Graded Neurologic Scale for Use in Acute Hemispheric Stroke Treatment Protocols

Robert J. Adams, Kimford J. Meador, Kapil D. Sethi, James C. Grotta, and David S. Thomson

A standardized neurologic assessment scoring instrument was developed and tested for use in a multicenter trial of hypervolemic hemodilution in acute hemispheric stroke. Components of the neurologic examination pertinent to hemispheric stroke syndromes were emphasized. The scale was evaluated using 16 acute stroke patients for concurrent validity (Pearson coefficient $r = 0.89$ compared with global assessments by neurologists or neurosurgeons) and interobserver reliability ($r = 0.95$ interobserver reliability estimate). Such a scale should prove useful in quantifying neurologic deficits in hemispheric stroke and in following changes in neurologic status during multicenter acute treatment protocols. (Stroke 1987;18:665–669)

In preparation for a multicenter controlled trial in the treatment of acute ischemic stroke, a graded neurologic assessment instrument (Hemispheric Stroke Scale, HSS) was developed, designed to emphasize aspects of the neurologic examination pertinent to acute hemispheric infarction. Existing indices such as the Barthel Index (BI)1 are useful in assessing functional status of stroke patients, while other simple assessment measures have been proposed that correlate with suitability for rehabilitation and independence in activities of daily living (ADL).2,3 Many acute stroke trials have used scales presented without documentation of the validity of the measuring instrument.4-9 A validated scale that could be applied soon after an infarction to document changes in the neurologic status of stroke patients undergoing acute treatment protocols was needed.

A scale appropriate for a multicenter trial should possess validity in that it measures deficits anticipated in the target population while minimizing "noise," i.e., items which score deficits that are unlikely to be encountered or that do not change with the evolution of the symptoms of interest. It should further possess reliability in that interobserver variation should be acceptably low. Recent reports10,11 have highlighted interobserver variability in neurologic assessments, a particular problem in hemispheric stroke patients with deficits in language or attention. Our objective was to produce and test a rating scale simple enough to be repeatedly applied by a neurologist or neurosurgeon in less than $1/2$ hour yet detailed enough to document and follow not only changes in extremity motor function but also major deficits in higher cortical function. The objective was not to replace other rating scales but to validate a scale designed for a specific group of experimental subjects.

**Subjects and Methods**

**Scale Design**

The scale was designed for noncomatose adults with signs of acute ischemia or infarction in the carotid circulation. From the combination of motor, sensory, and higher cortical function deficits encountered in this population, we tried to select the most pertinent deficits that could be reliably and rapidly assessed in most such patients. Initially, a graded neurologic examination was applied to 25 nonconsecutive hemispheric stroke patients at the Medical College of Georgia.

Based on this experience, individual cranial nerve testing was replaced with gaze paresis12 and visuospatial tasks.13 Tests of praxis and sensation, although important,2 were minimized because interobserver variability was high and because detailed testing was too often unreliable in dysphasic or inattentive patients. Motor and language function were emphasized. Functional measures, such as those used in the BI, were impractical in the early period immediately after the stroke and were not included.

The resulting scale is reproduced in Figures 1–3. A total of 100 deficit points is theoretically possible; the breakdown of deficit points by factors is shown in Table 1. The Glasgow Coma Scale, although not designed for stroke patients, was used because it is an accepted method for assessing the level of consciousness in neurologically impaired patients.

**Patient Population and Testing Methods**

This evaluation was conducted using 16 nonconsecutive hospitalized stroke patients at 6 centers in conjunction with a pilot study involving acute treatment with hypervolemic hemodilution. Inclusion criteria for

---

From the Department of Neurology, Medical College of Georgia, Augusta, Georgia (R.J.A., K.J.M., K.D.S.), the Department of Neurology, University of Texas Health Science Center at Houston, Houston, Texas (J.C.G.), and American Critical Care, Department of Biometrics, McGaw Park, Illinois (D.S.T.).

Supported in part by American Critical Care Study Number 606-83-33-02.

Address for reprints: Dr. Robert J. Adams, Department of Neurology, BIW 338, Medical College of Georgia, 1125 15th Street, Augusta, GA 30912-2366.

