Minimal Detectable Change and Clinically Important Difference of the Wolf Motor Function Test in Stroke Patients

Keh-chung Lin, ScD, OTR, Yu-wei Hsieh, MS, Ching-yi Wu, ScD, OTR, Chia-ling Chen, MD, PhD, Yuh Jang, PhD, OTR, and Jung-sen Liu, MD, PhD

Objectives. This study aimed to establish the minimal detectable change (MDC) and clinically important differences (CID) of the Wolf Motor Function Test (WMFT) in patients with stroke, and to assess the proportions of patients' change scores exceeding the MDC and CID after stroke rehabilitation. Methods. A total of 57 patients received 1 of the 3 treatments for 3 weeks and underwent clinical assessments before and after treatment. The MDC, at 90% confidence (MDC90), was calculated from the standard error of measurement to indicate a real change for individual patients. Anchor-based and distribution-based approaches were used to triangulate the values of minimal CID. The percentages of patients exceeding the MDC and minimal CID were also examined. Results. The MDC90 of the WMFT was 4.36 for the performance time (WMFT time) and 0.37 for the functional ability scale (WMFT FAS). The minimal CID ranged from 1.5 to 2 seconds on the WMFT time and from 0.2 to 0.4 points on the WMFT FAS. The MDC and CID proportions ranged from 14% to 30% on the WMFT time and from 39% to 65% on the WMFT FAS, respectively. Conclusions. The change score of an individual patient has to reach 4.36 and 0.37 on the WMFT time and WMFT FAS to indicate a real change. The mean change scores of a stroke group on the WMFT time and WMFT FAS should achieve 1.5 to 2 seconds and 0.2 to 0.4 points to be regarded as clinically important changes. Furthermore, the WMFT FAS may be more responsive than the WMFT time based on the results of proportions exceeding the threshold criteria.

Keywords: Minimal detectable change; Clinically important difference; Outcome measures; Upper extremity; Stroke rehabilitation; Wolf Motor Function Test

The Wolf Motor Function Test (WMFT) was originally developed by Wolf et al. to evaluate the effects of forced use in stroke patients. The WMFT has been widely used as an outcome measure in stroke rehabilitation trials, particularly in constraint-induced therapy studies. Although the reliability and validity of the WMFT have been established in subacute and chronic stroke, properties of the test to detect changes after interventions have not been examined yet.

Clinimetrics, one of methodological disciplines introduced by Feinstein, focuses on measurement issues in clinical research and practice. Clinimetric indexes fit into clinicians’ frame of thinking and thus may be more helpful and attractive to clinicians. For example, when considering the index of test-retest reliability, clinicians may prefer an index that indicates the size of measurement error rather than the intraclass correlation coefficient that is commonly reported in psychometric studies.

One of the more common measurement error indicators for individual use is the minimal detectable change (MDC), which is expressed in the same units as the outcome measure. The MDC, estimated from the standard error of measurement (SEM), is defined as the smallest amount of change that likely reflects true change rather than measurement error inherent in the score. One advantage of the MDC is that it considers both the reliability and responsiveness to change. That is, the MDC helps clinicians determine whether the change score in individual performance represents real and reliable change and simultaneously tells whether the outcome measure will be able to detect such a change.

A further concern of clinical relevance is the amount of change between scores that can be considered as a clinically important difference (CID). If the treatment-induced change can be identified as clinically important, it would help clinicians guide clinical care and make decisions. Pursuit of the CID is therefore an important area of work in interpretation of the change scores, especially for clinical practice. Several methods, such as anchor-based and distribution-based approaches, have been proposed to determine the CID. Because agreement has not yet been reached on a gold standard approach, combinations of the anchor-based and distribution-based approaches...
to triangulate on a single value or small range of values for the CID have been suggested.23,24

Reporting both the MDC and the CID values will provide information on 2 important benchmarks to assist in answering how much change is required to demonstrate a real and clinically important change. This information will allow clinicians and researchers to tell whether the patients actually experience meaningful change after an intervention. For individual-level use (ie, within-person change), the established minimal CID value should be greater than the MDC to indicate that the measure has the precision to indicate meaningful clinical change.25 In sum, interpretation of the meanings of change scores is a critical issue that should be considered while assessing the usefulness of outcome measures.26 The investigation of MDC and CID of the WMFT is warranted because it is a commonly used outcome measure in stroke rehabilitation. Therefore, we aimed to establish the estimates of the MDC and CID of the WMFT and assess the proportions of patients’ change scores on the WMFT exceeding the MDC and CID after stroke rehabilitation interventions.

