Responsiveness and Validity of Three Outcome Measures of Motor Function After Stroke Rehabilitation

Yu-wei Hsieh, MS; Ching-yi Wu, ScD, OTR; Keh-chung Lin, ScD, OTR; Ya-fen Chang, MS; Chia-ling Chen, MD, PhD; Jung-sen Liu, MD, PhD

**Background and Purpose**—This study investigated and compared the responsiveness and validity of the Fugl-Meyer Assessment (FMA), the Action Research Arm Test (ARAT), and the Wolf Motor Function Test (WMFT) for patients after stroke rehabilitation.

**Methods**—A total of 57 patients with stroke received 1 of 3 rehabilitation treatments for 3 weeks. At pretreatment and posttreatment, the 3 outcome measures, as well as the Functional Independence Measure (FIM) as the external criterion, were administered. The standardized response mean (SRM) and the Wilcoxon signed rank test were used to examine the responsiveness. Construct validity and predictive validity were examined by the Spearman correlation coefficient ($\rho$).

**Results**—The responsiveness of the FMA, ARAT, and WMFT functional ability scores was large (SRM = 0.95–1.42), whereas the WMFT performance time score was small (SRM = 0.38). The responsiveness of the FMA was significantly larger than those of the ARAT and the WMFT-TIME, but not the WMFT functional ability scores. With respect to construct validity, correlations between the FMA and other measures were relatively high ($\rho=0.42$–0.76). The FMA and the WMFT performance time scores at pretreatment had moderate predictive validity with the FIM scores at posttreatment ($\rho=0.42$–0.47). In addition, the ARAT and the WMFT functional ability scores revealed a low predictive validity with the FIM ($\rho=0.17$–0.26).

**Conclusions**—The results support the FMA and the WMFT-FAS are suitable to detect changes over time for patients after stroke rehabilitation. While simultaneously considering the responsiveness and validity attributes, the FMA may be a relatively sound measure of motor function for stroke patients based on our results. Further research based on a larger sample is needed to replicate the findings. ([Stroke. 2009;40:1386-1391.])

**Key Words:** cerebrovascular accident ■ rehabilitation ■ outcome ■ upper extremity ■ clinimetrics

Most stroke survivors suffer persistent impairment of upper extremity (UE) movement. Various contemporary rehabilitation strategies which aim to improve upper extremity motor deficits after stroke based on repetitive practice, such as constraint-induced therapy (CIT) and bilateral arm training (BAT), have come into wide use. CIT first proposed by Taub et al involves restraint of the less affected UE over an extended period, in combination with intensive task-specific training of the more affected limb. Numerous studies in stroke patients have shown substantial evidence of the efficacy of the CIT and its variants. BAT is an alternative treatment approach. The unaffected UE performs the spatiotemporal pattern of movement as that performed by the affected UE. Practice of the bilateral movement allows the activation of the intact hemisphere to facilitate the activation of the damaged hemisphere and subsequently improves the motor function of the affected UE.

Critical evaluation of the effects of these intervention strategies requires measures that have good clinimetric properties (eg, reliability, validity, and responsiveness) in assessing upper extremity motor function. Thus, extensive demonstrations of the clinimetric properties of outcome measures are important for their application in clinical trials. Reliable and valid measures are sufficient to ensure the usefulness of measures in detecting cross-sectional differences. However, responsiveness is essential for assessing the effectiveness of treatment as well as measuring longitudinal change over time. Responsiveness is defined as the ability of a measure to detect changes which have occurred accurately over time. There has been an increasing emphasis on the importance of investigating the responsiveness of a measure, but there is no consensus on the advantage of one method over another, and multiple methods have been suggested.
The Fugl-Meyer Assessment (FMA), Action Research Arm Test (ARAT), and Wolf Motor Function Test (WMFT) have been widely used as outcome measures in stroke rehabilitation research, including CIT and BAT trials. However, the relative capacity among these measures to detect change after interventions remains unclear, and there is no consensus concerning the optimal outcome measures for UE motor function in stroke patients. Inconsistent usage of outcome measures might lead to discrepancies in the interpretation of the data, as well as make comparisons or synthesis of the study results difficult.

Another concern is the lack of a study that directly compares the responsiveness and validity of the 3 measures under the supervision of 3 separate certified occupational therapists. These 3 therapists were trained in the administration of the intervention protocols and assessment was carried out by the senior authors, and a written competency test was administered before subject treatment and assessment. At pretreatment and posttreatment, the outcome evaluations were administered by 3 raters masked to the participant group and trained to properly administer the outcome measures.

