Neurorehabilitation and Neural Repair
Volume 23 Number 5
June 2009 429-434

10.1177/1545968308331144
http://nnr.sagepub.com

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 (MDC_{90}), 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 MDC_{90} 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[1] 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.[2-6] Although the reliability and validity of the WMFT have been established in subacute and chronic stroke,[7-10] properties of the test to detect changes after interventions have not been examined yet.

Clinimetrics, one of methodological disciplines introduced by Feinstein,[11] 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.[11-13] 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.[12]

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.[14] 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.[15] One advantage of the MDC is that it considers both the reliability and responsiveness to change.[16,17] That is, the MDC helps clinicians determine whether the change score in individual performance represents real and reliable change[17,18] and simultaneously tells whether the outcome measure will be able to detect such a change.[16]

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.[19] 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,[20,21] have been proposed to determine the CID. Because agreement has not yet been reached on a gold standard approach,[22] combinations of the anchor-based and distribution-based approaches

From the School of Occupational Therapy, College of Medicine, National Taiwan University, Taipei, Taiwan (KCL, YWH, YJ); Division of Occupational Therapy, Department of Physical Medicine and Rehabilitation, National Taiwan University Hospital, Taipei, Taiwan (KCL, YJ); Graduate Institute of Clinical Behavioral Science and Department of Occupational Therapy, Chang Gung University, Tao-yuan, Taiwan (CYW); Department of Physical Medicine and Rehabilitation, Chang Gung Memorial Hospital, Tao-yuan, Taiwan (CLC); and Department of Medical Education and Department of Surgery, Cathay General Hospital, Taipei, Taiwan (JSL). Address correspondence to Ching-yi Wu, ScD, OTR, Graduate Institute of Clinical Behavioral Science and Department of Occupational Therapy, Chang Gung University, 259 Wen-hwa 1st Road, Kwei-shan, Tao-yuan, Taiwan. E-mail: cywu@mail.cgu.edu.tw.

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 a 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 III[27] 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 score[28] $\leq$ 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 hand 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

level and the square root of 2, reflecting the additional uncertainty introduced from 2 difference measurements.[16] The SEM quantifies the within-subject variability and reflects the amount of error in measurements.[16,25] Furthermore, the MDC estimate is most often based on 90% CI ($z = 1.65$) or 95% CI ($z = 1.96$). MDC_{90} means that 90% of stable patients demonstrate a random variation of less than this amount when tested on 2 occasions. In other words, if a patient has a change score equal to or above the MDC_{90} threshold, it is possible to state with 90% confidence that the change is reliable rather than measurement error.[15] Therefore, a change greater than the MDC is interpreted as true change.

The MDC_{90} was calculated using the following formula: $MDC_{90} = 1.65 \times \sqrt{2} \times SEM = 1.65 \times SD \times \sqrt{(2\ [1 - r])}$, where 1.65 is the 2-tailed tabled z value for the 90% confidence interval, SD is the standard deviation, $\sqrt{2}$ represents the variance of 2 measurements, and 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 size[33] 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 MDC_{90} 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 MDC_{90} and CID estimates of the WMFT are given in Table 2. The MDC_{90} 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)

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

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

					CID Estimates	
Scale	SD	r	SEM	MDC_{90}	Anchor-Based	0.2 SD[a]
WMFT time	5.91	0.90	1.87	4.36	1.64	1.37
WMFT FAS	0.73	0.95	0.16	0.37	0.33	0.14

Abbreviations: CID, clinically important difference; MDC_{90}, 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.
[a]Square root of $[(SD_{pretreatment})^2 + (SD_{posttreatment})^2 / 2]$.

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 MDC_{90} 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

Table 3
Proportions of Patients Who Met the Criteria of the Minimal Detectable Change and Clinically Important Difference on the Wolf Motor Function Test

Scale	Proportion Exceeding the MDC_{90}	Proportion Exceeding the Anchor-Based CID Estimate	Proportion Exceeding the 0.2 SD[a]
WMFT time	14% (8/57)	26.3% (15/57)	29.8% (17/57)
WMFT FAS	38.6% (22/57)	52.6% (30/57)	64.9% (37/57)

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 $[(SD_{pretreatment})^2 + (SD_{posttreatment})^2 / 2]$.

