Development of a Novel, Weighted, Quantifiable Stroke Scale

Japan Stroke Scale

Fumio Gotoh, MD; Yasuo Terayama, MD; Takahiro Amano, MD; for the Stroke Scale Committee of the Japan Stroke Society

**Background and Purpose**—Several stroke scales are available for estimation of the severity of stroke, but none of them provides information regarding the relative weights of the observed variables. To define an integrated severity of stroke, we developed a quantifiable stroke scale with weighted variables that apply conjoint analysis to calculate the relative weight of each item.

**Methods**—We selected 10 variables (consciousness, language, neglect, hemianopsia, gaze, pupillary abnormality, facial palsy, plantar reflex, sensation, and weakness) based on the multivariate analysis of the Keio Stroke Patient Database Battery. The variables were categorized and evaluated for their distribution and sensitivity. The categorizations were then modified and rechecked. The procedure was repeated until the appropriate categorization was obtained from 198 patients. A temporary stroke scale without weight was then formulated, and the reliability of the scale was examined and revised with 80 new stroke patients. As a next step, 150 neurologists were asked to rank a set of 27 virtual patients, each with a different combination of variables, according to severity. From these rankings, conjoint analysis was used to derive utility scores (weights) for each factor level.

**Results**—The relative weights of each of the factors were as follows: consciousness 49.8%, language 9.9%, weakness of lower extremity 7.3%, pupillary abnormality 6.8%, gaze palsy 5.6%, weakness of arm 4.3%, weakness of hand 3.7%, neglect 3.7%, facial palsy 2.4%, plantar reflex 2.2%, hemianopsia 2.2%, and sensory impairment 2.1%. The total score for a patient could be calculated from the sum of the scores for each of the variables ranging from −0.38 to 27.86. Scoring of 100 patients with acute stroke was carried out, and the changes in scores were followed for validation. Longitudinal clinical monitoring of the patients correlated well with the scores in each patient. The interrater and intrarater reliabilities of the scale were excellent (weighted κ 0.83; Cronbach’s α 0.998).

**Conclusions**—The Japan Stroke Scale is a parametric stroke scale that provides a quantitative measure of the severity of stroke. Each of the variables of the scale has a relative weight according to the severity of stroke. Reliability and responsiveness were proved to be excellent. The present data revealed a potentiality for the Japan Stroke Scale to be a universally accepted and reliable standardized system from the clinimetrical point of view. *(Stroke. 2001;32:1800-1807.)*

**Key Words:** stroke assessment ■ stroke outcome

**Stroke** scales represent a useful tool for estimating the severity of stroke at onset and for assessing prognostic information in hospital. In general, a stroke scale consists of several variables for observing the signs and symptoms, and each variable is categorized for scoring. The total score for a patient can be calculated from an integration of the scores for each of the variables. Several stroke scales are available, but none of them provide information regarding the relative weights of the observed variables, which is essential for integration of the scores. Some scales use no weight for the variables, and others use arbitrary weights. This prevents the available stroke scales from providing quantification of the obtained scores, although they are often erroneously handled as a quantifiable measure. They remain at most a nonparametric scale and offer no scientific grounds for being parametric. The limitations of the available stroke scales are derived mainly from the lack of an appropriate method for calculation of the part-worth of each of the variables.¹ We have found that conjoint analysis can be applied to compute the part-worth for each of the items of the scale. Conjoint analysis has rarely been used in the field of medical science, but it has often been applied in the field of marketing research.
Subjects and Methods
The study was conducted with a total of 378 patients with acute stroke at the 13 Japanese medical centers (Appendix), at which the members of the Stroke Scale Committee are in clinical practice. The ages of the patients ranged from 42 to 88 years (mean ± SD 67 ± 13 years).

Procedures for Development of a New Stroke Scale
The procedures for developing the stroke scale can be summarized as follows: (1) select the variables, (2) categorize the variables, (3) evaluate the categorizations for their distribution and sensitivity, (4) modify and reevaluate the categorizations, (5) repeat procedures 1 through 4 until the appropriate categorizations are obtained, (6) formulate a temporary stroke scale without weight, (7) examine the interrater and intrarater reliabilities, (8) calculate the relative weights (part-worths) for each of the variables by conjoint analysis, (9) formulate the weighted stroke scale, and (10) apply the scale in actual patients with acute stroke.

