

Stroke

American Stroke
AssociationSM

JOURNAL OF THE AMERICAN HEART ASSOCIATION

A Division of American
Heart Association



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 Feb 19, 2009;

DOI: 10.1161/STROKEAHA.108.530584

Stroke is published by the American Heart Association, 7272 Greenville Avenue, Dallas, TX 75214

Copyright © 2009 American Heart Association. 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/cgi/content/full/40/4/1386>

Subscriptions: Information about subscribing to Stroke is online at

<http://stroke.ahajournals.org/subscriptions/>

Permissions: Permissions & Rights Desk, Lippincott Williams & Wilkins, a division of Wolters
Kluwer Health, 351 West Camden Street, Baltimore, MD 21202-2436. Phone: 410-528-4050. Fax:
410-528-8550. E-mail:

journalpermissions@lww.com

Reprints: Information about reprints can be found online at

<http://www.lww.com/reprints>

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 (ρ).

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 (ρ =0.42–0.76). The FMA and the WMFT performance time scores at pretreatment had moderate predictive validity with the FIM scores at posttreatment (ρ =0.42–0.47). In addition, the ARAT and the WMFT functional ability scores revealed a low predictive validity with the FIM (ρ =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.^{1,2} Various contemporary rehabilitation strategies which aim to improve upper extremity motor deficits after stroke based on repetitive practice, such as constraint-induced therapy (CIT)^{3,4} and bilateral arm training (BAT),^{5,6} have come into wide use. CIT first proposed by Taub et al³ 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^{7,8} and its variants.^{9–11} BAT is an alternative treatment approach.^{5,12} The unaffected UE performs the spatiotemporal pattern of movement as that performed by the affected UE.¹³ Practice of the bilateral movement allows the activation of the intact hemisphere to facilitate the activation of the damaged hemisphere^{14,15} 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.¹⁶ 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.¹⁷ However, responsiveness is essential for assessing the effectiveness of treatment as well as measuring longitudinal change over time.^{18,19} Responsiveness is defined as the ability of a measure to detect changes which have occurred accurately over time.²⁰ There has been an increasing emphasis on the importance of investigating the responsiveness of a measure,^{17,21–23} but there is no consensus on the advantage of one method over another, and multiple methods have been suggested.^{16,20}

Received July 3, 2008; final revision received September 7, 2008; accepted October 1, 2008.

From the School of Occupational Therapy, College of Medicine (Y.W.H., K.C.L.), National Taiwan University, Taipei, Taiwan; the Division of Occupational Therapy, Department of Physical Medicine and Rehabilitation (K.C.L.), National Taiwan University Hospital, Taipei, Taiwan; the Department of Occupational Therapy and Graduate Institute of Clinical Behavioral Science (C.Y.W.), Chang Gung University, Taoyuan, Taiwan; the Division of Occupational Therapy, Department of Rehabilitation (Y.F.C.), Shin Kong Wu Ho-Su Memorial Hospital, Taipei, Taiwan; the Department of Physical Medicine and Rehabilitation (C.L.C.), Chang Gung Memorial Hospital, Taoyuan, Taiwan; and the Departments of Surgery and Medical Education (J.S.L.), Cathay General Hospital, Taipei, Taiwan.

Correspondence to Keh-chung Lin, ScD, OTR, School of Occupational Therapy, College of Medicine, National Taiwan University and Division of Occupational Therapy, Department of Physical Medicine and Rehabilitation, National Taiwan University Hospital, 17, F4, Xu-Zhou Road, Taipei, Taiwan 100. E-mail kehchunglin@ntu.edu.tw

© 2009 American Heart Association, Inc.

Stroke is available at <http://stroke.ahajournals.org>

DOI: 10.1161/STROKEAHA.108.530584

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.^{2,10,11,24,25} 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.²⁶ Another concern is the lack of a study that directly compares the responsiveness and validity of the 3 measures in stroke patients undergoing rehabilitative therapies. Therefore, we aimed to examine and compare the responsiveness and validity properties of the FMA, ARAT, and WMFT in a stroke sample cohort that had received rehabilitation interventions.

