The Fugl-Meyer Assessment of Motor Recovery after Stroke: A Critical Review of Its Measurement Properties

David J. Gladstone, Cynthia J. Danells and Sandra E. Black

Neurorehabil Neural Repair 2002 16: 232
DOI: 10.1177/154596802401105171

The Fugl-Meyer Assessment of Motor Recovery after Stroke: A Critical Review of Its Measurement Properties

David J. Gladstone, Cynthia J. Danells, and Sandra E. Black

Measurement of recovery after stroke is becoming increasingly important with the advent of new treatment options under investigation in stroke rehabilitation research. The Fugl-Meyer scale was developed as the first quantitative evaluative instrument for measuring sensorimotor stroke recovery, based on Twitchell and Brunnstrom's concept of sequential stages of motor return in the hemiplegic stroke patient. The Fugl-Meyer is a well-designed, feasible and efficient clinical examination method that has been tested widely in the stroke population. Its primary value is the 100-point motor domain, which has received the most extensive evaluation. Excellent interrater and intrarater reliability and construct validity have been demonstrated, and preliminary evidence suggests that the Fugl-Meyer assessment is responsive to change. Limitations of the motor domain include a ceiling effect, omission of some potentially relevant items, and weighting of the arm more than the leg. Further study should test performance of this scale in specific subgroups of stroke patients and better define its criterion validity, sensitivity to change, and minimal clinically important difference. Based on the available evidence, the Fugl-Meyer motor scale is recommended highly as a clinical and research tool for evaluating changes in motor impairment following stroke.

Key Words: *Stroke recovery—Measurement—Scales*

The Fugl-Meyer (FM) Assessment[1] is considered by many in the field of stroke rehabilitation to be one of the most comprehensive quantitative measures of motor impairment following stroke, and its use has been recommended for clinical trials of stroke rehabilitation. This scale is a disease-specific objective impairment index designed specifically as an evaluative measure for assessment of recovery in the poststroke hemiplegic patient. In this article, we review the development of this scale and its measurement properties—sensibility, reliability, validity, and responsiveness. Limitations of the scale are discussed, and recommendations for future use are presented.

From the Division of Neurology, Department of Medicine, Sunnybrook and Women's College Health Sciences Centre, Toronto, Ontario, Canada (DJG, CJD, SEB); and Graduate Department of Rehabilitation Science, University of Toronto, Toronto, Ontario, Canada (CJD).

Address correspondence and reprint requests to Sandra Black, MD, Cognitive Neurology and Stroke Research Unit, A421, Sunnybrook and Women's College Health Sciences Centre, 2075 Bayview Avenue, Toronto, Ontario, Canada M4N 3P5.

Gladstone DJ, Dannells CJ, Black SE. The Fugl-Meyer assessment of motor recovery after stroke. Neurorehabil Neural Repair 2002;16:232–240.

DESCRIPTION OF THE SCALE

The FM scale is a 226-point multi-item Likert-type scale developed as an evaluative measure of recovery from hemiplegic stroke. It is divided into 5 domains: motor function, sensory function, balance, joint range of motion, and joint pain. Each domain contains multiple items, each scored on a 3-point ordinal scale (0 = *cannot perform*, 1 = *performs partially*, 2 = *performs fully*). The motor domain includes items measuring movement, coordination, and reflex action about the shoulder, elbow, forearm, wrist, hand, hip, knee, and ankle. The motor score ranges from 0 (hemiplegia) to a maximum of 100 points (normal motor performance), divided into 66 points for the upper extremity and 34 points for the lower extremity (Table 1). Similarly, there is a maximum of 24 points for sensation, 14 points for sitting and standing balance, 44 points for joint range of motion, and 44 points for joint pain. The FM assessment is best administered by a trained physical therapist on a one-to-one basis with the patient. It takes approximately 30 minutes to administer.

DEVELOPMENT OF THE SCALE

Very little has been written about the method of development of the FM scale. The original report, published in 1975, outlined the purpose and target population for the scale, its content, and instruc-

Table 1. Fugl-Meyer Assessment—Motor Domain Items[a]