Received July 14, 1986; accepted December 22, 1986.
The treatment study was 1) age 35–80 years, 2) a clinical picture including neurologic deficit consistent with an acute (<24 hours) ischemic infarction in the carotid circulation, 3) a cranial computed tomogram (CT) consistent with acute hemispheric infarction without hemorrhage, 4) Glasgow Coma Scale score of ≥9, and 5) the absence of a clinical picture suggestive of lacunar stroke, metabolic encephalopathy, seizures, rapidly improving deficits, or severe medical illness. The rating scale was tested on 7 patients entered in the treatment study and on 9 similar patients who did not qualify for the treatment protocol on the basis of late presentation (24–96 hours). Eleven of these patients had follow-up paired exams 5–15 days later by a neurologist or neurosurgeon within a 5-hour period (mean 2.25 hours) between 12 and 96 hours (mean 22 hours) after the onset of symptoms. The resulting 54 paired exams were used to determine interobserver reliability.

All 16 patients were examined by 2 clinicians (either a neurologist or neurosurgeon) within a 5-hour period (mean 2.25 hours) between 12 and 96 hours (mean 22 hours) after the onset of symptoms. Eleven of these patients had follow-up paired exams 5–15 days later by the same examiners. The resulting 54 paired exams were used to determine interobserver reliability.

In addition, 39 of the 54 exams included an additional page on which the examiner chose one of the following global, qualitative terms to describe the patient’s overall deficit severity: normal, mild, mild-moderate, moderate, moderate–severe, or severe. These data were used to compute concurrent validity, a measure of the correlation between the numeric estimate of deficit by one examiner and the qualitative description of deficit by the other.

Examiners were given the option of marking “not assessable” (NA) for any item except those in the Glasgow Coma Scale that could not be reliably tested. All such items were scored uniformly by the statistician as a maximal deficit for that item.

To estimate how well this scale, which was designed to score neurologic dysfunction, also estimated functional impairment, it was compared with the BI by correlating the scores from the HSS and the BI, which were simultaneously applied by the same examiners to patients entered into a randomized, controlled treatment trial. This trial involved acute ischemic stroke patients similar to, but distinct from, the 16 patients in the validation series. A total of 103 pairs of examinations were available for comparison, which included evaluations performed from 1 week to 3 months after stroke.

---

**FIGURE 1.** Hemispheric Stroke Scale, page 1. Glasgow Coma Scale (inverted) and language assessment.

**FIGURE 2.** Hemispheric Stroke Scale, page 2. Assessment of visual fields, neglect, and cranial nerve function.
Statistical Analysis

Interobserver reliability was estimated by computing a correlation coefficient between the numeric ratings of the paired examinations.14 Concurrent validity was evaluated by computing both a Pearson correlation coefficient and a Spearman rank correlation.15 HSS and BI were compared by computing a correlation coefficient for the HSS and the inverted Barthel score.

Results

The interobserver reliability for each component of the HSS and the total test is shown in Table 1. Tests of motor function showed the least and sensory function the most interobserver variability. The total scale interobserver reliability was \( r = 0.95 \) (\( n = 27, p < 0.001 \)). The scores for each set of exams, shown in Figure 4, demonstrate that the deficits ranged from mild to severe, and with the exception of 1 set of exams, show generally good agreement between examiners. Concurrent validity was estimated by Pearson’s coefficient at \( r = 0.89 \) (\( n = 39, p < 0.001 \)) and by Spearman’s rank coefficient at \( r = 0.84 \) (\( p < 0.001 \)). Although the correlation coefficients are significant, the data (shown in Figure 5) demonstrate a wide range of scores classified by examiners as moderate–severe.

As anticipated, all 25 items in the HSS could not be assessed in all patients. Of 1,350 items in 54 examinations, 110 (8%) were marked NA. More than 90% of these 110 were found in the tests for neglect (III-E), visual construction (III-F), and/or stereognosis (V-B). However, there were only 8 examples (1% of 675 opportunities for discrepancy) in which the examiners did not agree on the assessability of particular items, indicating good agreement between clinicians on which items could be accurately tested. Most of these discrepancies occurred in dysphasic patients. The HSS score correlated well with estimates of functional impairment in ADL, as assessed by the BI (Figure 6). The correlation coefficient between the two scores for 103 pairs of examinations was 0.87.

