Assessing Wolf Motor Function Test as Outcome Measure for Research in Patients After Stroke

Steven L. Wolf, PhD, PT; Pamela A. Catlin, EdD, PT; Michael Ellis, MPT; Audrey Link Archer, MPT;
Bryn Morgan, MPT; Aimee Piacentino, MPT

Background and Purpose—The Wolf Motor Function Test (WMFT) is a new time-based method to evaluate upper extremity performance while providing insight into joint-specific and total limb movements. This study addresses selected psychometric attributes of the WMFT applied to a chronic stroke population.

Methods—Nineteen individuals after stroke and with intact cognition and sitting balance were age- and sex-matched with 19 individuals without impairment. Subjects performed the WMFT and the upper extremity portion of the Fugl-Meyer Motor Assessment (FMA) on 2 occasions (12 to 16 days apart), with scoring performed independently by 2 random raters.

Results—The WMFT and FMA demonstrated agreement ($P < 0.0001$) between raters at each session. WMFT scores for the dominant and nondominant extremities of individuals without impairment were different ($P \leq 0.05$) from the more and less affected extremities of subjects after stroke. The FMA score for the more affected extremity of subjects after stroke was different ($P \leq 0.05$) from the dominant and nondominant extremities. However, the FMA score for the less affected upper extremity of individuals after stroke was not different ($P > 0.05$) from the dominant and nondominant extremities of individuals without impairment. The WMFT and FMA scores were related ($P < 0.02$) for the more affected extremity in individuals after stroke.

Conclusions—The interrater reliability, construct validity, and criterion validity of the WMFT, as used in these subject samples, are supported. (*Stroke*. 2001;32:1635-1639.)

Key Words: arm ■ motor activity ■ psychometrics ■ stroke

Many upper extremity motor function outcome measures do not produce data that provide obvious links between the basis for planning treatment and the emergent plan for functional restitution. The Wolf Motor Function Test (WMFT) quantifies upper extremity movement ability through timed single- or multiple-joint motions and functional tasks.[1] The tasks are arranged in order of complexity, progress from proximal to distal joint involvement, test total extremity movement and movement speed, and require few tools and minimal training for test execution.

The present study establishes the reliability and validity of the WMFT. The scores from the WMFT and the upper extremity portion of the Fugl-Meyer Motor Assessment (FMA) were compared to investigate the criterion validity of the WMFT. The FMA was chosen as the criterion test because it focuses on multijoint upper extremity function in patients after stroke and is reliable[2] and valid.[3,4] Yet the FMA is difficult to use and examines synergy patterns that no longer form the basis for many functionally oriented treatments.

Subjects and Methods

Subjects

Forty-seven subjects were recruited by convenience sampling in this repeated-measures design study. Twenty-one subjects had sustained a stroke. All subjects participating were between the ages of 42 and 76 years. Nineteen subjects after stroke (mean age 61.4 ± 9.5 years, mean time from stroke 4.9 ± 6.4 years, range 0.67 to 29 years) and 19 individuals without impairments (mean age 60.0 ± 9.3 years) could be matched by age and sex. All subjects met predetermined inclusion criteria indicative of medical stability, sitting and standing balance control, and intact cognition. Subjects neither took medications affecting motor performance, nor did they engage in strenuous activity before data collection sessions. All subjects could actively extend wrist, thumb, and at least 2 other digits $\geq 10°$ and signed an informed consent form approved by the Emory University Institutional Review Board. A signed consent also was obtained from the attending physician for subjects after stroke.

Measurements

Wolf Motor Function Test

Tasks 1 to 6 of the WMFT[1] (see General Description of the WMFT in the Appendix) involve timed joint-segment movements, and tasks 7 to 15 consisted of timed integrative functional movements.

Fugl-Meyer Motor Assessment

The FMA[5] (see FMA: Upper Extremity Portion in the Appendix) assesses voluntary movement, reflex activity, grasp, and coordination. Performance is measured on 33 tasks with a 3-point ordinal scale (0 to 2), with a maximum score of 66.