Upper Extremity (66 points)	Lower Extremity (34 points)
Shoulder retraction	Hip flexion
Shoulder elevation	Hip extension (supine)
Shoulder abduction	Hip adduction (supine)
Shoulder abduction to 90 degrees	Knee flexion (supine)
Shoulder adduction/internal rotation	Knee flexion (sitting)
Shoulder external rotation	Knee flexion (standing)
Shoulder flexion 0–90 degrees	Knee extension (supine)
Shoulder flexion 90–180 degrees	Ankle dorsiflexion (supine)
Elbow flexion	Ankle dorsiflexion (sitting)
Elbow extension	Ankle dorsiflexion (standing)
Forearm supination	Ankle plantar flexion (supine)
Forearm pronation	Heel-shin speed
Forearm supination/pronation (elbow at 0 degrees)	Heel-shin tremor
Forearm supination/pronation (elbow at 90 degrees, shoulder at 0 degrees)	Heel-shin dysmetria
Hand to lumbar spine	Knee reflex
Wrist flexion/extension (elbow at 0 degrees)	Hamstring reflex
Wrist flexion/extension (elbow at 90 degrees)	Ankle reflex
Wrist extension against resistance (elbow at 0 degrees)	
Wrist extension against resistance (elbow at 90 degrees)	
Wrist circumduction	
Finger flexion	
Finger extension	
Extension of MCP joints, flexion of PIPs/DIPs	
Thumb adduction	
Thumb opposition	
Grasp cylinder	
Grasp tennis ball	
Finger-nose speed	
Finger-nose tremor	
Finger-nose dysmetria	
Finger flexion reflex	
Biceps reflex	
Triceps reflex	

[a]See Fugl-Meyer and others[1] for details and scoring instructions.

tions for scoring. Although the rationale for such a scale is sound, the process of item generation and reduction appear to be based more on art than on science.

Rationale

Hemiparesis is the most common disabling deficit after stroke, affecting 70% to 80% of stroke survivors and is often the deficit most in need of rehabilitation. The rationale for introducing this index in 1975 was the lack of an existing scale that provided a specific and quantitative method for measuring recovery from hemiplegia. In their original paper, the Fugl-Meyer authors stated "the assessment methods applied by the advocates for the different (rehabilitation) methods all lack numerical scores and it is consequently difficult to assess the efficacy of treatment."[1] Rehabilitation specialists needed a reliable method for assessing recovery and the effects of the different physical therapy techniques that were being developed. Most existing scales at that time focused on measuring disability and independence in activities of daily living (ADL); they ignored "the neuromuscular capacity per se."[1] Therefore, there was a need not only for a quantitative measure but also for a measure of neurologic recovery at the impairment level. The authors also recognized that previous assessment methods lacked standardization of the patients' posture and motor performance and allowed for varying degrees of compensatory mechanisms by the patient. For these reasons, a new scale appeared justified.

Today, 25 years after the introduction of this scale, the rationale for this type of scale still stands. In fact, there is even greater demand now for such a scale because of the explosion in recent years of new treatment options for stroke and an increasing number of clinical trials investigating rehabilitation therapies. Outcome measures for stroke rehabilitation trials need to adequately reflect changes in the rate and quality of patients' performance during the course of recovery.

Recovery from stroke is extremely difficult to measure.[2] One of the greatest challenges is the tremendous individual variability in the rate and degree of spontaneous recovery. Stroke is a heterogeneous disorder and recovery depends on multiple factors, including the type of stroke; type, severity and number of accompanying neurological deficits; lesion size and location; patient characteristics and comorbid conditions; and rehabilitation therapy, among others.[3-5] Scales that use simple objective dichotomous outcomes (e.g., alive/dead, independent/dependent), disability scales (e.g., Barthel Index (BI), Functional Independence Measure [FIM]), or "global" deficit rating scales (e.g., the National Institutes of Health Stroke Scale [NIHSS]) are in common use, but are inadequate to capture the dynamic process of motor recovery and are less likely to detect changes in response to specific treatments. In fact, the failure of many recent clinical stroke trials may relate more to the choice of outcome measures rather than to the lack of efficacy of the agent under investigation.[6] Individuals considered "recovered" neurologically, as defined by the widely used NIHSS scale (NIHSS score 0), may still have significant weakness of the arm or leg that is not captured on this scale. The need for impairment level measures in addition to disability measures conforms to the World Health Organization concept of recovery encompassing impairment, activity limitations (disability), and participation restrictions (handicap).[7]