Methods

Participants

All subjects in this study were enrolled in a randomized controlled trial which aims to investigate the effectiveness of distributed constraint-induced therapy and bilateral arm training. In the trial, outcome measures included the FMA, ARAT, and WMFT for evaluating motor performance, and the FIM for assessing functional outcome. A cohort of 57 patients with no missing values for any of the 4 measures was included in the present study. Subjects were recruited from 3 medical centers. The inclusion criteria were: (1) a first-ever stroke onset at least 6 months previously; (2) demonstration of Brunnstrom stage III for the proximal UE or above (Supplemental Table I, available online at <http://stroke.ahajournals.org>); and (3) no excessive spasticity in the joints (shoulder and elbow) of the affected UE (Modified Ashworth Scale score ≤ 2.5 in each joint; Supplemental Table II). To avoid the confounding effects of cognitive and medical conditions, subjects were excluded if the medical or physical screening examination resulted in a score of less than 24 on the Mini-Mental State Examination, physician-determined major medical problems or poor physical conditions that would interfere with participation, or excessive pain in any joint that might limit participation. Institutional review board approval was obtained from the study sites, and written informed consent was obtained from each patient before inclusion.

The sample size for the original trial was determined based on a priori power analysis.²⁷ At least 20 subjects for each group would be required to detect an effect size d of 0.80 given a significance level of 5% (1-tailed) and 80% power. This effect size was estimated based on the findings of 2 previous trials^{9,11} on the distributed CIT in which the effect size d was 1.39 and 0.75 on the FMA, respectively.

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 occupational therapist informed of the group allocation.

Patients received 1 of 3 treatments for a 3-week training period. The distributed CIT group involved restriction on movement of the unaffected hand by placement in a mitt for a target of 6 hours per day and intensive training of the affected UE in functional tasks for 2 hours per weekday, including reaching forward or upward to move a cup, picking up coins, taking-up an utensil to take food, and other functional movements similar to daily activities needed. The BAT group concentrated on both affected and unaffected UE moving simultaneously in functional tasks by the symmetrical or alternated patterns for 2 hours per weekday. The functional tasks also emphasized UE movements involved in daily activities, but focused on both UEs

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.

Outcome Measures

Fugl-Meyer Assessment

The upper-extremity subscale of the Fugl-Meyer Assessment (FMA) was used to assess motor impairment.²⁸ The 33 UE items measure the movement and reflexes of the shoulder/elbow/forearm, wrist, hand, and coordination/speed. They are scored on a 3-point ordinal scale (0-cannot perform, 1-performs partially, 2-performs fully).²⁸ The maximum score is 66, indicating optimal recovery. The psychometric properties of the FMA have been shown to be satisfactory in stroke patients.^{28–30}

Action Research Arm Test

The Action Research Arm Test (ARAT) was designed for evaluation of upper extremity function.³¹ It consists of 19 items (maximum score 57) 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.^{32–34}

Wolf Motor Function Test

The Wolf Motor Function Test (WMFT) was originally developed to assess the impact of forced use on upper extremity function.³⁵ The WMFT contains 17 tasks (15 function-based and 2 strength-based).³⁶ The performance time (WMFT-TIME) and functional ability scale (WMFT-FAS) of the 15 function-based items were administered in this study. The maximum time allowed to complete an item was 120 seconds. For functional ability scoring, we used a 6-point ordinal scale where 0 indicates “does not attempt with the involved arm” and 5 indicates “arm does participate; movement appears to be normal.” The test-retest reliability, interrater reliability, criterion validity, and construct validity of the WMFT has been ascertained in stroke patients.^{36–39}

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.^{40,41} Compared to other functional assessments, the FIM revealed preferable or similar properties in rehabilitation patients.^{42,43} 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 ability.⁴⁴ 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

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
Brunnstrom stage of the proximal UE, median	5
Brunnstrom 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.