Methods

Participants

We recruited 57 participants from 3 medical centers. The inclusion criteria for the participants were as follows: (1) must be 6 months after onset of a first-ever stroke; (2) must be able to reach Brunnstrom stage III27 or above for the proximal upper extremity; and (3) must have no excessive spasticity in the shoulder and elbow joints of the affected upper extremity (Modified Ashworth Scale score28 ≤ 2.5). To avoid the confounding effects of cognitive and medical conditions, participants were excluded if the medical or physical screening examination resulted in a score of less than 24 on the Mini-Mental State Examination,29 if the patient was in poor physical condition or had a physician-determined major medical problem that would interfere with participation, or if there was 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.

Interventions and Procedures

Eligible patients were randomized to a 3-week session of treatment in 1 of the following 3 treatment groups: distributed constraint-induced therapy, bilateral arm training, or traditional rehabilitation. Distributed constraint-induced therapy involved restriction on movement of the unaffected arm by placing it in a mitt for a target of 6 hours per day and intensive training of the affected upper extremity in functional tasks for 2 hours per weekday, including reaching forward or upward to move a cup, picking up coins, picking up a utensil to take food, and other functional movements similar to daily activities needed. Bilateral arm training concentrated on both the affected and unaffected upper extremity, moving simultaneously in functional tasks by the symmetric or alternated patterns for 2 hours per weekday. The functional tasks also emphasized upper extremity movements involved in daily activities, but focused on both upper extremities moving synchronously, such as lifting 2 cups, picking up 2 pegs, reaching forward or upward to move blocks, grasping and releasing 2 towels, and similar activities. Traditional rehabilitation 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 to administer the intervention protocols by the senior authors, and a written competency test was administered before subject treatment. The clinical evaluations were administered before and after treatment by blinded raters.

Clinical Measures

Wolf Motor Function Test. The WMFT is a laboratory-based measurement to assess upper extremity motor function.1 The WMFT contains 15 function-based tasks and 2 strength-based tasks.10 The performance time (WMFT time) and functional ability scale (WMFT FAS) on the 15 function-based items were administered in this study. The speed at which functional tasks can be completed is measured by performance time and the movement quality when completing the tasks is measured by functional ability. The maximum time allowed to complete an item is 120 seconds. For functional ability scoring, we used a 6-point ordinal scale, where 0 = does not attempt with the involved arm and 5 = arm does participate/movement appears to be normal. The test-retest reliability, interrater reliability, criterion validity, and construct validity of the WMFT have been ascertained in patients with stroke.7,10

Fugl-Meyer Assessment. The 33-item upper extremity subscale of the Fugl-Meyer Assessment (FMA) was used to assess motor impairment.30 Items are scored on a 3-point ordinal scale (0 = cannot perform, 1 = performs partially, 2 = performs fully), with a maximum score of 66.30 The psychometric properties of the FMA have been shown to be satisfactory in stroke patients.30-32 The FMA scores of the patients were used to estimate the anchor-based CID of the WMFT in this study. If the patient’s change score from pretreatment to posttreatment on the FMA fell into 6 to 10 points, which is about 10% to 15% of the maximum score, the patient was classified into the group with minimal CID.

Data Analysis

MDC calculation. The MDC is calculated by multiplying the SEM by the z score associated with the desired confidence.
The MDC90 was calculated using the following formula: 
\[
MDC_{90} = 1.65 \times \sqrt{2} \times \text{SEM} = 1.65 \times \text{SD} \times \sqrt{2 \left[1 - r\right]},
\]
where 1.65 is the 2-tailed tabled \(z\) value for the 90% confidence interval, \(r\) is the coefficient of the test-retest reliability. The test-retest reliability coefficients, 0.90 for performance time and 0.95 for functional ability, of the WMFT were based on the study of Morris et al.7

Estimates of CID. The CIDs of the WMFT were determined by 2 approaches. The anchor-based CID estimate was calculated as the mean change score on the WMFT, corresponding to patients who were defined as having minimal CID; that is, those with their FMA change score of 6 to 10 points. In addition, the distribution-based CID estimate was determined using the Cohen effect size33 benchmark, which is widely accepted.34-36 An effect size of 0.2 was advocated as a reasonable method to estimate the minimal CID.35,36 Thus, an effect size of 0.2 (ie, 0.2 SD of baseline score), indicating small but important change, was used to estimate minimal threshold of CID in this study.

MDC and CID proportions. To assess the extent of patients’ changes after interventions detected by the WMFT, the proportions of patients with change scores exceeding the values of the MDC90 and CID estimates were examined. If the patients did not benefit from the treatment, they were excluded from the estimation of responsiveness. The greater the proportions of patients who exceeded the values, the more responsive the measure is.

Results

The clinical and demographic characteristics of the 57 participants (39 men, 18 women; mean age, 55 years) are summarized in Table 1. The mean scores of each item at baseline were 7.05 seconds on the WMFT time and 3.22 points on the WMFT FAS.