**Outcome Measures**

**Fugl-Meyer Assessment**

The upper-extremity subscale of the Fugl-Meyer Assessment (FMA) was used to assess motor impairment. The 33 UE items measure the movement and reflexes of the shoulder/elbow/forearm, wrist, hand, and coordination; they are scored on a 3-point ordinal scale (0—cannot perform, 1—performs partially, 2—performs fully). The maximum score is 66, indicating optimal recovery. The psychometric properties of the FMA have been shown to be satisfactory in stroke patients.

**Action Research Arm Test**

The Action Research Arm Test (ARAT) was designed for evaluation of upper extremity function. It consists of 19 items (maximum score 75) that are divided into 4 subscales: grasp, grip, pinch, and gross movement. The items are scored on a 4-level ordinal scale (0—can perform no part of test, 1—performs test partially, 2—completes test, but takes an abnormally long time or has great difficulty, 3—performs test normally). The reliability, validity and responsiveness of the ARAT have been established in patients with stroke.

**Wolf Motor Function Test**

The Wolf Motor Function Test (WMFT) was originally developed to assess motor function of upper extremity. It consists of 19 items (maximum score 66) that are divided into 4 subscales: grasp, grip, pinch, and gross movement. The items are scored on a 4-level ordinal scale (0—cannot perform, 1—performs partially, 2—performs fully). The maximum score is 66, indicating optimal recovery. The psychometric properties of the WMFT have been shown to be satisfactory in stroke patients.

**Interventions and Procedures**

Eligible participants were individually randomized into the distributed CIT, BAT, or traditional rehabilitation group, with the computerized (block) randomization scheme including prestratification according to participating hospital. One set of opaque, numbered envelopes was prepared for each site containing cards indicating the allocated group. When a new patient was registered, a card was extracted and the participant group and trained to properly administer the outcome measures.

**Criterion Measure**

**Functional Independence Measure**

For examining construct and predictive validity of measures, a gold standard or external criterion is needed. Independence in daily activities is the optimal goal of stroke rehabilitation and the Functional Independence Measure (FIM) is a widely used measure of functional independence in stroke rehabilitation. Compared to other functional assessments, the FIM revealed preferable or similar properties in rehabilitation patients. Besides, not only global disability scores but also scores defined disability in motor function can be obtained from the FIM. Thus, the FIM was selected as the functional criterion for validating the 3 outcome measures in this study. The FIM consists of 18 items grouped into 6 subscales measuring self-care, sphincter control, transfer, locomotion, communication, and social cognition abilities. Each item is rated from 1 to 7 (maximum score 126) based on the required level of assistance to perform the tasks (1—complete assistance, 2—maximal assistance, 3—moderate assistance, 4—minimal assistance, 5—supervision, 6—modified independence, 7—complete independence). The 13 items of self-care, sphincter control, transfer, and locomotion subscales requiring motor function were summed to create an FIM-motor score with a range moving synchronously, such as lifting 2 cups, picking up 2 pegs, reaching forward or upward to move blocks, grasping and releasing 2 towels, etc. The traditional rehabilitation group focused on neurodevelopmental techniques with emphasis on functional tasks when possible. Therapy included stretching of the more affected limb, training for strength, hand function, and coordination, and functional task practice for 2 hours per weekday. The interventions were provided at the participating sites under the supervision of 3 separate certified occupational therapists. These 3 therapists were trained in the administration of the intervention protocols and assessment was carried out by the senior authors, and a written competency test was administered before subject treatment and assessment. At pretreatment and posttreatment, the outcome evaluations were administered by 3 raters masked to the participant group and trained to properly administer the outcome measures.
of 13 to 91. A low score on any subscale indicates a more severe disability. Evaluation of the reliability, validity, and responsiveness of the FIM has been reported extensively.17,45–48

Data Analysis

Examination of Responsiveness

Various indicators such as the paired t test, effect size statistics, and receiver operating characteristic curve have been proposed for examining the responsiveness of a measure,49–51 but there is no consensus on the preferred method.16,17 In this study, the responsiveness of the 3 measures to change from pretreatment to posttreatment was determined with the following 2 indices. The Wilcoxon matched pairs signed rank test focuses on the statistical significance of the observed change in the measures. The standardized response mean (SRM), a type of effect size, is defined as the mean change in score divided by the standard deviation of the changed scores.16 According to Cohen criteria,27 an SRM greater than 0.8 is large, one of 0.5 to 0.8 is moderate, and one of 0.2 to 0.5 is small. In addition, the 95% confidence intervals for the SRMs were estimated using bootstrap 1000 samples with replacement.52 To test for the differences of SRMs between 2 measures, we took the differences in SRMs for 1000 paired bootstrap samples of 2 measures, then sorted the differences from lowest to highest, and then examined whether the value zero was included between the 25th and 975th observations (if the zero is not included, there is a significant difference between measures).53