TABLE 1.   Total Scores for WMFT and FMA

Variable	After Stroke		Without Impairment	
	Mean±SD	Min–Max	Mean±SD	Min–Max
WMFT, s				
More affected/dominant				
Session 1				
Rater 1	9.39±16.87	2.23–77.29	1.63±0.26	1.27–2.15
Rater 2	9.37±16.90	2.20–77.37	1.61±0.28	1.21–2.20
Session 2				
Rater 1	9.23±16.03*	2.15–73.00	1.57±0.32	1.09–2.14
Rater 2	9.20±16.04*	2.05–72.98	1.54±0.32	1.01–2.16
Less affected/nondominant				
Session 1				
Rater 1	2.55±1.85	1.67–10.09	1.68±0.37	1.15–2.48
Rater 2	2.53±1.85	1.65–10.05	1.66±0.37	1.14–2.40
Session 2				
Rater 1	2.51±1.78	1.60–9.75	1.58±0.31*	1.15–2.20
Rater 2	2.50±1.79	1.46–9.75	1.54±0.31*	1.14–2.28
FMA				
More affected/dominant				
Session 1				
Rater 1	51.1±9.9	22–63	65.5±1.1	63–66
Rater 2	50.7±10.6	21–65	65.6±0.8	64–66
Session 2				
Rater 1	51.2±11.7	23–65	65.2±1.1	63–66
Rater 2	51.7±10.8	24–66	65.5±0.7	64–66
Less affected/nondominant				
Session 1				
Rater 1	64.8±1.7	61–66	65.5±1.0	63–66
Rater 2	64.8±1.5	61–66	65.8±0.6	64–66
Session 2				
Rater 1	64.5±2.5	58–66	65.5±0.9	63–66
Rater 2	64.5±1.9	61–66	65.4±0.9	64–66

Values are mean±SD, minimum (Min), and maximum (Max) total scores across group (n=19 per group), extremity, session, and rater.

*Wilcoxon paired sample signed rank test indicated a difference ($P < 0.05$) between rater 1 and 2.

Training Raters

Rater training was completed for both tests by using a separate sample of 4 subjects (2 after stroke and 2 without stroke). Training concluded when all 4 raters scored, independently and concurrently, all tasks among all subjects within 0.20 seconds (WMFT) and with exact agreement (FMA).

Procedure

All instruments were calibrated before data collection and on every fourth subject. For each subject, testing sequence and rater pair, from among 4 raters, were 12 to 16 days apart.

Data Analysis

Nonparametric analyses were used for all data not normally distributed, on the basis of Shapiro-Wilk test results. Interrater reliability of the WMFT and FMA total scores per limb per session was determined by Intraclass Correlation Coefficient (ICC), model (1,1). Interrater reliability also was determined for WMFT (ICC) and FMA (κ statistic) individual tasks of the affected limb in subjects after stroke. Rater total scores for each test were compared by the

Wilcoxon signed rank (paired sample) test. Internal consistency of each test was determined by Chronbach's α. Each WMFT and FMA total score was compared between groups by the Wilcoxon 2-sample test. The WMFT and FMA total scores for the most affected poststroke limb were related by using the Spearman rank correlation coefficient. Only primary examiners were used in analyses, except for the reliability tests. For all analyses, the criterion α level was 0.05, and power was ≥0.90 for WMFT scores (effect size 1.22,[1] n=19) and for FMA scores (effect size 0.94,[2] n=19).