Results of the MDC90 and CID estimates of the WMFT are given in Table 2. The MDC90 of the WMFT time and WMFT FAS was 4.36 and 0.37, respectively. As calculated from the 22 patients whose FMA change score reached 6 to 10 points, the anchor-based CID estimate was 1.64 and 0.33 for the WMFT time and WMFT FAS, respectively. In addition, the distribution-based CID estimates (ie, 0.2 SD) for the WMFT time and WMFT FAS equated to a change of 1.37 and 0.14, respectively.

Table 1

Clinical and Demographic Characteristics of the Participants (N = 57)

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender (male/female)</td>
<td>39/18</td>
</tr>
<tr>
<td>Age, mean (SD), y</td>
<td>54.6 (11.5)</td>
</tr>
<tr>
<td>Side of stroke, no. (%)</td>
<td></td>
</tr>
<tr>
<td>Left</td>
<td>31 (54.4)</td>
</tr>
<tr>
<td>Right</td>
<td>26 (45.6)</td>
</tr>
<tr>
<td>Time since stroke, mean (SD), month</td>
<td>12.98 (7.62)</td>
</tr>
<tr>
<td>Brunnstrom stage of the proximal UE (median)</td>
<td>5</td>
</tr>
<tr>
<td>Mini-Mental State Examination, mean (SD)</td>
<td>28.54 (1.56)</td>
</tr>
<tr>
<td>WMFT time at pretreatment, mean (SD), s</td>
<td>7.05 (6.85)</td>
</tr>
<tr>
<td>WMFT FAS at pretreatment, mean (SD)</td>
<td>3.22 (0.72)</td>
</tr>
<tr>
<td>WMFT time at posttreatment, mean (SD), s</td>
<td>4.91 (4.78)</td>
</tr>
<tr>
<td>WMFT FAS at posttreatment, mean (SD)</td>
<td>3.52 (0.74)</td>
</tr>
</tbody>
</table>

Abbreviations: UE, upper extremity; WMFT FAS, functional ability scale of the Wolf Motor Function Test; WMFT time, performance time of the Wolf Motor Function Test.

Table 2

The Minimal Detectable Change and Clinically Important Difference Estimates of the Wolf Motor Function Test

<table>
<thead>
<tr>
<th>Scale</th>
<th>SD</th>
<th>(r)</th>
<th>SEM</th>
<th>MDC90</th>
<th>CID Estimates</th>
</tr>
</thead>
<tbody>
<tr>
<td>WMFT time</td>
<td>5.91</td>
<td>0.90</td>
<td>1.87</td>
<td>4.36</td>
<td>1.64</td>
</tr>
<tr>
<td>WMFT FAS</td>
<td>0.73</td>
<td>0.95</td>
<td>0.16</td>
<td>0.37</td>
<td>0.33</td>
</tr>
</tbody>
</table>

Abbreviations: CID, clinically important difference; MDC90, minimal detectable change at 90% confidence; SD, standard deviation; SEM, standard error of measurement; WMFT FAS, functional ability scale of the Wolf Motor Function Test; WMFT time, performance time of the Wolf Motor Function Test.

WMFT time and WMFT FAS, respectively. In addition, the distribution-based CID estimates (ie, 0.2 SD) for the WMFT time and WMFT FAS equated to a change of 1.37 and 0.14, respectively.

The mean group change scores were 2.14 for the WMFT time and 0.30 for the WMFT FAS. As summarized in Table 3, 14% and 38.6% of the patients had a positive change that exceeded the MDC90 of the WMFT time and WMFT FAS, respectively. Furthermore, there were 26.3% and 52.6% of the patients whose changes achieved the anchor-based CID estimates of the WMFT time and WMFT FAS. Approximately one third and two thirds of patients’ change scores exceeded the 0.2 SD of the WMFT time and WMFT FAS, respectively.

Discussion

To our knowledge, this is the first study to investigate the MDC and CID of the WMFT in stroke rehabilitation. Our
results provide useful benchmarks for clinicians to interpret whether patients’ change score on the WMFT after stroke rehabilitation can be interpreted as true or clinically meaningful changes and to make clinical judgments about the patients’ progress. The analyses show that the MDC90 was 4.36 and 0.37 for the WMFT time and WMFT FAS, respectively. It indicates that when the change scores of an individual stroke patient between 2 measurements reach 4.36 seconds and 0.37 points on the WMFT time and WMFT FAS, the clinicians may interpret the changes as true and reliable (ie, beyond measurement error), given the 90% confidence level. The requirements of measurement error for individual-level use in clinical practice are higher than that for group-level interpretation that establishes in this study, indicating that the results can be interpreted not only as statistically significant, but also as clinically important. Moreover, our results showed that the mean change WMFT time and WMFT FAS score was 2.14 seconds and 0.3 points, which also reached the threshold of clinical importance.

