Modified National Institutes of Health Stroke Scale Can Be Estimated From Medical Records

Scott E. Kasner, MD; Brett L. Cucchiara, MD; Michael L. McGarvey, MD; Jean M. Luciano, CRNP; David S. Liebeskind, MD; Julio A. Chalela, MD

Background and Purpose—The 15-item National Institutes of Health Stroke Scale (NIHSS) is a quantitative measure of stroke-related neurological deficit with established reliability and validity for use in clinical research. An abridged 11-item modified NIHSS (mNIHSS) has been described that simplifies or eliminates redundant and less reliable items. We aimed to determine whether the mNIHSS could be accurately abstracted from medical records to facilitate retrospective research.

Methods—We selected 39 patient records for which NIHSS scores were formally measured. Handwritten notes from medical records were abstracted, and NIHSS item scores were estimated by 5 raters blinded to actual scores. Estimated scores were compared among raters and with actual measured scores.

Results—Interrater reliability for total NIHSS on admission and discharge was excellent, with intraclass correlation coefficients (ICCs) of 0.85 and 0.79, respectively. However, ICCs for 2 items (facial palsy and dysarthria) were poor (<0.40). Interrater reliability for total mNIHSS was slightly greater, with ICCs of 0.87 and 0.89 on admission and discharge, respectively. None of the 11 mNIHSS items had poor reliability, 4 were moderate (ICC, 0.40 to 0.75), and 7 were excellent (ICC >0.75). Sixty-two percent of estimated total NIHSS scores were within 2 points of actual scores and 91% were within 5 points, whereas 70% of estimated total mNIHSS scores were within 2 points and 95% were within 5 points.

Conclusions—The mNIHSS can be estimated from medical records with a high degree of reliability and validity. In retrospective assessment of stroke severity, the mNIHSS performs better than the standard NIHSS and may be easier to use because it has fewer and simpler items. (Stroke. 2003;34:568-570.)

Key Words: outcome ■ retrospective studies ■ stroke assessment

Quantitative measures of stroke-related deficit are necessary for clinical research to account for baseline severity and to analyze outcomes. Retrospective studies often precede prospective cohort or randomized clinical trials to provide preliminary data and generate hypotheses. However, retrospective stroke studies that rely on chart review must transform qualitative clinical narratives of the neurological examination into quantitative measures for formal statistical analyses.

The National Institutes of Health Stroke Scale (NIHSS) is a quantitative measure of stroke-related neurological deficit that was developed for prospective clinical research. It includes key aspects of the neurological examination: level of consciousness, speech and language function, neglect, visual fields, eye movements, facial symmetry, motor strength, sensation, and coordination. The scale has proven intrarater and interrater reliability and has predictive validity for stroke outcome. It has gained widespread use as a measure of initial and final neurological deficit in acute stroke trials. Subsequent research has demonstrated that the NIHSS can also be estimated from medical records with a high degree of reliability and validity, although accuracy may be better in academic and teaching hospitals than in community hospitals.

A modified version of the NIHSS (mNIHSS) was recently developed that is simpler and easier to use than the standard NIHSS. Redundant items and those with lower reliability were deleted: level of consciousness (item 1a), facial palsy (item 4), limb ataxia (item 7), and dysarthria (item 10). The sensory item (item 8) was collapsed from 3 to 2 choices to improve its reliability. The mNIHSS has been shown to be clinometrically similar to the standard NIHSS with excellent reliability and validity in prospective studies and appeared to offer increased power for statistical analysis.

We hypothesized that the mNIHSS score could be abstracted from medical records with a degree of reliability and validity that was comparable to or better than retrospective estimation of the standard NIHSS.

Subjects and Methods
This study was performed at an academic university hospital after review and approval by our institutional review board.
Patients were selected for this study if they had suffered an acute ischemic stroke leading to enrollment in an experimental stroke protocol. To be included, patients must have had a formal measurement of the NIHSS score (as part of the experimental stroke protocol) at the time of admission. This information was collected prospectively during 3 separate clinical trials, providing 39 patients for analysis. The medical monitors for these clinical trials permitted the use of their case report forms to determine the actual NIHSS item and total scores for this study. For each eligible patient, the handwritten notes from the days of admission and discharge were photocopied, edited to remove any reference to the actual measured NIHSS scores, and then photocopied again for distribution to the raters. The raters estimated each component of the NIHSS for each patient on admission and discharge and calculated the total score. The admission and discharge notes were collated together so that raters could make inferences for a given patient about the discharge score based in part on the admission score, even if documentation was incomplete in the discharge note. This approach to handling poor or missing information was chosen because it approximates the actual process of retrospective chart abstraction. Actual and estimated mNIHSS scores were calculated by deletion of the 4 standard NIHSS items as described above and by dichotomization of sensory (item 8) scores as 0=normal and 1=abnormal.