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

Table 2. Responsiveness Indices of the Three Outcome Measures

Scale	SRM (95% CI)	The Wilcoxon Test Z-Value
FMA	1.42 (1.19, 1.80)	6.33*
ARAT	0.95 (0.75, 1.20)	4.64*
WMFT-TIME	0.38 (0.22, 0.59)	5.97*
WMFT-FAS	1.30 (1.03, 1.67)	5.59*
FIM-total	0.36 (0.17, 0.59)	3.39*
FIM-motor	0.37 (0.17, 0.58)	3.18*

**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.

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 3. Comparisons of the SRMs of the Three Outcome Measures

Measures	Difference in SRM (95% CI)
FMA vs ARAT	0.47 (0.09, 0.89)*
FMA vs WMFT-TIME	1.04 (0.70, 1.40)*
FMA vs WMFT-FAS	0.12 (-0.33, 0.68)
ARAT vs WMFT-TIME	0.57 (0.28, 0.86)*
WMFT-FAS vs ARAT	0.35 (-0.01, 0.78)
WMFT-FAS vs WMFT-TIME	0.92 (0.57, 1.30)*

**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

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

* $P < 0.05$.** $P < 0.01$.

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.

the ARAT (difference in SRM=0.57, 95% CI: 0.28, 0.86) and the WMFT-FAS (difference in SRM=0.92, 95% CI: 0.57, 1.30).

Table 4 shows construct validity of the 3 outcome measures at pretreatment. Correlations between the FMA and other outcome measures were relatively high (Spearman ρ =0.71–0.76). The ARAT had moderate to good correlations with the other outcome measures (Spearman ρ =0.63–0.77). Both the WMFT-TIME and the WMFT-FAS had moderate to good correlations with the FMA and the ARAT (Spearman ρ =0.63–0.77). However, the correlation between the WMFT-TIME and WMFT-FAS was relatively low (Spearman ρ =0.45). In addition, on comparing to the criterion measure at pretreatment, the FMA and the WMFT-TIME had a moderate correlation with the FIM-motor test, whereas the ARAT and the WMFT-FAS had low correlations.

Table 5 shows the construct validity results at posttreatment. The FMA and the ARAT had moderate to good correlations with the other outcome measures (FMA: Spearman ρ =0.51–0.74, ARAT: Spearman ρ =0.58–0.74). Both of the WMFT-TIME and the WMFT-FAS had moderate to good correlations with the FMA and the ARAT (Spearman ρ =0.51–0.71), whereas the relationship between the 2 measures of the WMFT was low (Spearman ρ =0.25). When comparing to the criterion measure at posttreatment, the FMA, WMFT-TIME, and ARAT had moderate correlation with the FIM-motor test, whereas the WMFT-FAS had a low correlation.

Table 6 shows that the FMA and the WMFT-TIME at pretreatment had moderate predictive validity with the FIM-total and FIM-motor scores at posttreatment (Spearman ρ =0.42–0.47, $P < 0.01$). In addition, the ARAT and the WMFT-FAS exhibited low predictive validity with the FIM-total and FIM-motor scores (Spearman ρ =0.17–0.26).

Discussion

To our knowledge, this study is the first to investigate and compare the clinimetric properties of the FMA, ARAT, and

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 pretreatment as well as at posttreatment. These findings of sound responsiveness and validity of the FMA are consistent with those demonstrated in previous studies.^{30,55,56} 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.²² 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.^{33,56} Our study, together with previous research, indicates that the ARAT has good responsiveness and construct validity in measuring upper extremity motor function.

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

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

** $P < 0.01$.

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 6. Relationships Between the Three Outcome Measures at Pretreatment and the FIM-Total and FIM-Motor Score at Posttreatment

Scale	Predictive Validity for FIM-Total Score (Spearman ρ) (95% CI)	Predictive Validity for FIM-Motor Score (Spearman ρ) (95% CI)
FMA	0.42** (0.18, 0.62)	0.42** (0.18, 0.62)
ARAT	0.22 (−0.04, 0.46)	0.26 (0.00, 0.49)
WMFT-TIME	0.47** (0.24, 0.66)	0.43** (0.19, 0.63)
WMFT-FAS	0.17 (−0.10, 0.42)	0.19 (−0.07, 0.43)

** $P < 0.01$.

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.

The responsiveness of the WMFT-FAS was larger than that of the WMFT-TIME in this study. This result is contrary to the findings of Caimmi et al,⁵⁷ who concluded that responsiveness measured by the effect size d was moderate for the WMFT-TIME and trivial for the WMFT-FAS. Although different indices and different treatment protocols were used in these 2 studies, the larger variance of the WMFT-TIME among our patients may have contributed to our results. In addition, the correlations of the 2 measures of the WMFT with the FMA were moderate to good, which is concordant with the previous findings^{38,39} and thus supports the construct validity of the WMFT. However, the correlation between the WMFT-TIME and the 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.⁵⁴ 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,^{57,58} 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-TIME 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-EX97-9742PI) in Taiwan.