Discussion

The HSS was designed for use in multicenter treatment protocols in acute hemispheric stroke. Although the motor examination is emphasized, the scale includes assessment of both dominant and nondominant hemisphere functions. Unlike a complete neurologic examination, it is specifically directed toward the deficits encountered in this population. The concurrent validity is excellent, indicating that the scale accurately reflects the overall impressions of the examining

Table 1. Division of the Hemispheric Stroke Scale by Factors, and Results of Interobserver Reliability Estimate

<table>
<thead>
<tr>
<th>Factor</th>
<th>Points</th>
<th>Reliability</th>
</tr>
</thead>
<tbody>
<tr>
<td>Level of consciousness</td>
<td>12</td>
<td>0.78</td>
</tr>
<tr>
<td>Language</td>
<td>20</td>
<td>0.86</td>
</tr>
<tr>
<td>Other cortical function</td>
<td>17</td>
<td>0.90</td>
</tr>
<tr>
<td>Motor</td>
<td>40</td>
<td>0.95</td>
</tr>
<tr>
<td>Sensory</td>
<td>11</td>
<td>0.75</td>
</tr>
<tr>
<td>Total</td>
<td>100</td>
<td>0.95</td>
</tr>
</tbody>
</table>

\( n = 27, p < 0.001 \) in all cases.
Severity scores of 1 examiner plotted against the global assessment of severity provided by a second examiner at concurrent applications of the rating scale. Examiners chose from the following descriptions of overall neurologic status: 1, normal; 2, mild; 3, mild-moderate; 4, moderate; 5, moderate-severe; and 6, severe.

Clinicians. However, what different examiners consider to be a moderately severe deficit (Grade 5, Figure 5) did vary considerably in numeric score, perhaps reflecting differences in personal bias as to the disabling effects of certain types of deficits such as language. The range of scores emphasizes the need for a numeric scale to more accurately grade changes in deficit. More significant is the high interobserver reliability. This is important for comparison of study results across several centers. It should be noted that the total reliability estimate exceeds any individual component except motor because the combination of multiple, somewhat independent, factors (i.e., motor vs. language) explains more of the total variance and improves the reliability estimate.

It is recognized that the Glasgow Coma Scale, designed for use in head injury, penalizes aphasic patients and overestimates their deficit. However, it does identify those patients whose responses are impaired beyond that expected for unilateral hemispheric infarction, in most cases because of brain edema. This sensitivity to the presence or development of significant mass effect and its familiarity are distinct advantages.

If a patient cannot perform a task, an examiner can simply score the patient as a maximal deficit or he can indicate, using the NA option, that the patient cannot be tested for that task. The latter approach is more representative and allows a central statistician to handle the data in a uniform manner, either scoring such items as maximal deficits or removing them from analysis and readjusting the scale. Either method applied consistently would be valid. The tasks most often found not assessable were line bisection, visual construction, stereognosis, and dysarthria. The first 3 were included to add sensitivity to the evaluation of cortical deficits, recognizing the difficulty with aphasic patients. Dysarthria cannot be assessed in the anarthric subject. These 4 items account for 10 points, indicating that an awake, globally aphasic subject can still be tested over 90% of the total scale. In this study, all examiners were aware of the NA option, but its use was not emphasized. It is recommended that all examiners be specifically instructed to indicate those tasks that cannot be assessed and why to minimize discrepancies between centers in future applications of the HSS.

Discrepancies in some of the paired examinations focus on the small but troublesome problem of NA items. Many treatment protocols require serial assessments by the same, usually blinded, observer throughout the study period. In such cases, consistency in the repeated application of this or any scale becomes the only important factor in describing the progress of an individual subject. In a multicenter study, however, discrepancies in the use of a scale make comparisons between centers less meaningful. The high interobserver reliability of the HSS makes it ideally suited in this regard.

Although not designed to measure impairment in ADL, the HSS scores correlated well with disability measured by the BI, which assesses the patient’s ability to perform ADL. In summary, the HSS is a valid and reliable scoring system that should prove useful in multicenter trials involving acute hemispheric infarction.

Acknowledgments

The authors wish to thank Drs. J. Rothrock, T. Keller, M. Pritz, and K. Von Roenn for their assistance in the evaluation of this scale and Pamm House for preparation of the manuscript.

References

6. Kaste M, Fogelholm R, Waltimo O: Combined dexamethasone...


**KEY WORDS** • neurologic scale • stroke
Graded neurologic scale for use in acute hemispheric stroke treatment protocols.
R J Adams, K J Meador, K D Sethi, J C Grotta and D S Thomson

Stroke. 1987;18:665-669
doi: 10.1161/01.STR.18.3.665

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/18/3/665

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/