Evaluation of Fatigue Scales in Stroke Patients

Gillian Mead, MD; Joanna Lynch, MA; Carolyn Greig, PhD; Archie Young, MD; Susan Lewis, PhD; Michael Sharpe, MD

Background and Purpose—There is little information on how to best measure poststroke fatigue. Our aim was to identify which currently available fatigue scale is most valid, feasible, and reliable in stroke patients.

Methods—Fatigue scales were identified by systematic search, and the 5 with the best face validity were identified by expert consensus. Feasibility (ie, did patients provide answers?) and internal consistency (an aspect of reliability) of these scales were evaluated by interviewing 55 stroke patients. Test-retest reliability was assessed by reinterviewing 51 patients, interrater reliability was assessed by rerating audio recordings, and convergent validity was assessed by measuring the correlation between scale scores.

Results—Of the 52 scales identified, the SF-36v2 (vitality component), the fatigue subscale of the Profile of Mood States, the Fatigue Assessment Scale, the general subscale of the Multidimensional Fatigue Symptom Inventory, and the Brief Fatigue Inventory had the best face validity. The Brief Fatigue Inventory was unfeasible to administer and was omitted. Of the remaining 4 scales, the Fatigue Assessment Scale had the poorest internal consistency. Test-retest reliability for individual scale questions ranged from fair to good; the Fatigue Assessment Scale had the narrowest limits of agreement for the total score, indicating the best test-retest reliability. Interrater reliability for individual questions ranged from good to very good, and there was no significant mean difference in total scores for any scale. Convergent validity was moderate to high for the total scores of the 4 scales.

Conclusions—All four scales were valid and feasible to administer to stroke patients. The Fatigue Assessment Scale had the best test-retest reliability but the poorest internal consistency. (Stroke. 2007;38:2090-2095.)

Key Words: complications ■ quality of life scales ■ stroke recovery ■ fatigue

Fatigue is defined as a feeling of lack of energy, weariness, and aversion to effort. Although it has been well studied in neurological diseases such as multiple sclerosis and Parkinson’s disease,1 emerging evidence indicates that fatigue is also common after stroke but to date has been relatively neglected.2–8 Previous cross-sectional studies3–8 have all used scales devised for use in conditions other than stroke, and consequently their validity in stroke patients is unknown. For example, the question “Do you feel weak?” does not have face validity for stroke because patients’ muscle weakness is generally due to the local weakness of hemiparesis rather than to generalized fatigue. Two scales9,10 ask patients to rate the extent to which fatigue interferes with physical functioning, even though some stroke patients cannot disentangle the effect of their fatigue from that of their neurological deficit.8 The reliability (ie, does the measure as a whole, and do the items included in the measure, give repeatable results when used under the same circumstances11?) of existing fatigue scales when used in stroke patients has not previously been determined. Our aims were to determine the best scales for measuring the severity of fatigue after stroke by determining face validity, feasibility, reliability (internal consistency, test-retest, and interrater), and convergent construct validity of currently available scales.

Methods

Identification of Currently Available Fatigue Scales

One researcher (G.M.) performed a search of MEDLINE (1966 to February 2004) in July 2004 by using the search terms “fatigue” (and related words such as “tiredness”), “scale,” “assessment,” “measurement.” Abstracts were screened, potentially relevant publications were obtained, and reference lists of these articles were screened. A second search was performed in October 2006 (after completion of the patient interviews) to check whether the literature had changed significantly.

Face Validity of Fatigue Scales

Full-text articles were then independently scrutinized by 4 observers to determine the face validity of each scale for stroke patients. One of these observers (G.M.) was a practicing stroke physician with research experience of fatigue in stroke patients; the other 3 all had clinical or research experience in assessing patients with stroke and/or fatigue. Face validity was assessed on the criteria that the
scale (1) captured the phenomenon of poststroke fatigue and (2) was free from items indistinguishable from the effects of the stroke, eg, “My limbs feel weak.” Each observer independently listed their 5 preferred scales and tabled their preferences at a consensus meeting. Any differences were resolved by discussion. As a further check of face validity, the 5 chosen scales were pilot tested in 13 stroke inpatients.