Selection of Variables
For the selection of items, multivariate analysis was applied to compute clinical features that could predict functional dependence or death for 1274 consecutive stroke patients during a 5-year period who were admitted to Keio University Hospital within 72 hours after onset and were registered with the Keio Stroke Patient Database. Table 1 summarizes the clinical profiles of these patients.

On the basis of the contribution of each item to the prognosis and of a review of currently available stroke scales (Table 2),1-33 10 items were selected as variables for development of the scale. Each of the variables was provisionally graded on a 2- to 5-point scale.

Categorization of Variables
Each of the variables was categorized into 2 to 5 categories depending on the nature of the variable. Each of the categories was expressed in a concrete way, avoiding abstractive expression, so that the same grade could be obtained regardless of the level of training of the rater.

Evaluation of Categorizations for Their Distribution and Sensitivity
For the purpose of content validation of the initial scale regarding whether the selected variables were reasonable for adequate measurement of the stroke severity, scoring of 133 patients with acute stroke (96 cerebral infarctions and 37 cerebral hemorrhages; mean age 68 ± 11 years old, M/F 80:53) was undertaken by 42 stroke specialists at the 13 institutes for 4 to 8 weeks (31 neurologists, 4 internists, and 7 neurosurgeons), and the distribution of the categorized variables was evaluated by the authors.

Modification and Reevaluation of the Categorization
After the adequacy of each variable had been examined with respect to the distribution and sensitivity of the categorized data, the scale was modified to improve these factors.

Repetition of the Procedure of Modification and Reevaluation of the Categorization
Modification and reevaluation of the categorization were repeated until the appropriate categorizations were obtained. For procedures 4 and 5, a total of 65 new patients (48 cerebral infarctions and 17 cerebral hemorrhages; mean age 65 ± 9 years old) were enlisted to formulate a temporary stroke scale without weight. Scoring of these patients was undertaken by 32 stroke specialists at the 13 institutes for 4 to 8 weeks (20 neurologists, 4 internists, and 8 neurosurgeons), and the distribution of the categorized variables was evaluated by the authors.

Formulation of the Temporary Stroke Scale Without Weight
After completion of the reevaluation, the revised scale was formulated as a temporary stroke scale without weight.

Examination of the Interrater and Intrarater Reliabilities
The interrater and intrarater reliabilities of this temporary scale were tested. The results were analyzed with $\kappa$ statistics.34 Eighteen new patients with stroke (14 cerebral infarctions and 4 cerebral hemorrhages; mean age 65 ± 12 years old) who were admitted consecutively to the National Cardiovascular Center (NCVC), Osaka, for a period of 2 weeks, were used to examine the interrater and intrarater reliabilities. For this purpose, 48 pairs of physicians (including 13 neurologists, 4 internists, and 5 neurosurgeons) from the 13 institutes gathered together at the NCVC and scored the patients. Based on this assessment, which was interpreted conventionally (Table 3),35 the variables that showed $\kappa$ values of <0.5 were modified to formulate a revised temporary scale. The interrater and intrarater reliabilities of the revised temporary scale were then retested by 62 pairs of
physicians with 62 new stroke patients (56 cerebral infarctions and 6 cerebral hemorrhages; mean age 66.8 years old) distributed among 11 facilities.

Calculation of Relative Weights for Each of the Variables by Conjoint Analysis

For estimation of weights, we focused on conjoint analysis. Such analysis has been extensively used in marketing research to estimate the impact of selected product characteristics on consumer preferences for products.2 The ORTHOPLAN program was applied to produce a set of 27 computed, hypothetical patient profiles with different combinations of neurological deficits that were sufficient to calculate the part-worth of each of the items. Each of the hypothetical patients was described individually on 3×5-inch cards (Figure 1). Ranking of the 27 virtual patients was undertaken by 150 board-certified neurologists and neurosurgeons who cared for stroke patients at the 13 institutes. The data obtained for preferences were applied to the part-worth utility functions of the neurological deficits with the use of conjoint analysis.

Application of the Weighted Scale in Patients With Acute Stroke

For the purpose of preliminary validation and assessment of the responsiveness of the scale, longitudinal monitoring of the JSS score for 4 weeks after onset was carried out for 100 new patients with acute stroke (17 cerebral hemorrhages, 21 cardiogenic embolisms, 52 atherothrombotic infarctions, and 10 lacunar infarctions) who were admitted to 1 of the 13 institutes. The time courses of the JSS scores after onset among these subtypes were compared. All of the patients received conventional supportive therapy for acute stroke.