Disclosures

None.

References

- Gowland C, deBruin H, Basmajian JV, Plews N, Burcea I. Agonist and antagonist activity during voluntary upper-limb movement in patients with stroke. *Phys Ther*. 1992;72:624–633.
- van der Lee JH, Wagenaar RC, Lankhorst GJ, Vogelaar TW, Deville WL, Bouter LM. Forced use of the upper extremity in chronic stroke patients: results from a single-blind randomized clinical trial. *Stroke*. 1999;30:2369–2375.
- Taub E, Miller NE, Novack TA, Cook EW, III, Fleming WC, Nepomuceno CS, Connell JS, Crago JE. Technique to improve chronic motor deficit after stroke. *Arch Phys Med Rehabil*. 1993;74:347–354.
- Taub E, Uswatte G, Pidikiti R. Constraint-induced movement therapy: a new family of techniques with broad application to physical rehabilitation—a clinical review. *J Rehabil Res Dev*. 1999;36:237–251.
- Cauraugh JH, Kim SB, Duley A. Coupled bilateral movements and active neuromuscular stimulation: intralimb transfer evidence during bimanual aiming. *Neurosci Lett*. 2005;382:39–44.
- McCombe Waller S, Whittall J. Fine motor control in adults with and without chronic hemiparesis: baseline comparison to nondisabled adults and effects of bilateral arm training. *Arch Phys Med Rehabil*. 2004;85:1076–1083.
- Taub E, Uswatte G, Morris DM. Improved motor recovery after stroke and massive cortical reorganization following constraint-induced movement therapy. *Phys Med Rehabil Clin N Am*. 2003;14:S77–91, ix.
- Wolf SL, Winstein CJ, Miller JP, Taub E, Uswatte G, Morris D, Giuliani C, Light KE, Nichols-Larsen D. Effect of constraint-induced movement therapy on upper extremity function 3 to 9 months after stroke: The EXCITE randomized clinical trial. *JAMA*. 2006;296:2095–2104.