Patient Interviews for Feasibility, Reliability, and Convergent Construct Validity

Ethics approval was granted by the local ethics committee. All participating patients gave written consent. Patients were recruited from hospital stroke wards (at least 1 week after stroke onset) and from the community via stroke clinics or community nurses. Patients with dysphasia or confusion severe enough to prevent them from understanding the rationale for the study or giving informed consent were excluded, as were those who were medically unstable because of another condition. The side of the brain lesion, stroke subtype (Oxfordshire Community Stroke Project classification), and computed tomography brain results were recorded from patient’s records. After verifying that the patient understood the meaning of the word “fatigue,” the scales were administered verbally and the interview was audio recorded.

Feasibility

As a measurement of feasibility, we recorded whether the patient provided an appropriate response to each item of each scale.

Reliability

Responses to individual scale items were used to determine the internal consistency of each scale. For test-retest reliability, the scales were readministered verbally by the same interviewer 3 days later. Responses from both interviews were used to determine the test-retest reliability of individual scale items and total scale scores. For interrater reliability, the audio recordings of the first interview were rerated by a second rater who was blinded to the initial rating to determine interrater reliability of individual scale items and total scale scores.

Internal consistency was calculated by Cronbach’s \( \alpha \). Test-retest and interrater agreements between the individual items of each scale were analyzed by percent agreement, the weighted kappa statistic (for items with \( >2 \) response levels, by Stat-Xact), and Cohen’s kappa (for items with \( 2 \) response levels). Interpretation of kappa was based on published guidelines. \(^{12} \) Test-retest and interrater agreement for total test scores was analyzed by the Bland-Altman method \(^{13} \) and intraclass correlation coefficients.

Convergent Construct Validity

Convergent construct validity was determined by the Spearman correlation coefficient of ratings on each of the scales completed at the first interview. To assess convergent construct validity, the total scores for each scale obtained at the first interview were correlated with each of the other scale scores.

Statistical Analysis

All analyses were done with SPSS version 11.0 unless stated otherwise.

Results

Identification of Fatigue Scales

Pertinent citations (N=6122) were identified. From these citations, 52 fatigue scales were identified, none of which had been designed for use in stroke patients. A second search in October 2006 identified 3 additional scales devised for conditions other than stroke.

Face Validity

Five scales were deemed to have adequate face validity: the vitality subscale of the SF-36v2, \(^{14} \) the fatigue subscale of the Profile of Mood States (POMS-fatigue), \(^{15} \) the Fatigue Assessment Scale (FAS), \(^{16} \) the general subscale of the Multidimensional Fatigue Symptom Inventory (MFSI-general), \(^{17} \) and the Brief Fatigue Inventory (BFI). \(^{18} \) After pilot testing in 13 stroke inpatients, one question on the MFSI-general (“I feel exhausted.”) was poorly understood and was changed to “I feel exhausted.”

Patient Interviews for Feasibility, Reliability, and Convergent Construct Validity

Sixty-four patients were invited to participate, of whom 55 consented. Their median age was 73 years (interquartile range, 66 to 81 years), 31 (56%) were male, 40 were inpatients, and 15 were resident in the community. Twenty-seven (49%) had a right hemisphere stroke. Three (6%) had a hemorrhagic stroke. Eleven (20%) had total anterior circulation syndromes, 22 (40%) had partial anterior circulation syndromes, 16 (29%) had lacunar syndromes, and 6 (11%) had posterior circulation syndromes. Forty-three (78%) had a relevant stroke lesion on brain computed tomography scans. The median time between stroke and first assessment was 23 days (interquartile range, 10 to 53 days) for inpatients and 137 days (interquartile range, 93 to 217 days) for community patients.

All 55 patients were successfully interviewed at time 1, and 51 of these were interviewed at time 2. Reasons for not having a second interview were discharge before it was due (n=1), deterioration in medical condition (n=2), and refusal (n=1). The mean duration between interviews (test-retest reliability) was 3.9 days (range, 3 to 7 days).

Feasibility

At the first interview, every patient answered every item of the SF-36v2 (vitality component), POMS-fatigue, and MFSI-general. Three patients could not complete the BFI, an additional 8 patients could not answer the question on the interference of fatigue with “walking ability,” and one was unable to answer the question on the interference of fatigue with “normal work” (both items in the BFI). Three patients each could not answer one of three different items on the FAS.