Results

Total scores for the WMFT and FMA are presented in Table 1. Interrater reliability for the WMFT ranged from 0.97 to 0.99. Reliability for the more affected extremity of subjects after stroke for the FMA was ICC 0.96 ($P < 0.0001$). A ceiling effect was observed in FMA scores for the less affected extremity of subjects after stroke and for both extremities of subjects without impairment, prohibiting interrater reliability

TABLE 2. Total Number and Percentages of FMA Tasks Partitioned Into 3 κ Ranges Across Sessions and Within Task Categories for FMA Scores of Poststroke Subjects (n=19)

Task Category	κ Level 1.0–0.7		κ Level 0.69–0.4		κ Level <0.4	
	Session 1 (12 Tasks)	Session 2 (12 Tasks)	Session 1 (14 Tasks)	Session 2 (12 Tasks)	Session 1 (7 Tasks)	Session 2 (9 Tasks)
Grasp (n=5)	4 (80%)	3 (60%)	1 (20%)	2 (40%)	0 (00%)	0 (00%)
Reflex (n=3)	2 (67%)	0 (00%)	0 (00%)	0 (00%)	1 (33%)	3 (100%)
Coordination (n=3)	1 (33%)	1 (33%)	1 (33%)	2 (67%)	1 (33%)	0 (00%)
Synergy (n=22)	5 (23%)	8 (36%)	12 (55%)	8 (36%)	5 (23%)	6 (27%)

n equals total tasks for session 1 (3 κ levels combined) and session 2 (3 κ levels combined). Values are total number (percentages) of FMA tasks.

(ICC) calculations. The WMFT scores were not different ($P>0.05$) between raters, except in the more affected extremity of subjects after stroke and in the nondominant extremity of subjects without impairment, both at session 2 (Table 1). These differences are of minimal clinical importance (<0.04 seconds). The FMA scores were not different ($P>0.05$) between raters. Interrater reliability for each of the 15 WMFT tasks was 0.95 to 0.99 ($P<0.0001$). Interrater reliability for 24% of the 33 FMA tasks had κ values <0.40 (Table 2). Internal consistency for the WMFT and FMA were 92.4% and 90.7%, respectively.

The WMFT scores were different ($P=0.0006$) between groups for all upper extremity comparisons (Table 3). The FMA scores were different ($P=0.0001$) only between the more affected extremity of subjects after stroke and each extremity of individuals without impairment.

WMFT and FMA scores were related for the more affected arm of subjects after stroke for both raters at session 1 ($r_s=-0.57$ [$P=0.0115$] and -0.54 [$P=0.02$], respectively) and at session 2 ($r_s=-0.67$ [$P=0.0015$] and -0.68 [$P=0.001$], respectively).

Discussion

Interrater reliability was supported for the WMFT within both subject groups for both sessions. Similar results were reported (E. Taub, oral communication, December 1999) for the interrater reliability of the WMFT ($r=0.97$, $P<0.05$) in subjects after stroke who met similar upper extremity movement criteria. Although a ceiling effect occurred in FMA scores for the less affected arm of subjects after stroke (score range 58 to 66) and in both arms of the subjects without impairments (score range 63 to 66), no difference in scores between raters was observed, supporting interrater reliability (Table 1). For individual tasks, interrater reliability was observed for both tests, with the WMFT demonstrating agreement between raters on all 15 tasks. Although the FMA demonstrated internal consistency (90.7%) and a relationship ($\kappa >0.40$) between raters on a majority of the tasks (Table 2), tasks 3, 11, and 14 (FMA: Upper Extremity Portion in Appendix) had low agreement ($\kappa <0.40$) at both sessions. Likewise, the reflex elicitation task (task 18) had low agreement at the second visit and was attributable to 1 rater, an observation also reported by Duncan et al.[6] The low interrater agreement for the 3 FMA tasks may be due to the 3-point

ordinal scale. The raters often interpreted and scored a "partial" or "faultless" movement differently. Lindmark and Hamrin[7] changed the FMA to a 4-point scale in the BL Motor Assessment (BLMA). Benaim et al[8] subsequently revised the BLMA because some of the FMA-based tasks were not functionally appropriate.

The WMFT scores appeared to differentiate the more affected extremity and the less affected extremity from either extremity of subjects without impairment (Table 3). Findings also support previous observations in that the FMA scores are different between the more affected extremity of individuals after stroke and either extremity of individuals without impairment (Table 3).[3,9] However, deficits are also present in the less affected extremity in individuals after stroke. Sunderland et al[10] attributed impaired function in the less affected extremity to deficits affecting perception and control of action. But the FMA scores did not differentiate changes that may have occurred in the less affected extremity of subjects after stroke (Table 3). The lack of sensitivity of the FMA may be due to use of a 3-point scale versus use of performance time for the WMFT. The WMFT scores of increased performance time in the less impaired extremity among patients with cognitive deficits require further study. However, the difference in WMFT scores between groups supports test construct validity.