Examination of Validity

We examined the construct and predictive validity of the 3 measures using bivariate correlational analyses. Construct validity refers to the ability of a measure to assess an abstract concept or construct.54 Predictive validity attempts to establish that a measure will be a valid predictor of certain future criterion measures.54 The Spearman rank correlation coefficient (ρ) was used to compare the 3 outcome measures with each other as well as to examine their relationships with the FIM-motor score at pretreatment and posttreatment, respectively. In addition, we examined whether the scores of the 3 measures at pretreatment could predict the FIM-total and the FIM-motor scores at posttreatment using the Spearman rank correlation coefficient (ρ). Correlations between 0 and 0.25 were considered low; those between 0.25 and 0.5 were considered fair; those between 0.5 and 0.75 were considered moderate to good; and those greater than 0.75 were considered good to excellent.54

Results

Table 1 lists the demographic and clinical features of the 57 participants. Their mean age was 55 years, and 68% were men. The study sample included patients with mild to moderate stroke, as assessed by their motor function and disability scores at pretreatment.

The responsiveness indices of the 3 outcome measures are shown in Table 2. The change in responsiveness of the FMA, ARAT, and WMFT-FAS was large from pretreatment to posttreatment (SRM=0.95–1.42, and Wilcoxon Z=4.64–6.33, P<0.001), whereas the WMFT-TIME was small (SRM=0.38 and Wilcoxon Z=5.97, P<0.001). Table 3 shows results of the comparisons of the SRMs among the outcome measures. The responsiveness of the FMA was significantly larger than those of the ARAT (difference in SRM=0.47, 95% confidence interval, CI: 0.09, 0.89) and the WMFT-TIME (difference in SRM=1.04, 95% CI: 0.70, 1.40), but not the WMFT-FAS (difference in SRM=0.12, 95% CI: -0.33, 0.68). In addition, the responsiveness of the WMFT-TIME was significantly smaller than those of

| Table 1. Demographics and Clinical Characteristics of the Participants (n=57) |
|---------------------------------|--------|
| Characteristics               | Value  |
| Age, y                        | 54.56±11.52 |
| Gender                        |        |
| Female                        | 18 (31.6%) |
| Male                          | 39 (68.4%) |
| Side of stroke                |        |
| Left                          | 31 (54.4%) |
| Right                         | 26 (45.6%) |
| Time since stroke, months     | 12.98±7.62 |
| Brunstrom stage of the proximal UE, median | 5 |
| Brunstrom stage of the distal UE, median | 5 |
| Mini-Mental State Examination | 28.54±1.56 |
| FMA at pretreatment           | 49.35±9.30 |
| ARAT at pretreatment          | 42.72±12.11 |
| WMFT-TIME at pretreatment, seconds | 7.05±6.85 |
| WMFT-FAS at pretreatment      | 3.22±0.72  |
| FIM at pretreatment           | 117.88±11.02 |
| FIM-motor at pretreatment     | 84.79±8.97 |

FMA indicates Fugl-Meyer Assessment; ARAT, Action Research Arm Test; WMFT-TIME, performance time of the Wolf Motor Function Test; WMFT-FAS, functional ability scale of the Wolf Motor Function Test; FIM, Functional Independence Measure.

<table>
<thead>
<tr>
<th>Table 2. Responsiveness Indices of the Three Outcome Measures</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scale</td>
</tr>
<tr>
<td>FMA</td>
</tr>
<tr>
<td>ARAT</td>
</tr>
<tr>
<td>WMFT-TIME</td>
</tr>
<tr>
<td>WMFT-FAS</td>
</tr>
<tr>
<td>FIM-total</td>
</tr>
<tr>
<td>FIM-motor</td>
</tr>
</tbody>
</table>

*P<0.001.

SRM indicates standardized response mean; CI, confidence interval; FMA, Fugl-Meyer Assessment; ARAT, Action Research Arm Test; WMFT-TIME, performance time of the Wolf Motor Function Test; WMFT-FAS, functional ability scale of the Wolf Motor Function Test; FIM, Functional Independence Measure.