Of those attempting the second interview (n=51), all items on the SF-36v2 (vitality component) and FAS were answered by every patient. One patient was unable to answer “In the past week I feel sluggish?” (POMS-fatigue) and “In the past week I feel sluggish” (MFSI-general). Three patients did not attempt the BFI at the second interview, and an additional three could not answer the BFI question on the interference of fatigue with “walking ability.” These data demonstrated that the BFI is the least feasible scale to administer, so the BFI was dropped from further analysis for this reason.

Internal Consistency

With respect to internal consistency, for the first and second interview, Cronbach’s \( \alpha \) was 0.91 and 0.93, respectively, for the MFSI-general; 0.89 and 0.88, respectively, for the POMS-...
fatigue; 0.76 and 0.78, respectively, for the SF-36v2 vitality score; and 0.58 and 0.62, respectively, for the FAS.

**Test-Retest Reliability**

The kappa values for the individual items of all 4 scales ranged from fair to good (Table 1). The agreement between total scores for each scale is shown in Table 2 and as Bland-Altman plots (the Figure). The horizontal axes of the Bland-Altman plots span the entire range of the scale, whereas the vertical axes cover the range of possible differences. The scale with the narrowest scatter of differences (ie, limits of agreement) was the FAS (Figure, c). The 95% CIs for the mean difference between the first interview and the second for the total scores of POMS-fatigue and MFSI-general did not include zero, demonstrating that there was a significant mean difference between the 2 interviews, whereas there was no significant mean difference between interviews for the FAS and SF-36v2-vitality (Table 2).

### Table 1. Test-Retest and Interrater Agreement for Individual Questions in Each Scale: Reliability (Test-Retest and Interrater) for Individual Scale Items