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 MDC_{90} 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 looks at the average of a group of patients.[13,16] As in our data, the MDCs of the WMFT used for individual patients are higher than the SEMs used for a group of patients.

Because different minimal CID estimates may be obtained from different methods, triangulation of the results to identify the CID is suggested.[14,23] We adopted the combination of the anchor-based and distribution-based approaches. This practice has been advocated for use in recent studies.[34,37-39] In this present study, the minimal CID estimate derived from an anchor of clinical measure was 1.64 and 0.33, and the distribution-based minimal CID was 1.37 and 0.14 for the WMFT time and WMFT FAS, respectively. Combining the 2 approaches to establish the minimal CID, we recommend that the changes have to be in the range of 1.5 to 2 seconds on the WMFT time and 0.2 to 0.4 points on the WMFT FAS to meet the requirements for minimal CID. In other words, if a stroke group achieves a mean score of 1.5 to 2 seconds on the WMFT time or a mean score of 0.2 to 0.4 points on the WMFT FAS, they are likely to have clinically important change in their speed or quality of movements while performing the tasks of the WMFT. For instance, a previous study of constraint-induced therapy (the EXCITE Trial)[5] used the WMFT as an outcome measure to evaluate the intervention effects of the therapy. Their data showed that the mean WMFT time and WMFT FAS improvement scores from preintervention to postintervention were 8.5 seconds and 0.3 points, respectively.[5] The mean change scores exceeded the values of minimal CIDs established 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. Since the EXCITE trial was a large trial, the statistically and clinically significant results may enhance the validity of this study's conclusions.

Of note, there is concern over the differences between group and individual clinical importance.[40] Average effects across a group may not be meaningful to the individual patient.[16] Group-derived CID values[40,41] are suitable to interpret the results of clinical trials or group studies, but they are often 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.[42,43] However, some researchers have suggested that this may not always be the case.[42,44] 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.[43] 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.[42,43]

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.[14,16] 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.[16] 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.[14,44,45] 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).[23] 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-EX97-9742PI) in Taiwan.