Results

Reliability of the JSS Variables

Table 4 summarizes the results of the multivariate analysis showing the clinical features that contributed to the prognosis of patients with cerebral infarction and hemorrhage. On the basis of the contribution of each item to the prognosis and a review of currently available stroke scales, 10 items were selected as variables for development of the scale. Table 5 lists the 10 variables that were selected. After completion of the above-mentioned procedures, a resultant scale that involves these 10 variables with 2- or 3-graded categories (A, B, C) was developed (Figure 2). Table 6 lists the interrater and intrarater reliabilities of the variables of the scale. Clearly, the interrater reliability (weighted \( \kappa \)) of each of the variables was excellent (mean value 0.83). Also, Cronbach’s \( \alpha \) (mean value 0.998) indicated a high internal consistency (intrarater reliability) of the variables.

TABLE 3. Assessment of Interrater Reliability

<table>
<thead>
<tr>
<th>Weighted ( \kappa )</th>
<th>Measure of Agreement</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;0</td>
<td>Poor</td>
</tr>
<tr>
<td>0–0.20</td>
<td>Slight</td>
</tr>
<tr>
<td>0.21–0.40</td>
<td>Fair</td>
</tr>
<tr>
<td>0.41–0.60</td>
<td>Moderate</td>
</tr>
<tr>
<td>0.61–0.80</td>
<td>Substantial</td>
</tr>
<tr>
<td>0.81–1.00</td>
<td>Almost perfect</td>
</tr>
</tbody>
</table>
Relative Importance and Weights of the JSS Variables

As a result of the conjoint analysis, the relative importance and utility relative to the severity of the stroke were calculated as shown in Table 7. Across all responders in the study, level of consciousness (49.8%) was clearly the most important factor for determining the severity of acute stroke patients. Language (9.9%), weakness of leg (7.3%), and pupillary abnormality (6.8%) were the next most important factors for determining the stroke severity. Utility indicates the relative weight of the categories of each of the variables (Figure 2 and Table 7). The total score for a patient can be calculated as the sum of the utility scores for each of the variables and a constant (−14.71), and ranges from −0.38 (the best) to 27.86 (the worst).

Scales developed via conjoint analysis have been proved to be parametric. Our scale was developed by following this exact method, so the total score can be mathematically assumed to be a parametric value that represents the severity of stroke, although there is no gold standard for stroke severity. The methodological concepts and algorithm have been described precisely in previous publications.

Application of JSS in Patients With Acute Stroke: Serial Assessment of Acute Stroke With JSS

Figure 3 illustrates the time course of the JSS score after stroke onset in patients with cerebral hemorrhage, cardiogenic embolism, atherothrombotic infarction, or lacunar infarction. All patients received conventional supportive therapy for acute stroke. The mean admission scores for the patients with cerebral hemorrhage, cardiogenic embolism, atherothrombotic infarction, and lacunar infarction were 11.01, 11.98, 9.04, and 1.89, respectively. As shown, among the patients with cerebral hemorrhage and cardiogenic embolism, the JSS score was highest on the day of the attack and then gradually declined, indicating that stroke severity is worst on the day of the attack and then gradually improves with lapse of time. On the other hand, among the patients with atherothrombotic infarction, the JSS score at 3 to 5 days after the onset was the highest, indicating that stroke severity reaches its peak at 3 to 5 days after onset. These findings accurately reflected the overall impressions of the examining physicians.

Discussion

A reproducible and valid method for quantification of the neurological deficit that occurs after stroke is essential for monitoring patients; many stroke scales have been proposed for this purpose. Table 2 lists the stroke scales that are
available. As shown, stroke scales can be classified as stroke impairment scales on the basis of physical deficit, which are usually recognized as “stroke scales,” and as functional scales, which are used to evaluate functional recovery after stroke and can be subdivided into global outcome scales and activities of daily living scales. The JSS used in the present study can be classified as a physical deficit scale. The measurement of impairments at the acute phase of stroke is important for diagnosis and prognosis, whereas the measurement of impairments at the chronic phase of stroke is considered to be less important.37 de Haan et al38 concluded that the impact of impairments on functional outcomes tends to be underestimated by stroke scales.