Table 3

<table>
<thead>
<tr>
<th>Scale</th>
<th>Proportion Exceeding the MDC$_{90}$</th>
<th>Proportion Exceeding the Anchor-Based CID Estimate</th>
<th>Proportion Exceeding the 0.2 SD(^a)</th>
</tr>
</thead>
<tbody>
<tr>
<td>WMFT time</td>
<td>14% (8/57)</td>
<td>26.3% (15/57)</td>
<td>29.8% (17/57)</td>
</tr>
<tr>
<td>WMFT FAS</td>
<td>38.6% (22/57)</td>
<td>52.6% (30/57)</td>
<td>64.9% (37/57)</td>
</tr>
</tbody>
</table>

Abbreviations: CID, clinically important difference; MDC$_{90}$, minimal detectable change at 90% confidence; SD, standard deviation; WMFT FAS, functional ability scale of the Wolf Motor Function Test; WMFT time, performance time of the Wolf Motor Function Test.

\(^a\)Square root of [(SD$_{pre-visit}$)$^2 + (SD$_{post-visit}$)$^2 / 2$.

Of note, there is concern over the differences between group and individual clinical importance. Average effects across a group may not be meaningful to the individual patient. Group-derived CID values are suitable to interpret the results of clinical trials or group studies, but they often are not directly applied to interpret the individual’s change. For individual-level use, it may be reasonable to expect that the MDC would be less than or equal to the minimal CID. However, some researchers have suggested that this may not always be the case. In our study, the MDC of the WMFT FAS (0.37) was similar to the minimal CID of the WMFT FAS (0.2-0.4), whereas the MDC of WMFT time (4.37) was higher than the corresponding minimal CID (1.5-2). When the MDC exceeds the minimal CID, both values are suggested to be considered in clinical decision making. We provided the proportions of study participants exceeding the minimal threshold of MDC and CID for a comparison of the 2 measures. The proportions of the participants exceeding the threshold of MDC were lower than the proportions exceeding the CID for both subscales of the WMFT, indicating the need to consider both levels of analysis (ie, individual vs group) when interpreting clinical significance of change on this test. Further research is needed to identify the optimal methods for establishing the values and to examine the relationship of the MDC and minimal CID.

We used both the MDC and CID estimates to interpret the change scores on the WMFT of our patients who underwent stroke rehabilitation. Reporting the proportion of patients that met the MDC and CID requirements provided more insightful and intuitive clinical interpretations than considering the overall mean change scores. In addition, the proportions can be used in the determination of responsiveness of outcome measures, analogous to the way used to calculate the number needed to treat in the evidence-based practice terminology. On the basis of our results, 14% and 39% of patients achieved a degree of improvement beyond measurement error on the WMFT time and WMFT FAS. Nearly one third and more than half of the patients, respectively, exceeded the minimal CID of the WMFT time and WMFT FAS, depending on the criteria. Because more patients exceeded the values of the MDC and CID of the WMFT FAS than the values of the WMFT time, the...
WMFT FAS appears more responsive to detect changes than the WMFT time.

Some potential limitations of this study should be mentioned. First, previous studies have pointed out that the MDC and CID may vary, depending on the patient’s baseline level of disability. A larger sample of stroke patients differing in level of motor impairment is necessary to verify whether the MDC and CID of the WMFT differ in accordance with the severity of impairment. Second, the minimal CID may be affected by the direction of change (ie, getting better or worse). Because our CID estimates of the WMFT were derived from patients who improved in their performance, the CID results are only applicable for patients who improved. Third, the results of this study may not be generalized to stroke patients with cognitive impairments because we recruited only patients with a MMSE score > 24.

Conclusions

This study provides a preliminary investigation of the MDC and CID of the WMFT, which is useful for interpretation of change scores in stroke patients. The MDC and CID estimates add information about clinimetric properties of the WMFT and may facilitate the interpretation of performance change on the test after stroke rehabilitation. Our findings suggest that the change score of an individual patient has to reach 4.36 and 0.37 on the WMFT time and WMFT FAS, respectively, to indicate a true change. If the mean change scores within a stroke group fall into a range of 1.5 to 2 on the WMFT time and 0.2 to 0.4 on the WMFT FAS, the changes may be clinically important. Furthermore, the WMFT FAS appears more responsive than the WMFT time based on the proportions of patients who met the MDC and CID criteria. Further research based on a larger sample is warranted to establish the clinimetric properties of the WMFT in stroke rehabilitation.

Acknowledgments

This project was supported in part by the National Science Council (NSC 96-2628-B-002-033-MY2, NSC 96-2320-B-182-029, and NSC 97-2314-B-002-008-MY3) and the National Health Research Institutes (NHRI-EX97-9742PJ) in Taiwan.

References