Five raters of various levels of experience reviewed the records, including a stroke specialist–attending physician, stroke fellow, senior neurology resident, junior neurology resident, and nurse coordinator. All were previously trained and certified in the administration of the NIHSS. All were blinded to the actual NIHSS score and to each other’s ratings.

Statistical analysis was performed by use of STATA version 7.0 (Stata Corp). Scores were compared among raters, and interrater reliability was determined by calculation of an intraclass correlation coefficient (ICC) through analysis of variance. The ICC reflects the proportion of the total variance that is due to the “true” variance among patients and is calculated as \( \frac{s_r^2}{s_r^2 + s_e^2 + s_a^2} \), where \( s_r^2 \) is the variance component for subjects, \( s_a^2 \) is the variance component for the raters, and \( s_e^2 \) is the variance component for residual error.\(^{11}\) The ICC can be interpreted as a weighted \( \kappa \) statistic: an ICC of 1 suggests perfect reliability, and an ICC >0.75 is generally considered to represent excellent reliability.\(^{11}\) Pairwise comparisons between raters were also assessed with the ICC. Criterion validity was determined by comparison of estimated scores with actual scores as recorded on the case report forms of the clinical trials in which these patients had been enrolled.

### Results

Actual NIHSS item and total scores were available on admission for 39 patients, and of these 39, 30 had formal NIHSS measurements on discharge, for a total of 69 patient records. Notes were written by at least 1 neurologist (attending or resident physician) for all patients on admission and 28 of the 30 patients on discharge. Complete or nearly complete neurological examinations were documented in 38 of 39 admission notes but in only 15 of 30 of discharge notes.

Actual NIHSS total scores ranged from 0 to 23, with a median of 9; estimated standard NIHSS total scores ranged from 0 to 21, with a median of 9. Interrater reliability for estimated total standard NIHSS on admission and discharge was excellent, with ICCs of 0.85 and 0.79, respectively. Agreement between pairs of raters was also excellent, with ICCs ranging from 0.81 to 0.94. ICCs for individual items on admission are summarized in the Table. Similar results were found for discharge (data not shown). ICCs for 2 items were poor (<0.40), 6 were moderate (ICC, 0.40 to 0.75), and 7 were excellent (ICC >0.75).

### Reliability of Estimated NIHSS and mNIHSS Item Scores on Admission

<table>
<thead>
<tr>
<th>Item</th>
<th>ICC NIHSS 95% CI</th>
<th>ICC mNIHSS 95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>1a. LOC</td>
<td>0.74 (0.63–0.84)</td>
<td>...</td>
</tr>
<tr>
<td>1b. LOC questions</td>
<td>0.71 (0.60–0.82)</td>
<td>0.71 (0.60–0.82)</td>
</tr>
<tr>
<td>1c. LOC commands</td>
<td>0.78 (0.69–0.87)</td>
<td>0.78 (0.69–0.87)</td>
</tr>
<tr>
<td>2. Gaze</td>
<td>0.74 (0.64–0.84)</td>
<td>0.74 (0.64–0.84)</td>
</tr>
<tr>
<td>3. Visual fields</td>
<td>0.76 (0.67–0.86)</td>
<td>0.76 (0.67–0.86)</td>
</tr>
<tr>
<td>4. Facial palsy</td>
<td>0.27 (0.12–0.43)</td>
<td>...</td>
</tr>
<tr>
<td>5a. Left arm motor</td>
<td>0.91 (0.86–0.95)</td>
<td>0.91 (0.86–0.95)</td>
</tr>
<tr>
<td>5b. Right arm motor</td>
<td>0.90 (0.85–0.95)</td>
<td>0.90 (0.85–0.95)</td>
</tr>
<tr>
<td>6a. Left leg motor</td>
<td>0.90 (0.85–0.95)</td>
<td>0.90 (0.85–0.95)</td>
</tr>
<tr>
<td>6b. Right leg motor</td>
<td>0.89 (0.84–0.94)</td>
<td>0.89 (0.84–0.94)</td>
</tr>
<tr>
<td>7. Limb ataxia</td>
<td>0.70 (0.59–0.82)</td>
<td>...</td>
</tr>
<tr>
<td>8. Sensory</td>
<td>0.63 (0.50–0.76)</td>
<td>0.58 (0.44–0.72)</td>
</tr>
<tr>
<td>9. Language</td>
<td>0.92 (0.88–0.96)</td>
<td>0.92 (0.88–0.96)</td>
</tr>
<tr>
<td>10. Dysarthria</td>
<td>0.36 (0.21–0.52)</td>
<td>...</td>
</tr>
<tr>
<td>11. Neglect</td>
<td>0.59 (0.45–0.73)</td>
<td>0.59 (0.45–0.73)</td>
</tr>
<tr>
<td>Total score</td>
<td>0.85 (0.78–0.91)</td>
<td>0.87 (0.81–0.93)</td>
</tr>
</tbody>
</table>