Item Generation and Reduction

Development of a new scale involves gathering a pool of all potential items that will provide appropriate and comprehensive content and then choosing the best items through the process of item reduction. An evaluative scale must contain suitable content and items that are sensitive to change. In constructing the FM scale, item selection was based on empirical observations of normal recovery patterns in hemiplegic stroke patients. Fugl-Meyer and colleagues referred to the work of Twitchell, whose classic 1951 paper provides a detailed description of the natural history of motor recovery following stroke.[8] Twitchell documented that restoration of motor function in the arm and leg proceeds through an orderly, predictable, stepwise sequence after an initial areflexic flaccid paralysis: reflexes return and become hyperactive; muscle tone increases and spasticity develops; voluntary movement appears as part of stereotyped, reflexive flexor and extensor muscle "synergies"; voluntary movement may then be achieved "out of synergy"; finally, normalization of muscle tone and reflexes may then occur. Recovery may plateau and cease at any stage along the way, leaving residual impairment. In the arm, the first movement to return is usually at the shoulder, followed by elbow, wrist and fingers; initially any attempt at movement results in a flexor synergy with flexion of shoulder, elbow, wrist, and fingers together as a unit. Similarly, an extensor synergy can develop, appearing as shoulder adduction/internal rotation with extension of elbow and wrist and finger flexion/adduction. Voluntary movement is characterized by slowness in initiation, execution, and relaxation; limited range of motion; and weak power. To summarize, recovery from hemiplegia progresses from proximal to distal muscles in the arm and leg, return of reflexes always precedes the return of voluntary movement, and the first movement to return at any joint is flexor movement in the arm and extensor movement in the leg.[8] Based on these observations, Reynolds and others[9] described methods for monitoring motor recovery in hemiplegic stroke patients. Brunnstrom and others[10,11] similarly defined 6 sequential stages of motor recovery and described how the hemiplegic arm and leg progress though these stages as a method for assessing recovery. These assessments, however, measured motor recovery only in a qualitative fashion that was time-consuming and without numerical scoring; therefore, it was difficult to determine efficacy of different rehabilitation treatment methods. The Fugl-Meyer scale adopted a similar approach to motor assessment and quantified it. Item generation for the motor domain of the FM scale aimed at incorporating all the relevant components in the stages of recovery described by Twitchell and Brunnstrom and therefore appeared well grounded. Bobath[12] emphasized the important relationship between posture (certain body parts stabilized) and ability to perform selective movements of the limb. This concept was also incorporated into the development of the FM scale (e.g., the ability to sit tall and stabilize the shoulder at 90 degrees of flexion while pronating/supinating the forearm). The design of the other 5 domains of the scale lacks sufficient rationale or explanation of item generation; this is discussed further below.

Scoring of each item is on a 3-point ordinal scale. The only justification for this, based on unpublished observations, was stated as follows: "we have found in a pilot study that scales consisting of

5 or 7 steps were likely to elicit less reliable judgements, while the 3-graded scale gave a good degree of reliability."[1] On the contrary, expanding the grading system of each item may have actually been beneficial in maximizing the ability of the scale to detect change—which is the main purpose of the scale. As an example, the widely used Medical Research Council muscle strength scale uses a 6-point scale: 0 = *no movement*; 1 = *flicker of movement*; 2 = *movement with gravity eliminated*; 3 = *movement against gravity*; 4 = *movement against resistance*; 5 = *full power*. If a similar scoring system were applied to the items in the FM motor scale, the scale might be able to discern not only the ability to perform a movement, but also different grades of muscle strength. This, then, could make the scale more sensitive to change for those patients at the top end (i.e., it could reduce the "ceiling effect" for patients with good motor recovery). In clinimetrics, the optimal number of categories recommended is usually 5 to 7.[13]

Fugl-Meyer and colleagues administered their scale to a test population of 20 hemiparetic acute stroke patients aged 65 or younger who were admitted to a university hospital in Sweden in 1971.[1] Assessments were performed on the following schedule: an average of 4 days poststroke, at regular intervals for 6 weeks, monthly thereafter until 6 months, and again at 1 year. They graphed motor recovery in the arm and leg using the motor scores as a percentage of the maximum obtainable score. Using this scale, they were able to demonstrate the expected evolution of the sequential stages of motor recovery over time. In most patients, successive restoration of motor function was observed. Recovery was not limited to the first weeks or months but continued to 1 year. No recovery of hand function took place when motor function of the shoulder-arm did not reach at least 40% of maximum. Progression from one stage to the next began when motor function of the previous stage reached at least 60% in the arm and 80% in the leg. Without any formal testing of interrater or intrarater reliability, Fugl-Meyer and colleagues[1] stated that the scale was "reliable" because "the rigidly standardized procedures and the chosen scales allow little chance of error, and thus the procedures are reliable." The authors did not make explicit modifications to their scale after their pilot study. One approach now frequently used to improve the development of a scale is to assemble a panel of experts to provide additional clinical judgment with respect to appropriateness, relative importance and weighting of the items, and elimination of redundant items. Ideally, such a panel would sample a wide spectrum of specialists, such as physical therapists, occupational therapists, physiatrists, neurologists, and stroke experts. Methods to achieve consensus, such as the Delphi method, could be applied.