9. Lin KC, Wu CY, Liu JS, Chen YT, Hsu CJ. Constraint-induced therapy versus dose-matched control intervention to improve motor ability, basic/extended daily functions, and quality of life in stroke. *Neurorehabil Neural Repair*. In press.
10. Page SJ, Levine P, Leonard AC. Modified constraint-induced therapy in acute stroke: a randomized controlled pilot study. *Neurorehabil Neural Repair*. 2005;19:27–32.
11. Wu CY, Chen CL, Tang SF, Lin KC, Huang YY. Kinematic and clinical analyses of upper-extremity movements after constraint-induced movement therapy in patients with stroke: a randomized controlled trial. *Arch Phys Med Rehabil*. 2007;88:964–970.
12. Rose DK, Winstein CJ. Bimanual training after stroke: are two hands better than one? *Top Stroke Rehabil*. 2004;11:20–30.
13. Summers JJ, Kagerer FA, Garry MI, Hiraga CY, Loftus A, Cauraugh JH. Bilateral and unilateral movement training on upper limb function in chronic stroke patients: a TMS study. *J Neurol Sci*. 2007;252:76–82.
14. Cauraugh JH, Summers JJ. Neural plasticity and bilateral movements: A rehabilitation approach for chronic stroke. *Prog Neurobiol*. 2005;75:309–320.
15. Goble DJ. The potential for utilizing inter-limb coupling in the rehabilitation of upper limb motor disability due to unilateral brain injury. *Disabil Rehabil*. 2006;28:1103–1108.
16. Husted JA, Cook RJ, Farewell VT, Gladman DD. Methods for assessing responsiveness: a critical review and recommendations. *J Clin Epidemiol*. 2000;53:459–468.
17. Wallace D, Duncan PW, Lai SM. Comparison of the responsiveness of the Barthel index and the motor component of the Functional Independence Measure in stroke: the impact of using different methods for measuring responsiveness. *J Clin Epidemiol*. 2002;55:922–928.
18. Guyatt GH, Deyo RA, Charlson M, Levine MN, Mitchell A. Responsiveness and validity in health status measurement: a clarification. *J Clin Epidemiol*. 1989;42:403–408.
19. Kirshner B, Guyatt G. A methodological framework for assessing health indices. *J Chronic Dis*. 1985;38:27–36.
20. Wright JG, Young NL. A comparison of different indices of responsiveness. *J Clin Epidemiol*. 1997;50:239–246.
21. Hsueh IP, Hsieh C-L. Responsiveness of two upper extremity function instruments for stroke inpatients receiving rehabilitation. *Clinical Rehabilitation*. 2002;16:617–624.
22. van der Lee JH, Beckerman H, Lankhorst GJ, Bouter LM. The responsiveness of the Action Research Arm Test and the Fugl-Meyer Assessment scale in chronic stroke patients. *J Rehabil Med*. 2001;33:110–113.
23. Stratford PW, Binkley FM, Riddle DL. Health status measures: strategies and analytic methods for assessing change scores. *Phys Ther*. 1996;76:1109–1123.
24. Duncan P, Studenski S, Richards L, Gollub S, Lai SM, Reker D, Perera S, Yates J, Koch V, Rigler S, Johnson D. Randomized clinical trial of therapeutic exercise in subacute stroke. *Stroke*. 2003;34:2173–2180.
25. Whittall J, McCombe Waller S, Silver KH, Macko RF. Repetitive bilateral arm training with rhythmic auditory cueing improves motor function in chronic hemiparetic stroke. *Stroke*. 2000;31:2390–2395.
26. Bonaiuto D, Rebasti L, Sioli P. The constraint-induced movement therapy: a systematic review of randomised controlled trials on the adult stroke patients. *Eura Medicophys*. 2007;43:139–146.
27. Cohen J. *Statistical power analysis for the behavior sciences*. II ed. Hillsdale, NJ: Lawrence Erlbaum Associates; 1988.
28. Fugl-Meyer AR, Jaasko L, Leyman I, Olsson S, Stegling S. The post-stroke hemiplegic patient. 1. A method for evaluation of physical performance. *Scand J Rehabil Med*. 1975;7:13–31.
29. Gladstone DJ, Danells CJ, Black SE. The Fugl-Meyer assessment of motor recovery after stroke: A critical review of its measurement properties. *Neurorehabil Neural Repair*. 2002;16:232–240.
30. Platz T, Pinkowski C, van Wijck F, Kim IH, di Bella P, Johnson G. Reliability and validity of arm function assessment with standardized guidelines for the Fugl-Meyer Test, Action Research Arm Test and Box and Block Test: a multicentre study. *Clin Rehabil*. 2005;19:404–411.
31. Lyle RC. A performance test for assessment of upper limb function in physical rehabilitation treatment and research. *Int J Rehabil Res*. 1981;4:483–492.
32. Hsieh CL, Hsueh IP, Chiang FM, Lin PH. Inter-rater reliability and validity of the action research arm test in stroke patients. *Age Ageing*. 1998;27:107–113.
33. Lang CE, Wagner JM, Dromerick AW, Edwards DF. Measurement of upper-extremity function early after stroke: properties of the action research arm test. *Arch Phys Med Rehabil*. 2006;87:1605–1610.