LOC indicates level of consciousness.

Actual mNIHSS total scores ranged from 0 to 21, with a median of 6, whereas estimated mNIHSS total scores ranged from 0 to 18, with a median of 6. Interrater reliability for total mNIHSS was slightly greater than for the standard NIHSS, with ICCs of 0.87 and 0.89 on admission and discharge, respectively. Agreement between pairs of raters was also excellent, with ICCs ranging from 0.84 to 0.94. None of the 11 mNIHSS items had poor reliability, 4 were moderate (ICC, 0.40 to 0.75), and 7 were excellent (ICC >0.75).

Estimated scores were compared with actual scores. Sixty-two percent of estimated total NIHSS scores were within 2 points of actual scores and 91% were within 5 points, whereas 70% of estimated total mNIHSS scores were within 2 points and 95% were within 5 points.

### Discussion

The mNIHSS appears to offer a number of advantages over the standard NIHSS.\(^8,9\) It is simpler and shorter yet provides comparable reliability and validity. Our analysis suggests that the mNIHSS can be retrospectively estimated from the review of medical records with a high degree of reliability and validity, comparable to and perhaps slightly better than the standard NIHSS.

Given the inherent difficulties of chart review, estimation of the mNIHSS seems to be more “user friendly” than for the standard NIHSS and should therefore make this task less onerous in retrospective research. Observational retrospective cohort and case-control studies cannot substitute for carefully designed prospective trials but are often important for hypothesis generation and for situations in which randomized clinical trials and prospective cohorts are not feasible. The findings from the present study will be most useful for those retrospective studies in which information about stroke-related neurological deficits must be abstracted qualitatively and then transformed into a quantitative format for analysis.
These data were obtained from a single academic university setting in which the vast majority of patients were evaluated by neurologists. A recent study confirmed the reliability of chart abstraction of the NIHSS in academic settings and in community hospitals with acute neurological consultative services available, but this method proved much less reliable at community hospitals without such services. The simpler but less comprehensive and less commonly used Canadian Neurological Scale\textsuperscript{12,13} was comparable to the standard NIHSS in the first 2 hospital types but clearly outperformed the standard NIHSS in the community hospitals without acute neurological consultation.\textsuperscript{5} Further studies may determine whether the mNIHSS offers an improvement in this setting by eliminating and simplifying items that are characterized by poor interobserver reliability and may be lacking in medical records written by non-neurologists.

**Acknowledgment**

Dr Kasner was supported by NIH/NINDS grant K23-NS02147.

**References**


Modified National Institutes of Health Stroke Scale Can Be Estimated From Medical Records
Scott E. Kasner, Brett L. Cucchiara, Michael L. McGarvey, Jean M. Luciano, David S. Liebeskind and Julio A. Chalela

*Stroke.* 2003;34:568-570; originally published online January 16, 2003;
doi: 10.1161/01.STR.0000052630.11159.25

*Stroke* is published by the American Heart Association, 7272 Greenville Avenue, Dallas, TX 75231
Copyright © 2003 American Heart Association, Inc. All rights reserved.
Print ISSN: 0039-2499. Online ISSN: 1524-4628

The online version of this article, along with updated information and services, is located on the
World Wide Web at:
http://stroke.ahajournals.org/content/34/2/568

Permissions: Requests for permissions to reproduce figures, tables, or portions of articles originally published in *Stroke* can be obtained via RightsLink, a service of the Copyright Clearance Center, not the Editorial Office. Once the online version of the published article for which permission is being requested is located, click Request Permissions in the middle column of the Web page under Services. Further information about this process is available in the Permissions and Rights Question and Answer document.

Reprints: Information about reprints can be found online at:
http://www.lww.com/reprints

Subscriptions: Information about subscribing to *Stroke* is online at:
http://stroke.ahajournals.org//subscriptions/