SENSIBILITY

The motor domain of the FM scale makes sense. Based on the rehabilitation literature, clinical experience using this scale in hospitalized stroke patients, and Feinstein's checklist for appraising the sensibility of a clinical index,[14] this subscale fulfills a high degree of sensibility. The purpose of the scale, target population, and setting are all clearly specified. It is based on the well-defined stages of motor recovery.

Face validity and content validity for the motor domain are very good. The items included and the breakdown of categories (i.e., movement ability, reflexes, speed, etc.) are appropriate. However, the scaling is weighted heavily for the upper extremity (66 points) compared to the lower extremity (34 points), and reflexes appear to be overrepresented in the scoring system. A notable omission is the assessment of individual finger movements. Although gross hand function (grasping tasks) is included, recovery of distal fine motor function (a hallmark of good motor recovery) is underrepresented. Inclusion of more complex tasks, including finger extension, dexterity, and speed would likely increase the scale's responsiveness among patients with good recovery who have achieved the maximum obtainable scores on the hand component. From our own experience, the FM scale appears to have a ceiling effect for both the hand and lower extremity items. Impairment of the trunk muscles is also not assessed in the motor domain.

In contrast to the motor subscale, the nonmotor domains—sensation, balance, joint range of motion, and joint pain—appear less well suited for the intended purpose of this instrument. The choice of these items and their justification are not stated. The joint range of motion domain is taken from the 1965 American Academy of Orthopedic Surgeons standard method for measuring and recording joint motion. Although it is reasonable to document range of motion as a confounding factor (because pain and contracture can limit movement), inclusion of a separate domain for joint pain appears unnecessary and redundant for the purpose of this scale. The sensory domain appears on face value to be insufficient, including only light touch and proprioception and not other sensory modalities. The balance domain makes sense and provides a general idea of how patients are progressing (i.e., righting reactions are assessed before balance, and tasks progress from static positions with a wide base of support to positions over a narrow base). However, better quality balance scales are now available if one is primarily interested in assessing this feature after stroke (e.g., Berg balance scale[15]). Inclusion of irrelevant or excessive items can reduce the ability of a scale to detect change. A strength of the scale is that it is predominantly an objective measure rated by a trained physical therapist. However, inclusion of the subjective domains of sensation and joint pain may detract from this objectivity and could potentially reduce the scale's reliability and ability to detect change. Light touch and proprioception may be difficult or impossible to assess in the presence of aphasia or hemispatial inattention, common accompaniments in this patient population. Fugl-Meyer and colleagues[1] admit that in their test population the sensory ratings were either unreliable or could not be assessed in many patients due to poor cooperation or the presence of aphasia. Realizing this, the authors should have decided to modify or eliminate this section and then administer the scale to another test population. Instead, the scale was left as is, and no attempts at item reduction were reported. It is not surprising, then, that the FM scale is often used exclusively as a motor scale without the other domains.

The FM scale has gained international acceptance as a feasible and appropriate assessment of motor recovery in stroke rehabilitation. The items and scoring are easy to follow. The assessment procedure is simple to administer by a trained physical therapist and is acceptable to the patients being tested. The instructions are relatively straightforward, and the assessment requires no special equipment, in contrast to other assessment scales. The order of administration of the items can be modified to avoid repeated changes in patient positioning (e.g., from supine to sitting and standing), making the assessment more efficient and easier for patients. The scale lends itself to administration at the patient's bedside in an acute care hospital or in an inpatient or outpatient rehabilitation setting. The FM scale, like most clinical examinations, can be dependent on patient attention, motivation, cooperation, and effort, as the patient must be an active participant and give full effort in attempting the required tasks. The complete scale, in its original version, can be quite long and time-consuming to administer, which may limit its widespread applicability to hospitalized patients in clinical practice. However, when only the motor scale is performed, it takes 20 minutes for patients with a mild motor deficit and 10 minutes for those with a severe motor deficit. It is readily applicable for serial assessments during rehabilitation by physical therapists as part of routine patient care to monitor progress and guide therapy or for research purposes.

RELIABILITY

The ability to measure attributes in a reproducible and consistent manner is a necessary requirement of a valid and responsive measurement instrument. Excellent intrarater and interrater reliability have been established for the FM scale.