<table>
<thead>
<tr>
<th>Table 3. Comparisons of the SRMs of the Three Outcome Measures</th>
</tr>
</thead>
<tbody>
<tr>
<td>Measures</td>
</tr>
<tr>
<td>FMA vs ARAT</td>
</tr>
<tr>
<td>FMA vs WMFT-TIME</td>
</tr>
<tr>
<td>FMA vs WMFT-FAS</td>
</tr>
<tr>
<td>ARAT vs WMFT-TIME</td>
</tr>
<tr>
<td>WMFT-FAS vs ARAT</td>
</tr>
<tr>
<td>WMFT-FAS vs WMFT-TIME</td>
</tr>
</tbody>
</table>

*P<0.05.

SRM indicates standardized response mean; CI, confidence interval; FMA, Fugl-Meyer Assessment; ARAT, Action Research Arm Test; WMFT-TIME, performance time of the Wolf Motor Function Test; WMFT-FAS, functional ability scale of the Wolf Motor Function Test.
Table 4. Construct Validity (Spearman ρ) of the Three Outcome Measures at Pretreatment

<table>
<thead>
<tr>
<th>Scale</th>
<th>FMA (95% CI)</th>
<th>ARAT (95% CI)</th>
<th>WMFT-TIME (95% CI)</th>
<th>WMFT-FAS (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ARAT</td>
<td>0.73** (0.58, 0.83)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WMFT-TIME</td>
<td>0.76** (0.63, 0.86)</td>
<td>0.63** (0.44, 0.77)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>WMFT-FAS</td>
<td>0.71** (0.56, 0.82)</td>
<td>0.77** (0.64, 0.87)</td>
<td>0.45** (0.22, 0.64)</td>
<td></td>
</tr>
<tr>
<td>FIM-motor</td>
<td>0.49** (0.27, 0.66)</td>
<td>0.27* (0.01, 0.50)</td>
<td>0.40** (0.16, 0.60)</td>
<td>0.29* (0.03, 0.52)</td>
</tr>
</tbody>
</table>

*P<0.05.
**P<0.01.

CI indicates confidence interval; FMA, Fugl-Meyer Assessment; ARAT, Action Research Arm Test; WMFT-TIME, performance time of the Wolf Motor Function Test; WMFT-FAS, functional ability scale of the Wolf Motor Function Test; FIM, Functional Independence Measure.

Table 5. Construct Validity (Spearman ρ) of the Three Outcome Measures at Posttreatment

<table>
<thead>
<tr>
<th>Scale</th>
<th>FMA (95% CI)</th>
<th>ARAT (95% CI)</th>
<th>WMFT-TIME (95% CI)</th>
<th>WMFT-FAS (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ARAT</td>
<td>0.74** (0.60, 0.84)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WMFT-TIME</td>
<td>0.71** (0.56, 0.82)</td>
<td>0.58** (0.38, 0.74)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>WMFT-FAS</td>
<td>0.51** (0.29, 0.68)</td>
<td>0.68** (0.51, 0.80)</td>
<td>0.25 (−0.01, 0.48)</td>
<td></td>
</tr>
<tr>
<td>FIM-motor</td>
<td>0.42** (0.18, 0.62)</td>
<td>0.39** (0.15, 0.60)</td>
<td>0.43** (0.19, 0.63)</td>
<td>0.16 (−0.11, 0.41)</td>
</tr>
</tbody>
</table>

**P<0.01.

CI indicates confidence interval; FMA, Fugl-Meyer Assessment; ARAT, Action Research Arm Test; WMFT-TIME, performance time of the Wolf Motor Function Test; WMFT-FAS, functional ability scale of the Wolf Motor Function Test; FIM, Functional Independence Measure.

Discussion

To our knowledge, this study is the first to investigate and compare the clinimetric properties of the FMA, ARAT, and the WMFT after stroke rehabilitation based on 1 sample of patients. The findings show the FMA to be a relatively sound outcome measure with good responsiveness and validity properties, as compared to the ARAT and the WMFT, for patients undergoing stroke rehabilitation intervention. The FMA displayed a large degree of responsiveness from pretreatment to posttreatment, indicating it can accurately detect patients’ motor function changes over this intervention period. With moderate predictive validity, the pretreatment score of the FMA can predict, to some extent, the patient’s functional independence level after intervention. Further, the FMA had moderate to good construct validity with other 2 outcome measures and the criterion measure at posttreatment as well as at posttreatment. These findings of sound responsiveness and validity of the FMA are consistent with those demonstrated in previous studies.