The relationship between the tests' scores for the more affected extremity of subjects after stroke supports criterion validity and is consistent with values reported elsewhere.[4,11] These validity findings for the WMFT might encourage its use by clinicians and researchers to quantify upper extremity performance in individuals after stroke with motor characteristics similar to subjects in the present study. Additionally, inferences from WMFT scores may be made pertaining to the patient's level of function and potential motor recovery, because the WMFT is based on examining the time for completion of single joint or interjoint movements that frequently are engaged to either assess existing capabilities or to plan treatment for functional activities. Minimally, the test can show if interventions improve motor performance attributes, such as speed to complete tasks. Feys et al[12] recently demonstrated that motor performance may be the greatest predictor of motor recovery in individuals after stroke and that it, along with overall disability, predicted motor recovery 2, 6, and 12 months after stroke. Hence, further thought

TABLE 3. Wilcoxon 2-Sample Test Values Comparing Test Scores for WMFT and FMA Between Subjects Without Impairment and Poststroke Subjects (n=38) per Upper Extremity per Session

Test/Arm/Session	z	P
WMFT		
More affected vs dominant		
Session 1	5.26	0.0001
Session 2	5.26	0.0001
More affected vs nondominant		
Session 1	−5.20	0.0001
Session 2	−5.17	0.0001
Less affected vs dominant		
Session 1	−4.20	0.0002
Session 2	−4.15	0.0002
Less affected vs nondominant		
Session 1	3.45	0.0006
Session 2	4.06	0.0001
FMA		
More affected vs dominant		
Session 1	−5.37	0.0001
Session 2	−4.74	0.0001
More affected vs nondominant		
Session 1	5.40	0.0001
Session 2	5.01	0.0001
Less affected vs dominant		
Session 1	1.44	0.1506
Session 2	0.55	0.5815
Less affected vs nondominant		
Session 1	−1.53	0.1270
Session 2	−1.41	0.1598

should be given to delineation and development of quantitative performance-based functional tests and measures, such as used in the present study.

Appendix

General Description of the WMFT

All tasks are performed as quickly as possible and are truncated at 120 seconds. Tasks are as follows:

1. Forearm to table (side): Subject attempts to place forearm on the table by abduction at the shoulder.
2. Forearm to box (side): Subject attempts to place a forearm on the box by abduction at the shoulder.
3. Extend elbow (side): Subject attempts to reach across the table by extending the elbow (to the side).
4. Extend elbow (to the side), with weight: Subject attempts to push the sandbag against outer wrist joint across the table by extending the elbow.
5. Hand to table (front): Subject attempts to place involved hand on the table.
6. Hand to box (front): Subject attempts to place hand on the box.
7. Reach and retrieve (front): Subject attempts to pull 1-lb weight across the table by using elbow flexion and cupped wrist.
8. Lift can (front): Subject attempts to lift can and bring it close to lips with a cylindrical grasp.
9. Lift pencil (front): Subject attempts to pick up pencil by using 3-jaw chuck grasp.

10. Pick up paper clip (front): Subject attempts to pick up paper clip by using a pincer grasp.
11. Stack checkers (front): Subject attempts to stack checkers onto the center checker.
12. Flip cards (front): Using the pincer grasp, patient attempts to flip each card over.
13. Turning the key in lock (front): Using pincer grasp, while maintaining contact, patient turns key fully to the left and right.
14. Fold towel (front): Subject grasps towel, folds it lengthwise, and then uses the tested hand to fold the towel in half again.
15. Lift basket (standing): Subject picks up basket by grasping the handles and placing it on bedside table.