Duncan and others[16] examined intrarater reliability in a group of 19 chronic stroke patients at the Durham Veterans Administration Medical Center more than 1 year following hemiplegic stroke. The mean age was 56 ± 13 years (range 34–76), and the patients' scores ranged from 50% to 98% of maximal performance. Assessments were conducted by a single physical therapist on 3 occasions at 3-week intervals. For these patients with stable deficits, intrarater Pearson correlation coefficients were high for the total score (0.98–0.99), upper extremity motor subscore (0.995–0.996), lower extremity motor subscore (0.96), sensation subscore (0.95–0.96), joint range of motion/pain subscore (0.86–0.996), and balance subscore (0.89–0.98), ($P < 0.001$ for all). In addition, this study reported the interrater reliability among 4 physical therapists for the motor performance subsection: 0.98–0.995 for upper extremity motor function ($n = 8$) and 0.89–0.95 for lower extremity motor function ($n = 10$, $P < 0.001$). This study employed Pearson correlation, which does not correct for chance agreement, rather than the intraclass correlation coefficient (ICC) test of concordance for continuous data.

Sanford and others[17] extended these results by investigating interrater reliability for all subsections of the FM scale during the first 6 months poststroke. Twelve consecutive stroke patients (mean age 66) undergoing active rehabilitation were studied. Total scores ranged from 79 to 220, with a good distribution of severe, marked, and moderate severity stroke, and 1 patient with slight impairment. Patients were evaluated 1 day apart by 3 experienced physical therapists. The ICCs were used to quantify reliability. High interrater reliability was shown for the total score (0.96), upper extremity motor subscore (0.97), lower extremity motor subscore (0.92), balance (0.93), and range of motion (0.85). The sensation subscore ICC was

0.85 but had the greatest variance. Pain assessment was the least reliable section with an ICC score of 0.61.

These assessments of reliability were conducted in the appropriate patient populations and in similar settings for which the measurement was intended. There were criticisms concerning the small number of subjects tested and the lack of information available on patients with mild strokes. Separate reliability assessments have not been reported for patients stratified by stroke severity into mild, moderate, and severe categories. It is anticipated that reliability may be highest for those with the most severe strokes, as they will score 0 for most items; conversely, greater measurement variability is likely for patients with mild and moderate severity motor deficits who score in the middle to upper range of the scale. To aid in the interpretation of clinical trials that often stratify subjects according to stroke severity, future reliability studies should be performed in groups of patients homogeneous with respect to disease severity.

Sources of measurement variability include both the patient and the clinician administering the FM scale. Procedure-related variability is not a factor because no technical procedure or equipment is involved. As with most clinical assessments, patient variability may be minimized by examining patients at an appropriate time of day in comfortable and consistent testing conditions, preferably a quiet examination room without distractions, to ensure maximum patient cooperation. Clinician variability may be reduced by ensuring consistent examination techniques through training sessions and open discussion among different examiners. The patient's normal, unaffected limbs can serve as a natural control for comparison to help ensure consistent ratings of the hemiparetic side.

VALIDITY

There is good evidence from several validation studies that the FM scale is indeed measuring what it is intended to measure. Most studies have established construct validity by comparing the FM scale with measures of stroke recovery based on various functional scales reflecting independence in ADL or disability level following stroke. For example, Fugl-Meyer and others[18] compared the FM motor subscores to ADL capacity in 64 hemiplegic stroke patients more than 6 months after discharge from hospital. ADL performance was rated on its own scale comprising 52 items, each on a 5-point ordinal scale from dependent to fully independent; the items included feeding, hygiene, dressing, locomotion, housework, and psychosocial functioning. Significant correlations were found between the degree of motor impairment measured on the FM motor scale and ADL total score (0.75), hygiene (0.89), locomotion (0.76), feeding (0.72), and dressing (0.76). To further explore convergent validity, Dettmann and others[19] administered the FM scale together with the BI, a well-validated and widely used disability measure in stroke rehabilitation. Fifteen patients with hemiplegic stroke (mean age 64 years, range 46–87) were assessed at an average of 2 years poststroke (range 1 month to 11 years). All assessments were conducted by a single physical therapist. Pearson correlation coefficients between FM and BI scores were 0.67 for the FM total score, 0.74 for the FM motor subscore (mean score 54/100, range 2–97), 0.75 for the upper extremity motor subscore, and 0.76 for the balance subscore ($P < 0.01$ for all). The investigators also conducted their own objective measurements of walking performance using photographic analysis of gait pattern and velocity, as well as objective measurements of postural stability while standing on a force platform. The FM lower extremity motor subscore correlated well with most of the gait measurements, and the FM balance subscore correlated well with their stability index. The sensation subscore did not correlate significantly with any of these measures of gait or upright stability. In a similar study, Wood-Dauphinee and others[20] compared the FM scale to the BI in a large series of stroke patients ($n = 167$) at 2 time points—acute stage (3 to 5 days poststroke) and 5 weeks poststroke. The mean age was 74 ± 10 years, and the majority had ischemic hemispheric strokes. The FM upper extremity motor subscore was highly correlated with the BI total score in the acute stage (Pearson correlation coefficient, $r = 0.75$) and at 5 weeks ($r = 0.82$). Similarly, the FM lower extremity motor subscore correlated highly with the BI acute stage ($r = 0.77$) and at 5 weeks ($r = 0.89$), $P < 0.01$ for all. Most recently, Shelton and others[21] compared the FM scale to the FIM, another widely used and well-validated disability scale. The FIM has 18 items, each on a 7-point scale (total score ranges from 18 to 126) assessing mobility, locomotion, self-care, sphincter control, communication, and social cognition. They assessed 172 inpatients in a stroke rehabilitation hospital within 90 days of ischemic hemispheric stroke. Total FM scores correlated with total FIM ($r = 0.63$); FM upper extremity motor scores correlated with FIM self-care scores ($r = 0.61$); and lower extremity motor scores correlated with FIM mobility scores ($r = 0.74$), $P < 0.0001$ for all.