In general, the important requirements for clinical measurements are (1) reliability, (2) validity, (3) responsiveness, and (4) quantitativeness.1 Some of the scales listed in Table 2 do satisfy the requirements 1 through 3. However, none have been universally accepted, validated, or standardized as a quantitative measure for the severity of stroke. This is due mainly to the lack of information concerning the relative weights, which would provide delineation among the categories of observed variables.1

Historically, the importance of a scoring system for proper assessment of the severity of stroke was initially addressed by Tuthill et al.3 Figure 4 summarizes the results of a MEDLINE survey showing the numbers of studies that used each of the physical deficit scales in the period of 1997 to 1999. The National Institutes of Health (NIH) Stroke Scale14 appeared most frequently in journals of neuroscience (31 journals), followed by the Scandinavian Stroke Scale10 (23 journals), Hemispheric Stroke Scale12 (19 journals), Canadian Stroke Scale11 (15 journals), Fugl-
Meyer Assessment Scale\(^5\) (7 journals), Toronto Stroke Scale\(^7\) and Mathew Stroke Scale\(^4\) (5 journals each), and Orgogozo Scale\(^15,16\) (4 journals), although none completely satisfied the important requirements for a stroke scale.

The NIH Stroke Scale is a 15-item scale in which higher scores represent greater deficit. It is widely used because of its simplicity and high interrater reliability.\(^39,40\) This scale has been well validated by showing a good concurrent correlation with the volume of cerebral infarction measured on CT scans at 7 days and a good predictive relation between initial NIH score and 7-day CT scans.\(^14,41\) However, it is not an ideal stroke scale from the clinimetric standpoint because the items are not weighted, although in practice they are arbitrarily weighted. The calculated score is thus not a quantitative measure in the strict sense.

Table 8 summarizes the main features of each of the stroke scales listed here. None of these scales are quantitative measures for the severity of stroke, as mentioned earlier. Focusing on this point, we attempted to develop a quantifiable stroke scale with objectively weighted variables to define an integrated severity of stroke that would meet the standards of high-quality clinimetrics. As a means of calculating relative weights for each of the variables based on the severity of stroke, we applied conjoint analysis.\(^2\) This method is derived from mathematical psychology and psychometrics. It has been used extensively in marketing research and is recognized as a useful method for measuring the relative weights that consumers place on a product, such as for estimating the impact of selected product characteristics on consumer preferences. The procedure is also of value for making global judgments of multifactorial phenomena, including quality of life, difficulty in patient care, etc. However, few reports have used this method in the medical field.\(^42–44\) We applied this method to estimate weights for each of the neurological signs and symptoms, on the basis of the preferences of physicians, for the severity of stroke.

Certain limitations associated with the study should be noted. In principle, because the method is based on the preferences of physicians for assessment of the severity of stroke, the weighting of the items could be biased by either the cultural background or the diagnostic habits of the group of physicians who responded to the questionnaire. Although there were no significant differences among the Japanese physicians involved, physicians’ senses of values tend to vary among nations. For this reason, when applying this scale to international studies, it is imperative to recheck the physicians’ sense of values in the countries concerned. This can be easily done through a questionnaire to the physicians involved to establish their preferences for assessment of the severity of hypothetical stroke patients.

Our scale contains 10 variables for examination and can be completed within several minutes in acutely ill patients. The scoring system can be applied easily by physicians or nurses to the stroke patients. The scores for each of the variables (Figure 2 and Table 7) and the constant of \(-14.71\) are added to give a total score. The total score for patients ranges from \(-0.38\) (the best) to \(27.86\) (the worst). This total score is a parametric value that represents the severity of stroke.

Regarding the concurrent validity and responsiveness, the present study revealed that the JSS accurately reflected the overall impressions of the examining physicians regarding the severity of stroke, although this holds true for some existing, nonparametric stroke scales, including NIH Stroke Scale, among others. One remarkable difference from other existing stroke scales is that the JSS can perform quantitative differentiation of the stroke scale at onset among stroke subtypes.

More significant was the high interrater reliability. This is important for making comparisons of study results across several centers. Concerning interrater reliability, the levels of agreement for the examination variables of our scale ranged from \(\kappa 0.67\) to \(\kappa 0.91\) (mean 0.83). In relation to previously published scales, the present scale is the most reliable scale in terms of interrater reliability.
TABLE 8. Main Features of Currently Available Stroke Scales

