countries may limit the generalizability of these findings. Remove or revise the sentence on page 13, "Participation of different EMS systems...around the world."

Response: We agree with the Reviewer, and moved the modified sentence to the limitations in the discussion:

"Despite involvement of EMS systems from three European countries, generalizability to further EMS systems around the world cannot be concluded."

Reviewer: 2

Very interesting study and might be very useful.

Response: We appreciate the reviewer's very positive overall assessment.

As stressed by the authors, it needs prospective validation. Furthermore, the low specificity of the test for stroke recognition might overload local or regional high-specialized centres. This needs further prospective evaluation. As probably will need the low sensitivity to predict LOV.

Response: We agree.

How many false negatives the test will have in the (real) prehospital population?

Response: In our discussion, we stressed the need for future prospective validation in the prehospital target population. As an estimate, according to the test characteristics calculated in our validation cohort - which are similar to those reported for existing scales – the false-negative rate would be 29.7% (1-sensitivity).

How much time is it going to add to the treatment time?

Response: We thank the reviewer for this question. Time for assessment depends on (1) personal skills and (2) compliance of the patient. Due to the lack of prospective evaluation of the sNIHSS-EMS, we are not able to provide the answer (mean time) for now. To account for this limitation, which was also pointed out by reviewer 1, we removed any assumption on time from the manuscript.

As a conclusion, it is a very good paper/research that might improve the treatment of patients with acute stroke. Further prospective evaluation to validate those above raised concerns is mandatory.

Response: We fully agree.

Reviewer: 3

This manuscript uses a 2-stage process to evaluate a new method to use NIHSS elements in a prehospital screening for LVO.

While I see that your panel rejected some of the items, could you please add the shortened versions from the Tirschwell paper to your analysis? SNIHSS-8 and SNIHSS-5? <https://www.ncbi.nlm.nih.gov/pubmed/12468773/>

Response: Following the reviewer's suggestion, we added the sNIHSS-8 and -5 to our external validation of LVO prediction. It turned out, that the sNIHSS-EMS is non-inferior to the sNIHSS-5. However, the more complex sNIHSS-8 was statistically superior to the sNIHSS-EMS in terms of LVO prediction (AUC[sNIHSS-EMS] 0.808 vs. AUC[sNIHSS-8] 0.823;  $p=0.02$ ). Tirschwell et al. designed the scales with a focus on prediction of stroke outcome rather than stroke recognition and LVO prediction. Other than the sNIHSS-EMS, the sNIHSS-8 and -5 include items that were rejected by EMS personnel in our survey due to the complexity of correct assessment (such as 'gaze', requiring knowledge about performance of a oculocephalic maneuver). Furthermore, the sNIHSS-8 and -5 lack the item 'motor arm', which provide important additional information about the functional status, and in rare cases, isolated dysfunction of an arm is the only acute stroke symptom (e.g. in infarctions of the cortical "hand knob" area).

In order to share the additional data with the readers, we modified Figure 1 by adding the test characteristics of the sNIHSS-8 and -5. We furthermore modified the corresponding part in the discussion (page 9), which now reads

"Test characteristics regarding identification of patients with large vessel occlusion are non-inferior to existing LVO prediction scales."

And on page 11, the following sentence was added:

"The sNIHSS-8, which had a higher AUC in the ROC analysis than the sNIHSS-EMS, was developed for stroke outcome prediction and includes complex items rejected by stroke physicians and EMS personnel in our survey."

Can you also comment on how well the LVO prediction occurs if comatose patients are excluded? I apologize if you did this and I missed this. It would be useful to note how many patients with LVO actually presented with overt coma in your cohort.

Response: We thank the reviewer for his question. In our LVO prediction cohort, there were 5 patients with coma (NIHSS LOC = 3). Excluding these patients from the analysis, the optimal cut-off according to the Youden's index did not change (6). The sensitivity was 70.0%, Specificity 81.1%. Thus, only marginally differences in test characteristics were found.

We added these new data to the results, page 9:

"Exclusion of patients with coma (n=5) did not change the optimal cut-off and test characteristics

(sensitivity 70.0% [64.4–75.3], specificity 81.1% [77.1–84.6]).”

Additional manuscript improvements:

All references were re-checked for updates and updated accordingly.

## VERSION 2 – REVIEW

<b>REVIEWER</b>	Will Meurer University of Michigan, USA
<b>REVIEW RETURNED</b>	12-Jun-2017

<b>GENERAL COMMENTS</b>	<p>Thank you for your resubmission.</p> <p>Could you please add the Tirschwell shortened NIHSS-5 and NIHSS-8 to Table 3? My main reason for requesting this is that some agencies and localities may be interested in using those versions for various reasons. It would be nice to have all the operating characteristics.</p> <p>Also, I do not disagree that important stroke syndromes could involve isolated arm weakness, but I would respectfully suggest that LVOs infrequently have isolated arm weakness.</p> <p>Finally, I don't disagree that EMS won't want to do oculoccephalic maneuvers. While technically part of the NIHSS, I find that most of the time, patients with large strokes have fixed gaze deviation anyway. I understand that is not how he shortened it but I can't imagine that with the correlation with other items it by itself was a very contributory aspect of the scale.</p>
-------------------------	---

## VERSION 2 – AUTHOR RESPONSE

Response to Reviewer

First, we would like to thank the reviewer for his additional effort to improve our manuscript. Below, we respond to his comments.

Reviewer: #3

Thank you for your resubmission.

Could you please add the Tirschwell shortened NIHSS-5 and NIHSS-8 to Table 3? My main reason for requesting this is that some agencies and localities may be interested in using those versions for various reasons. It would be nice to have all the operating characteristics.

Response: We added the scale and test characteristics of both the sNIHSS-8 and -5 to table 3, including the external validation data analyzed in our cohort.