Other studies have specifically examined the validity of the FM scale for measuring upper limb motor recovery. For instance, DeWeerdt and others[22] compared FM upper extremity motor scores to the Action Research Arm (ARA) test, a validated scale of upper limb motor function. Both assessments were administered to 53 hospitalized stroke patients with a motor deficit. The 2 tests were highly correlated both at 2 weeks (r = 0.91) and at 8 weeks poststroke (r = 0.94), P < 0.01. These results, then, demonstrate that the impairment scores on the FM scale correlate with clinically meaningful function of the upper limb. However, many patients in this study still had poor hand function at 8 weeks; therefore, it would be important to establish the validity over a longer period of follow-up (i.e., 3 to 6 months) because, as hand function recovers, the scores on the ARA test may continue to improve, whereas the FM scores may plateau (ceiling effect). Berglund and Fugl-Meyer[15] compared the FM scale to another assessment of upper limb function, the DeSouza scale. Fifty consecutive stroke patients with a motor deficit, who were discharged home, were assessed by a single occupational therapist. Upper extremity motor scores were highly correlated (Spearman coefficient = 0.90). However, there was clustering of 10 patients at the lowest score of the DeSouza scale, and 6 achieved the maximum score, suggesting ceiling and floor effects, whereas there was no such clustering of scores on the FM scale for these patients. Therefore, the FM motor scale appeared to have more discriminative power toward the endpoints of the scale.

Thus, a high degree of construct validity has been established for the FM scale. In particular, the motor subscale correlates well with disability measures—the BI in the acute, subacute, and chronic stages poststroke, and the FIM—as well as with upper extremity functional recovery. There is a lack of information, however, regarding criterion validity, since at the time the FM scale was devised no other similar impairment instrument existed against which the FM scale could be tested. More recently, however, the Chedoke-McMaster Stroke Assessment has been developed and tested against the FM scale.[23,24] This scale assesses attributes similar to the FM scale in its impairment inventory and has gained acceptance in clinical stroke rehabilitation practice in Canada because it also contains a disability inventory and is useful for both discriminative and evaluative purposes. Concurrent validity testing reveals high correlations between all impairment subscores on the FM and the corresponding impairment subscores on the Chedoke impairment inventory (r = .95, P < 0.001).[24] Further research, perhaps using electromyography, could clarify the criterion of "synergism" in the stages of motor control as the underlying basis for the FM motor scale.

RESPONSIVENESS

Demonstration of responsiveness, the ability of an instrument to detect clinical change over time, is essential for any evaluative measure. In contrast to reliability and validity, however, formal assessment of responsiveness for outcome measures used in stroke rehabilitation has received relatively little attention.

For the FM scale, responsiveness can be inferred based on its demonstrated reliability in stable patients and on the wide range of impairments captured by the 100-point motor domain (from complete paralysis to recovery or very mild motor impairment). However, because a ceiling effect may exist, the scale may be most responsive to changes in those patients with severe and moderate deficits who will not achieve the maximum possible scores.