<table>
<thead>
<tr>
<th>Scale</th>
<th>Item</th>
<th>Test-Retest</th>
<th></th>
<th>Interrater</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Percentage</td>
<td>Kappa (95% CI),*</td>
<td>n</td>
<td>Level of Agreement</td>
</tr>
<tr>
<td>SF36v2</td>
<td>In the past 4 weeks, did you feel full of life?</td>
<td>43%</td>
<td>0.35 (0.07–0.63)</td>
<td>51</td>
<td>Fair</td>
</tr>
<tr>
<td></td>
<td></td>
<td>59%</td>
<td>0.46 (0.14–0.77)</td>
<td>51</td>
<td>Moderate</td>
</tr>
<tr>
<td></td>
<td></td>
<td>43%</td>
<td>0.47 (0.25–0.70)</td>
<td>51</td>
<td>Moderate</td>
</tr>
<tr>
<td></td>
<td></td>
<td>43%</td>
<td>0.43 (0.22–0.63)</td>
<td>51</td>
<td>Moderate</td>
</tr>
<tr>
<td>POMS (fatigue)</td>
<td>In the past week, do you feel tired?</td>
<td>65%</td>
<td>0.51 (0.29–0.73)</td>
<td>51</td>
<td>Moderate</td>
</tr>
<tr>
<td></td>
<td></td>
<td>57%</td>
<td>0.52 (0.32–0.72)</td>
<td>51</td>
<td>Moderate</td>
</tr>
<tr>
<td></td>
<td></td>
<td>55%</td>
<td>0.59 (0.41–0.76)</td>
<td>51</td>
<td>Moderate</td>
</tr>
<tr>
<td></td>
<td></td>
<td>56%</td>
<td>0.61 (0.42–0.80)</td>
<td>50</td>
<td>Good</td>
</tr>
<tr>
<td></td>
<td></td>
<td>57%</td>
<td>0.52 (0.30–0.73)</td>
<td>51</td>
<td>Moderate</td>
</tr>
<tr>
<td></td>
<td></td>
<td>55%</td>
<td>0.45 (0.19–0.72)</td>
<td>51</td>
<td>Moderate</td>
</tr>
<tr>
<td>FAS</td>
<td>I am bothered by fatigue</td>
<td>37%</td>
<td>0.37 (0.17–0.56)</td>
<td>51</td>
<td>Fair</td>
</tr>
<tr>
<td></td>
<td></td>
<td>47%</td>
<td>0.49 (0.27–0.71)</td>
<td>51</td>
<td>Moderate</td>
</tr>
<tr>
<td></td>
<td></td>
<td>44%</td>
<td>0.50 (0.30–0.71)</td>
<td>50</td>
<td>Moderate</td>
</tr>
<tr>
<td></td>
<td></td>
<td>46%</td>
<td>0.44 (0.21–0.66)</td>
<td>50</td>
<td>Moderate</td>
</tr>
<tr>
<td></td>
<td>Physically I feel exhausted</td>
<td>51%</td>
<td>0.64 (0.48–0.81)</td>
<td>51</td>
<td>Good</td>
</tr>
<tr>
<td></td>
<td></td>
<td>52%</td>
<td>0.53 (0.31–0.75)</td>
<td>50</td>
<td>Moderate</td>
</tr>
<tr>
<td></td>
<td></td>
<td>71%</td>
<td>0.77 (0.61–0.92)</td>
<td>51</td>
<td>Good</td>
</tr>
<tr>
<td></td>
<td></td>
<td>49%</td>
<td>0.33 (0.05–0.61)</td>
<td>51</td>
<td>Fair</td>
</tr>
<tr>
<td></td>
<td></td>
<td>51%</td>
<td>0.65 (0.46–0.84)</td>
<td>51</td>
<td>Good</td>
</tr>
<tr>
<td></td>
<td></td>
<td>47%</td>
<td>0.48 (0.25–0.71)</td>
<td>51</td>
<td>Moderate</td>
</tr>
<tr>
<td>MFSI (general)</td>
<td>In the past week, I feel</td>
<td>41%</td>
<td>0.58 (0.41–0.74)</td>
<td>51</td>
<td>Moderate</td>
</tr>
<tr>
<td></td>
<td>exhausted</td>
<td>49%</td>
<td>0.64 (0.47–0.81)</td>
<td>51</td>
<td>Good</td>
</tr>
<tr>
<td></td>
<td></td>
<td>45%</td>
<td>0.48 (0.27–0.69)</td>
<td>51</td>
<td>Moderate</td>
</tr>
<tr>
<td></td>
<td></td>
<td>39%</td>
<td>0.56 (0.37–0.74)</td>
<td>51</td>
<td>Moderate</td>
</tr>
<tr>
<td></td>
<td></td>
<td>60%</td>
<td>0.66 (0.47–0.84)</td>
<td>50</td>
<td>Good</td>
</tr>
<tr>
<td></td>
<td></td>
<td>59%</td>
<td>0.69 (0.53–0.85)</td>
<td>51</td>
<td>Good</td>
</tr>
</tbody>
</table>

*No. of possible responses: for SF36 (vitality), 5; for POMS (fatigue), 4; for FAS, 5; for MFSI (general), 5.

*Weighted kappa: all P values <0.01.
interclass correlation coefficient for total test-retest FAS scores was higher than for SF-36v2-vitality (0.77 vs 0.51; Table 2).

**Interrater Reliability**
Forty-three interviews were available for rerating. Interrater reliability for the individual items of each scale is shown in Table 1. The kappa values indicated good to very good interrater reliability. For total test scores, there was no significant mean difference between raters (Table 2).

**Convergent Construct Validity**
Convergent construct validity of the total scale scores for the POMS-fatigue, FAS, and MFSI-general scale was moderate to high (Table 3). The construct validity for FAS and MFSI-general was higher than for SF-36v2 and MFSI-general (0.71 vs 0.47).

**Discussion**
This is the first evaluation of available fatigue scales to determine the most valid, feasible to administer, and reliable for use in stroke patients. Of the 52 fatigue scales identified by our literature search, many contained items that could be confused with the neurological effect of stroke. We selected 5 with the best face validity in stroke and tested their psychometric properties.

We deliberately recruited both hospital and community patients to improve the generalizability of our results. When we applied our 5 chosen scales by interviewing stroke patients, the least feasible was the BFI, perhaps because patients had to grade their fatigue on a numeric scale of 0 to 10 and quantify the extent to which fatigue interfered with different activities. Patients frequently had difficulty distinguishing the effect of their neurological impairment from the effect of fatigue. The BFI was therefore dropped from further analysis.