34. van der Lee JH, De Groot V, Beckerman H, Wagenaar RC, Lankhorst GJ, Bouter LM. The intra- and interrater reliability of the action research arm test: a practical test of upper extremity function in patients with stroke. *Arch Phys Med Rehabil*. 2001;82:14–19.
35. Wolf SL, Lecraw DE, Barton LA, Jann BB. Forced use of hemiplegic upper extremities to reverse the effect of learned nonuse among chronic stroke and head-injured patients. *Exp Neurol*. 1989;104:125–132.
36. Wolf SL, Thompson PA, Morris DM, Rose DK, Winstein CJ, Taub E, Giuliani C, Pearson SL. The EXCITE trial: attributes of the Wolf Motor Function Test in patients with subacute stroke. *Neurorehabil Neural Repair*. 2005;19:194–205.
37. Morris DM, Uswatte G, Crago JE, Cook EW, III, Taub E. The reliability of the wolf motor function test for assessing upper extremity function after stroke. *Arch Phys Med Rehabil*. 2001;82:750–755.
38. Whittall J, Savin DN Jr, Harris-Love M, Waller SM. Psychometric properties of a modified Wolf Motor Function test for people with mild and moderate upper-extremity hemiparesis. *Arch Phys Med Rehabil*. 2006;87:656–660.
39. Wolf SL, Catlin PA, Ellis M, Archer AL, Morgan B, Piacentino A. Assessing Wolf motor function test as outcome measure for research in patients after stroke. *Stroke*. 2001;32:1635–1639.
40. Granger CV. The emerging science of functional assessment: Our tool for outcomes analysis. *Arch Phys Med Rehabil*. 1998;79:235–240.
41. Stucki G, Ewert T, Cieza A. Value and application of the ICF in rehabilitation medicine. *Disabil Rehabil*. 2002;24:932–938.
42. Cohen ME, Marino RJ. The tools of disability outcomes research functional status measures. *Arch Phys Med Rehabil*. 2000;81:S21–S29.
43. Hsueh IP, Lin JH, Jeng JS, Hsieh CL. Comparison of the psychometric characteristics of the functional independence measure, 5 item Barthel index, and 10 item Barthel index in patients with stroke. *J Neurol Neurosurg Psychiatry*. 2002;73:188–190.
44. Hamilton BB, Granger CV, Sherwin FS, Zielesny M, Tashman JS. A uniform national data system for medical rehabilitation. In: Fuhrer MJ, ed. *Rehabilitation outcomes: Analysis and measurement*. Baltimore: Paul H. Brookes Publishing Company; 1987. p. 137–147.
45. Chau N, Daler S, Andre JM, Patris A. Inter-rater agreement of two functional independence scales: the Functional Independence Measure (FIM) and a subjective uniform continuous scale. *Disabil Rehabil*. 1994;16:63–71.
46. Dodds TA, Martin DP, Stolov WC, Deyo RA. A validation of the functional independence measurement and its performance among rehabilitation inpatients. *Arch Phys Med Rehabil*. 1993;74:531–536.
47. Gosman-Hedstrom G, Svensson E. Parallel reliability of the functional independence measure and the Barthel ADL index. *Disabil Rehabil*. 2000;22:702–715.
48. Hamilton BB, Laughlin JA, Fiedler RC, Granger CV. Interrater reliability of the 7-level functional independence measure (FIM). *Scand J Rehabil Med*. 1994;26:115–119.
49. Deyo RA, Centor RM. Assessing the responsiveness of functional scales to clinical change: an analogy to diagnostic test performance. *J Chronic Dis*. 1986;39:897–906.
50. Kazis LE, Anderson JJ, Meenan RF. Effect sizes for interpreting changes in health status. *Med Care*. 1989;27:S178–189.
51. Liang MH, Fossel AH, Larson MG. Comparisons of five health status instruments for orthopedic evaluation. *Med Care*. 1990;28:632–642.
52. Efron B, Tibshirani RJ. *An Introduction to the Bootstrap*. New York: Chapman and Hall; 1993.
53. Stratford PW, Kennedy DM. Does parallel item content on WOMAC's pain and function subscales limit its ability to detect change in functional status? *BMC Musculoskelet Disord*. 2004;5:17.
54. Portney LG, Watkins MP. *Foundations of Clinical Research: Applications to Practice*. III ed. Upper Saddle River, NJ: Pearson/Prentice Hall; 2009.
55. Chae J, Johnston M, Kim H, Zorowitz R. Admission motor impairment as a predictor of physical disability after stroke rehabilitation. *Am J Phys Med Rehabil*. 1995;74:218–223.
56. Rabadi MH, Rabadi FM. Comparison of the action research arm test and the Fugl-Meyer assessment as measures of upper-extremity motor weakness after stroke. *Arch Phys Med Rehabil*. 2006;87:962–966.
57. Caimmi M, Carda S, Giovannana C, Maini ES, Sabatini AM, Smania N, Molteni F. Using kinematic analysis to evaluate constraint-induced movement therapy in chronic stroke patients. *Neurorehabil Neural Repair*. 2008;22:31–39.
58. Grotta JC, Noser EA, Ro T, Boake C, Levin H, Aronowski J, Schallert T. Constraint-induced movement therapy. *Stroke*. 2004;35:2699–2701.