Responsiveness of the FM scale can be seen graphically in stroke recovery curves plotted as a function of time. For example, in a prospective cohort of 459 stroke patients, Duncan and others[2] graphed the rate and degree of change over the first 6 months following stroke as measured by the FM scale and other common stroke scales. The FM scores show that spontaneous improvement occurs for mild, moderate, and severe strokes; maximum recovery occurs within the first month; and that recovery plateaus around 6 months poststroke. The recovery curves for the FM upper and lower extremity motor subscales closely parallel the course of recovery as measured by the BI and NIHSS. Similarly, Dettman and others[19] followed 19 patients longitudinally over 3 time intervals, and their FM motor recovery graphs also closely paralleled ADL recovery over time.

Other studies have calculated correlations of change scores between the FM and other stroke scales. For example, in the study by Wood-Dauphinee and others[20] described above, the correlation between mean change scores for FM upper and lower extremity motor subscores and total BI scores was 0.57 (P < 0.01). Shelton and others[21] found the correlation of FM motor change scores with total FIM change scores was 0.44 (P < 0.0001). Linear regression analysis demonstrated a 24-point increase in FIM score for every 10-point increase in FM motor score, indicating that improvement in motor impairment is associated with significant functional recovery. Further studies like this are needed to determine how motor recovery at the impairment level corresponds to measures at the disability level.

Knowledge of response stability is important to interpret the significance of change scores. Van der Lee and others[25] assessed the stability of the FM upper extremity motor score in chronic stroke patients. The limits of test-retest agreement found were –5 to 6.6 on the 66-point upper limb scale. In this study, the ARA test was more responsive to improvement in upper limb function in response to a forced use treatment intervention. Similarly, Sanford and others[17] showed that the range of measurement error (± 2 standard error of the mean [*SEM*]) was ± 7.2 for the FM upper limb motor score and ± 6.4 for the FM lower limb motor score, although this study used multiple raters. To conclude that true change has occurred, the observed improvement should exceed such error threshold values.[17,25]

Although the minimal clinically important difference on the FM scale is not yet known, a greater than 10-point (10%) change in FM motor scores may represent a clinically meaningful improvement based on clinical experience with this scale and consultation with physical therapists and stroke neurologists. It would also be relevant to know what change constitutes a clinically important difference from the patient's perspective, but this issue has not been addressed for the FM scale. Certainly, the responsiveness of the FM scale deserves more formal evaluation because its primary purpose is as an evaluative tool, and it is becoming increasingly utilized in clinical trials.

DISCUSSION

The FM scale is a much-needed instrument for monitoring the course of recovery from hemiplegic stroke. Its design, content, and measurement properties strongly favor the use of the 100-point motor domain to evaluate changes in motor impairment following stroke. Excellent intrarater and interrater reliability have been demonstrated for the entire scale and each of its subsections. Construct validity, particularly of the motor domain, is supported by numerous studies; criterion validity has been less well studied.

Although developed in 1975 without the formal methodology that would be expected of new scales developed today, item generation based on empirical knowledge and clinical judgment appears to have been successful in creating a motor domain that reflects the phenomena of interest. Inclusion of the sensation, balance, joint motion, and joint pain domains was based on an ad hoc approach; more rigorous development with expert consensus panels might improve the content of these sections. The sensation domain is likely to be less reliable in patients with aphasia or other cognitive impairments.

The FM scale should have particular value in clinical trials designed to evaluate changes in motor impairment following stroke, with some limitations kept in mind. It must be remembered that the FM scale is strictly an impairment index; how its scores relate to disability remains to be determined more completely. As with most scales, its use as a measurement of recovery for patients with mild motor impairment is limited by a ceiling effect. Thus, we recommend against inclusion of patients with mild impairments into clinical trials using the FM scale as the sole outcome measure. Using the FM scale in combination with a general activity measure (e.g., Chedoke-McMaster Disability Inventory[23]) or a specific activity measure (e.g., Chedoke Arm and Hand Activity Inventory[26]) might provide additional information to improve the measurement of recovery for patients who plateau or reach the maximum scores on the FM items. The combination of the FM motor impairment scale with functional scales should be evaluated in future studies.

Most studies have tested the FM scale in heterogeneous stroke populations without clearly specifying the stroke type, lesion localization, and patient characteristics; it would be valuable to know more about how this scale performs in patients stratified according to stroke severity and in those with different stroke subtypes and accompanying deficits. Further evaluation should aim to assess responsiveness in specific subgroups of stroke patients, compare its efficiency to other instruments, and better define the minimal clinically important difference for patients with varying degrees of hemiparesis. Administration of the scale requires a trained physical or occupational therapist. To our knowledge, there are no certification training programs available. We are currently developing a videotape demonstration of this scale to be made available for educational purposes.