<table>
<thead>
<tr>
<th>Scale</th>
<th>Year</th>
<th>No. of Items</th>
<th>Weights on Items</th>
<th>Quantitativeness</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>NIHSS</td>
<td>1989</td>
<td>15</td>
<td>Arbitrarily weighted</td>
<td>No (ordinal scale)</td>
<td>Designed for nonneurologists; designed for multicenter hemodilution trials</td>
</tr>
<tr>
<td>Scandinavian Stroke Scale</td>
<td>1985</td>
<td>9</td>
<td>Arbitrarily weighted</td>
<td>No (ordinal scale)</td>
<td>Designed for hemodilution trials of ischemic stroke</td>
</tr>
<tr>
<td>Hemispheric Stroke Scale</td>
<td>1987</td>
<td>20</td>
<td>No</td>
<td>No (ordinal scale)</td>
<td>Designed to measure motor recovery in stroke patients</td>
</tr>
<tr>
<td>Canadian Stroke Scale</td>
<td>1986</td>
<td>8</td>
<td>Arbitrarily weighted</td>
<td>No (ordinal scale)</td>
<td>Designed for study of steroid therapy of stroke</td>
</tr>
<tr>
<td>Fugl-Meyer Assessment Scale</td>
<td>1975</td>
<td>28</td>
<td>No</td>
<td>No (ordinal scale)</td>
<td>Designed for study on glyceral and hemodilution for acute stroke</td>
</tr>
<tr>
<td>Toronto Stroke Scale</td>
<td>1976</td>
<td>11</td>
<td>Arbitrarily weighted</td>
<td>No (ordinal scale)</td>
<td>Designed for patients with middle cerebral artery infarction</td>
</tr>
<tr>
<td>Mathew Stroke Scale</td>
<td>1972</td>
<td>10</td>
<td>No</td>
<td>No (ordinal scale)</td>
<td>Applied in research on nimodipine and hemodilution for acute stroke</td>
</tr>
<tr>
<td>Modified Mathew Scale</td>
<td>1988</td>
<td>10</td>
<td>No</td>
<td>No (ordinal scale)</td>
<td>Designed for patients with middle cerebral artery infarction</td>
</tr>
<tr>
<td>Orgogozo Scale</td>
<td>1991</td>
<td>10</td>
<td>No</td>
<td>No (ordinal scale)</td>
<td>Designed for patients with middle cerebral artery infarction</td>
</tr>
</tbody>
</table>

Further assessments of the scale validity must be performed in a wide variety of situations, such as the selection of patients for drug trials, referral of acute stroke patients from physician to physician, evaluation of therapeutic modalities using the scale, etc. The additional prognostic value of the associated neuroimaging findings and volume of the lesions that cause stroke are under investigation.

In conclusion, the present study has shown that the JSS to be a parametric stroke scale that provides a quantitative measure of the severity of each variable. The reliability and responsiveness were found to be excellent, indicating that the JSS is the first novel and weighted severity scale for acute stroke patients that satisfies all of the important requirements for clinical measurements.

Appendix: JSS Committee

Chair: F. Gotoh (Department of Neurology, Keio University, Tokyo).

Committee members: M. Fujishima (Department of Neurology, Kyushu University, Fukuoka), Y. Fukuchi (Department of Neurology, Keio University, Tokyo), K. Hashi (Department of Neurosurgery, Sapporo Medical College, Sapporo), S. Hirai (Department of Neurology, Shimane Medical College, Shimane), E. Otomo (Department of Neurology, Yokohama City University, Yokohama), M. Kameyama (Department of Neurology, Sumitomo Hospital, Osaka), S. Kobayashi (Department of Neurology, Shimane Medical College, Shimane), T. Sawada (Department of Neurology, National Cardiovascular Center, Osaka), Y. Shinohara (Department of Neurology, Tohoku University, Sendai), A. Tamura (Department of Neurosurgery, Teikyo University, Tokyo), T. Terashi (Department of Neurology, Nihon Medical College, Tokyo), T. Yanagisawa (Department of Neurology, National Cardiovascular Center, Osaka), T. Yanagihara (Department of Neurology, Osaka University, Osaka), and T. Yoshimoto (Department of Neurosurgery, Tohoku University, Sendai).

Secretaries: T. Amano (Department of Neurology, Keio University, Tokyo) and Y. Terayama (Department of Neurology, Yokohama Stroke and Brain Center, Yokohama).

References


Development of a Novel, Weighted, Quantifiable Stroke Scale: Japan Stroke Scale
Fumio Gotoh, Yasuo Terayama, Takahiro Amano and for the Stroke Scale Committee of the Japan Stroke Society

Stroke. 2001;32:1800-1807
doi: 10.1161/01.STR.32.8.1800

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/32/8/1800

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/