FMA: Upper Extremity Portion
I. Reflex activity
  1. Biceps
  2. Triceps
II. Flexor synergy
  3. Shoulder retraction
  4. Shoulder elevation
  5. Shoulder abduction
  6. Shoulder outward rotation
  7. Elbow flexion
  8. Forearm supination
III. Extensor synergy
  9. Shoulder adduction/inward rotation
  10. Elbow extension
  11. Forearm pronation
IV. Movements combining synergies
  12. Hand move to lumbar spine
  13. Shoulder flexion 0° to 90°
  14. Elbow 90°, pronation/supination
V. Movements out of synergy
  15. Shoulder abduction 0° to 90°
  16. Shoulder flexion 90° to 180°
  17. Elbow 0°, pronation/supination
VI. Reflex activity
  18. Normal reflex activity, biceps and triceps
VII. Wrist
  19. Elbow 90°, wrist stability
  20. Elbow 90°, wrist flexion/extension range of motion
  21. Elbow 0°, wrist stability
  22. Elbow 0°, wrist flexion/extension range of motion
  23. Wrist circumduction
VIII. Hand
  24. Fingers, mass flexion
  25. Fingers, mass extension
  26. Grasp a: First and radial surface of second digit pinch paper.
  27. Grasp b: First and second digit pinch paper.
  28. Grasp c: First and third digit pinch pencil.
  29. Grasp d: First, second, and third digit grip coke can.
  30. Grasp e: All digits grip tennis ball.
IX. Coordination/speed
  31. Tremor
  32. Dysmetria
  33. Speed

References

1. 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.
2. Sanford J, Moreland J, Swanson LR, Stratford PW, Gowland C. Reliability of the Fugl-Meyer assessment for testing motor performance in patients following stroke. *Phys Ther*. 1993;73:447–454.
3. DeWeerdt W, Harrison M. Measuring recovery of arm-hand function in stroke patients: a comparison of the Brunnstrom-Fugl-Meyer test and Action Research Arm test. *Physiotherapy*. 1985;70:542–548.
4. Wood-Dauphinee S, Williams J, Shapiro S. Examining outcome measures in a clinical study of stroke. *Stroke*. 1990;21:731–739.
5. Fugl-Meyer A, Jaasko L, Leyman I, Olsson S, Steglind S. The post-stroke hemiplegic patient: a method for evaluation of physical performance. *Scand J Rehabil Med*. 1975;7:13–31.
6. Duncan P, Propst M, Nelson S. Reliability of the Fugl-Meyer assessment of the sensorimotor recovery following cerebrovascular accident. *Phys Ther*. 1983;63:1606–1610.
7. Lindmark B, Hamrin E. Evaluation of functional capacity after stroke as a basis for intervention: presentation of a modified chart for motor capacity assessment and its reliability. *Scand J Rehabil Med*. 1988;20:103–109.
8. Benaim C, Perennou D, Villy J, Rousseaux M, Pelissier J. Validation of a standardized assessment of postural control in stroke patients: the postural assessment scale for stroke patients (PASS). *Stroke*. 1999;30:1862–1868.
9. Badke M, Duncan P. Patterns of rapid motor responses during postural adjustments when standing. *Phys Ther*. 1983;63:13–20.
10. Sunderland A, Bowers M, Sluman S, Wilcock D, Ardron M. Impaired dexterity of the ipsilateral hand after stroke and the relationship to cognitive deficit. *Stroke*. 1999;30:949–955.
11. Dettman M, Linder M, Sepic S. Relationships among walking performance, postural stability, and functional assessments of the hemiplegic patient. *Am J Phys Med*. 1987;66:77–90.
12. Feys H, De Weerdt W, Nuyens G, van de Winckel A, Selz B, Kiekens C. Predicting motor recovery of the upper limb after stroke rehabilitation: value of a clinical examination. *Physiother Res Int*. 2000;5:1–18.

Assessing Wolf Motor Function Test as Outcome Measure for Research in Patients After Stroke

Steven L. Wolf, Pamela A. Catlin, Michael Ellis, Audrey Link Archer, Bryn Morgan and Aimee Piacentino

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/32/7/1635