ACKNOWLEDGMENT

The authors acknowledge funding from the Canadian Institutes of Health Research (CIHR). Dr Gladstone is supported by the Royal College of Physicians and Surgeons of Canada Clinician-

Investigator Program in the Department of Medicine and Institute of Medical Sciences, University of Toronto, and by a research fellowship from the CIHR and Heart and Stroke Foundation of Ontario.

REFERENCES

1. Fugl-Meyer AR, Jaasko L, Leyman I, Olsson S, Steglind S. The post-stroke hemiplegic patient. *Scand J Rehabil Med* 1975;7:13-31.
2. Duncan PW, Lai SM, Keighley J. Defining post-stroke recovery: implications for design and interpretation of drug trials. *Neuropharmacology* 2000;39:835-41.
3. Kwakkel G, Wagenaar RC, Kollen BJ, Lankhorst GJ. Predicting disability in stroke. A critical review of the literature. *Age Ageing* 1996;25:479-89.
4. Giaquinto S, Buzzelli S, Di Francesco L, et al. On the prognosis of outcome after stroke. *Acta Neurol Scand* 1999;100:202-08.
5. Counsell C, Dennis M. Systematic review of prognostic models in patients with acute stroke. *Cerebrovasc Dis* 2001;12:159-70.
6. Duncan PW, Jorgensen HS, Wade DT. Outcome measures in acute stroke trials: a systematic review and some recommendations to improve practice. *Stroke* 2000;31:1429-38.
7. Gray DB, Hendershot GE. The ICIDH-2: developments for a new era of outcomes research. *Arch Phys Med Rehabil* 2000;81:S10-S14.
8. Twitchell TE. The restoration of motor function following hemiplegia in man. *Brain* 1951;74:443.
9. Reynolds G, Archibald KC, Brunnstrom S, Thompson N. Preliminary report on neuromuscular function testing of the upper extremity in adult hemiplegic patients. *Arch Phys Med* 1958;39:303.
10. Brunnstrom S. Motor testing procedures in hemiplegia: based on sequential recovery stages. *Phys Ther* 1966;46: 357-75.
11. Brunnstrom S. *Movement therapy in hemiplegia*. New York: Harper & Row; 1970.
12. Bobath B. *Adult hemiplegia: evaluation and treatm*ent. London: William Heinemann Medical Books; 1970.
13. Streiner DL, Norman GR. *Health measurement scales: a practical guide to their development and use*. 2nd ed. New York: Oxford Medical Publications; 1995.
14. Feinstein AR. *Clinimetrics*. New Haven (CT): Yale University Press; 1987.
15. Berglund K, Fugl-Meyer AR. Upper extremity function in hemiplegia. A cross-validation study of two assessment methods. *Scand J Rehab Med* 1986;18:155-7.
16. Duncan PW, Propst M, Nelson SG. Reliability of the Fugl-Meyer assessment of sensorimotor recovery following cerebrovascular accident. *Phys Ther* 1983;63:1606-10.
17. Sanford J, Moreland J, Swanson LR, Stratford PW, Gowland C. Reliability of the Fugl-Meyer assessment for testing motor performance in patients following stroke. *Phy Ther* 1993;73:447-54.
18. Fugl-Meyer AR, Jaasko L. Post-stroke hemiplegia and ADL performance. *Scand J Rehabil Med* 1980;7:140-52.
19. Dettmann MA, Linder MT, Sepic SB. Relationships among walking performance, postural stability, and functional assessments of the hemiplegic patient. *Amer J Phys Med* 1987;66:77-90.
20. Wood-Dauphinee SL, Williams JI, Shapiro SH. Examining outcome measures in a clinical study of stroke. *Stroke* 1990;21:731-9.
21. Shelton FNAP, Volpe BT, Reding MJ. The effect of motor impairment on disability following stroke [abstract]. *Stroke* 2000;31(1):291.
22. De Weerdt WJG, Harrison MA. Measuring recovery of arm-hand function in stroke patients: a comparison of the Brunnstrom-Fugl-Meyer test and Action Research Arm test. *Physiother Can* 1985;37:65-70.
23. Gowland C, VanHullenaar S, Torresin W, et al. Chedoke-McMaster stroke assessment: development, validation and administration manual. Hamilton (ON), Canada: Chedoke-McMaster Hospitals and McMaster University; 1995.
24. Gowland C, Stratford P, Ward M, et al. Measuring physical impairment and disability with the Chedoke-McMaster Stroke Assessment. *Stroke* 1993;24:58-63.
25. 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;33:110-3.
26. Barreca S, Gowland C, Stratford P, et al. Development of the Chedoke arm and hand activity inventory: item selection. *Physiother Can* 1999;209-11.