Given the growing number of clinical trials for stroke rehabilitation, it is important to identify the appropriate tools to assess the outcome of upper extremity motor function intervention. Findings of our study support the use of FMA in motor outcomes evaluation.

Although the degree of responsiveness of the ARAT and the FMA are both large, the FMA is more responsive than the ARAT in this study. This result is not consistent with the findings of van der Lee et al, in which they concluded that the ARAT is more responsive than the FMA. The differential results may be attributable to differences in the usage of responsiveness indices, intervention protocols, and time after onset of stroke of the patients (1.1 year versus 3.6 years of stroke onset). In addition, the findings of construct validity of the ARAT in this study are similar to those found in previous studies. Our study, together with previous research, indicates that the ARAT has good responsiveness and construct validity in measuring upper extremity motor function.
enable investigators to make sound prognostic decisions and when executing the tasks might be different. The WMFT-FAS was low, indicating the underlying construct validity of the WMFT. However, the correlation between the WMFT-TIME and WMFT-FAS was low, indicating the underlying constructs of speed of task completion and quality of movements when executing the tasks might be different.

An outcome measure with good predictive validity would enable investigators to make sound prognostic decisions and facilitate planning on placement after discharge. To our knowledge, there is a paucity of studies that have investigated whether the baseline scores of the 3 outcome measures predict stroke patients’ outcome in functional independence after intervention. Our preliminary findings suggest that the FMA and the WMFT-TIME have moderate predictive validity with the FIM scores, whereas the ARAT and the WMFT-FAS do not. It should, however, be noted that the results of this study may not generalize to other populations with different in patient characteristics (eg, cognitive status, severity of motor impairments, etc). Future studies should more widely investigate the impact of these potential moderating factors on outcome prediction after stroke.

Further research is warranted to investigate the predictive validity of these measures of motor function against outcomes in instrumental activities of daily livings and community reintegration at long-term follow-up (eg, 6 and 12 months after intervention) to investigate the longitudinal validity of the measures. In addition, because functional imaging and kinematic analysis have been suggested to be surrogate outcome measures for evaluating the efficacy of CIT and other rehabilitation techniques, the correlations between changes in the clinical measures of motor function and changes in functional imaging and kinematic performance need to be studied in future investigations.

When interpreting findings of this study, some potential limitations warrant considerations. First, this study was based on a modest sample size and the findings should be validated using a larger sample. Secondly, only patients with MMSE score larger than or equal to 24 were included in this study and the results may not be generalized to stroke patients with cognitive impairments. Third, this study included chronic stroke patients and sensitivity of outcome measures to change after rehabilitation may vary depending on time after stroke onset. Further research may study the clinimetric properties of these measures in patients within 6 months after onset of stroke who may have more spontaneous recovery and show different patterns of change.

### Conclusion

This study comprehensively examined and compared the responsiveness, construct validity, and predictive validity of 3 outcome measures widely used in stroke rehabilitation trials. The findings demonstrate the relative strengths and clinimetric attributes of the FMA, ARAT, and WMFT, which thus provide critical information for selection of outcome measures for stroke patients. The FMA demonstrated a large degree of responsiveness, along with good construct and predictive validity properties. In addition, the ARAT had good responsiveness and construct validity, but its predictive validity was low. The construct validity of the 2 measures of the WMFT was supported; the WMFT-FAS had a large responsiveness, and the WMFT-TIME had moderate predictive validity. Nevertheless, the responsiveness of the WMFT-FAS was small, and the predictive validity of the WMFT-FAS was low. Thus, compared to the ARAT and the WMFT, the FMA is a relatively sound outcome measure of motor function after stroke rehabilitation.

### Sources of Funding

This project was supported in part by the National Science Council (NSC-95-2314-B-002-225-MY2 and NSC-96-2628-B-002-033-MY2) and the National Health Research Institutes (NHRI-EX97-1021F) in Taiwan.

### Disclosures

None.

### References


Responsiveness and Validity of Three Outcome Measures of Motor Function After Stroke Rehabilitation
Yu-wei Hsieh, Ching-yi Wu, Keh-chung Lin, Ya-fen Chang, Chia-ling Chen and Jung-sen Liu

Stroke. 2009;40:1386-1391; originally published online February 19, 2009;
doi: 10.1161/STROKEAHA.108.530584
Stroke is published by the American Heart Association, 7272 Greenville Avenue, Dallas, TX 75231
Copyright © 2009 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/40/4/1386

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/