We assessed 3 aspects of reliability. First, internal consistency was lowest for the FAS, possibly because the FAS asks about different facets of fatigue, whereas the other scales ask several similar questions with only slight differences in vocabulary. Although an “ideal” scale may have high internal consistency, a scale measuring slightly different facets of a symptom as complex as fatigue is arguably more useful in practice. Second, test-retest reliability for individual items (as determined by kappa values) was similar for all 4 scales tested, indicating that no scale outperformed another in this respect (Table 1). However, for total test scores, the FAS performed the best, as it had the narrowest limits of agreement (Figure, c), there was no significant mean difference between the first and second interviews (Table 2), and the intraclass correlation coefficient was high, at 0.77 (Table 2). The SF-36v2-vitality had the widest limits of agreement, perhaps reflecting differences in the questions: the SF-36v2-vitality asks about the previous 4 weeks, whereas the POMS-fatigue and MFSI-general ask about the “last week” and the FAS asks about the present time. Third, analysis of interrater reliability indicated that no 1 scale outperformed another: the proportion of kappa scores in each category was similar for all scales (Table 1), and there was no significant mean difference between observers for total test scores (Table 2).

Construct validity of the total scale scores was moderate to high (Table 3). Because some of the scales contained the same (or very similar) questions, this is not unexpected. We noted that scores in our patients indicated higher levels of fatigue than in nonstroke subjects, eg, SF-36v2-vitality and FAS.16,19

There are some weaknesses to the study. First, our patients were not consecutive and may not be representative of stroke patients as a whole. Second, although a larger sample size would have given more precise estimates, a sample size of 50 is usually considered sufficient for studies of agreement.12 Third, although we assessed face validity and convergent construct validity, the absence of any “gold standard” for fatigue after stroke means we could not assess criterion validity. Fourth, when assessing test-retest reliability, the interviewer may have remembered the results of the first interview when performing the second, thereby artificially increasing apparent reliability. Fifth, when assessing interrater reliability, we used audio recordings rather than repeat...
interviews, again potentially increasing apparent reliability. Finally, not all of the interviews could be analyzed for interrater reliability because of the poor quality of some recordings, mainly due to background noise on hospital wards.

The four scales are all usable. Our “best buy” depends to some extent on the intended use, but on the basis of our data, we would recommend the use of the FAS to measure fatigue after stroke because it had face validity, it was feasible for most patients, and it had the best test-retest reliability and high construct validity. However, if high internal consistency is a priority, one of the other three scales should be considered.

Acknowledgments
We are grateful to the patients who participated and to the staff who assisted us in identifying patients.

Source of Funding
The study received funding from the Chief Scientist Office of the Scottish Executive Health Department (reference CZG/2/161).

Disclosures
None.

References

**TABLE 3. Construct Validity* for Comparison of Total Scale Scores: Convergent Construct Validity of Total Scale Scores**

<table>
<thead>
<tr>
<th>Scale</th>
<th>POMS (fatigue)</th>
<th>FAS</th>
<th>MFSI (general)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SF-36v2 (vitality)</td>
<td>-0.58 (&lt;0.001)</td>
<td>-0.41 (0.003)</td>
<td>-0.47 (&lt;0.001)</td>
</tr>
<tr>
<td></td>
<td>n=55</td>
<td>n=52</td>
<td>n=55</td>
</tr>
<tr>
<td>POMS (fatigue)</td>
<td>0.59 (&lt;0.001)</td>
<td>0.75 (&lt;0.001)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>n=52</td>
<td>n=55</td>
<td></td>
</tr>
<tr>
<td>FAS</td>
<td>0.71 (&lt;0.001)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>n=52</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Spearman correlation coefficient (ρ) for time 1 (first rater).
Evaluation of Fatigue Scales in Stroke Patients
Gillian Mead, Joanna Lynch, Carolyn Greig, Archie Young, Susan Lewis and Michael Sharpe

Stroke. 2007;38:2090-2095; originally published online May 24, 2007;
doi: 10.1161/STROKEAHA.106.478941
Stroke is published by the American Heart Association, 7272 Greenville Avenue, Dallas, TX 75231
Copyright © 2007 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/38/7/2090