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. Caimmi M, Carda S, Giovanzana C, et al. Using kinematic analysis to evaluate constraint-induced movement therapy in chronic stroke patients. *Neurorehabil Neural Repair*. 2008;22:31-39.
3. Kunkel A, Kopp B, Muller G, et al. Constraint-induced movement therapy for motor recovery in chronic stroke patients. *Arch Phys Med Rehabil*. 1999;80:624-628.
4. Page SJ, Sisto SA, Levine P, Johnston MV, Hughes M. Modified constraint induced therapy: a randomized feasibility and efficacy study. *J Rehabil Res Dev*. 2001;38:583-590.
5. Wolf SL, Winstein CJ, Miller JP, et al. 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.
6. Park SW, Wolf SL, Blanton S, Winstein C, Nichols-Larsen DS. The EXCITE trial: predicting a clinically meaningful motor activity log outcome. *Neurorehabil Neural Repair*. 2008;22:486-493.
7. Morris DM, Uswatte G, Crago JE, Cook EW 3rd, 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.
8. Whitall 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.
9. 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.
10. Wolf SL, Thompson PA, Morris DM, et al. The EXCITE trial: attributes of the Wolf Motor Function Test in patients with subacute stroke. *Neurorehabil Neural Repair*. 2005;19:194-205.
11. Feinstein AR. *Clinimetrics*. New Haven, CT: Yale University Press; 1987.
12. de Vet HCW, Terwee CB, Bouter LM. Clinimetrics vs psychometrics: two sides of the same coin. *J Clin Epidemiol*. 2003;56:1146-1147.
13. de Vet HC, Terwee CB, Bouter LM. Current challenges in clinimetrics. *J Clin Epidemiol*. 2003;56:1137-1141.
14. Haley SM, Fragala-Pinkham MA. Interpreting change scores of tests and measures used in physical therapy. *Phys Ther*. 2006;86:735-743.
15. Portney LG, Watkins MP. *Foundations of Clinical Research: Applications to Practice*. 3rd ed. Upper Saddle River, NJ: Pearson/Prentice Hall; 2009.
16. Schmitt JS, Di Fabio RP. Reliable change and minimum important difference (MID) proportions facilitated group responsiveness comparisons using individual threshold criteria. *J Clin Epidemiol*. 2004;57:1008-1018.
17. Beckerman H, Roebroeck ME, Lankhorst GJ, Becher JG, Bezemer PD, Verbeek AL. Smallest real difference, a link between reproducibility and responsiveness. *Qual Life Res*. 2001;10:571-578.
18. Flansbjer UB, Holmback AM, Downham D, Lexell J. What change in isokinetic knee muscle strength can be detected in men and women with hemiparesis after stroke? *Clin Rehabil*. 2005;19:514-522.
19. Beaton DE, Boers M, Wells GA. Many faces of the minimal clinically important difference (MCID): a literature review and directions for future research. *Curr Opin Rheumatol*. 2002;14:109-114.
20. Copay AG, Subach BR, Glassman SD, Polly DW, Schuler TC. Understanding the minimum clinically important difference: a review of concepts and methods. *Spine J*. 2007;7:541-546.
21. Crosby RD, Kolotkin RL, Williams GR. Defining clinically meaningful change in health-related quality of life. *J Clin Epidemiol*. 2003;56:395-407.
22. Kelleher CJ, Pleil AM, Reese PR, Burgess SM, Brodish PH. How much is enough and who says so? *BJOG*. 2004;111:605-612.
23. Hays RD, Woolley JM. The concept of clinically meaningful difference in health-related quality-of-life research. How meaningful is it? *Pharmacoeconomics*. 2000;18:419-423.
24. Revicki DA, Cella D, Hays RD, Sloan JA, Lenderking WR, Aaronson NK. Responsiveness and minimal important differences for patient reported outcomes. *Health Qual Life Outcomes*. 2006;4:70.
25. Wagner JM, Rhodes JA, Patten C. Reproducibility and minimal detectable change of three-dimensional kinematic analysis of reaching tasks in people with hemiparesis after stroke. *Phys Ther*. 2008;88:652-663.
26. Dobkin B. Progressive staging of pilot studies to improve phase III trials for motor interventions. *Neurorehabil Neural Repair*. In press.
27. Brunstrom S. *Movement Therapy in Hemiplegia*. New York, NY: Harper and Row; 1970.

28. Bohannon RW, Smith MB. Interrater reliability of a Modified Ashworth Scale of muscle spasticity. *Phys Ther*. 1987;67:206-207.
29. Folstein MF, Folstein SE, McHugh PR. "Mini-mental state." A practical method for grading the cognitive state of patients for the clinician. *J Psychiatr Res*. 1975;12:189-198.
30. Fugl-Meyer AR, Jaasko L, Leyman I, Olsson S, Steglind S. The poststroke hemiplegic patient. 1. A method for evaluation of physical performance. *Scand J Rehabil Med*. 1975;7:13-31.
31. 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.
32. 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.
33. Cohen J. *Statistical Power Analysis for the Behavior Sciences*. 2nd ed. Hillsdale, NJ: Lawrence Erlbaum Associates; 1988.
34. Eton DT, Cella D, Yost KJ, et al. A combination of distribution- and anchor-based approaches determined minimally important differences (MIDs) for four endpoints in a breast cancer scale. *J Clin Epidemiol*. 2004; 57:898-910.
35. Hawkes WG, Williams GR, Zimmerman S, et al. A clinically meaningful difference was generated for a performance measure of recovery from hip fracture. *J Clin Epidemiol*. 2004;57:1019-1024.
36. Samsa G, Edelman D, Rothman ML, Williams GR, Lipscomb J, Matchar D. Determining clinically important differences in health status measures: a general approach with illustration to the health utilities index mark ii. *Pharmacoeconomics*. 1999;15:141-155.
37. Crosby RD, Kolotkin RL, Williams GR. An integrated method to determine meaningful changes in health-related quality of life. *J Clin Epidemiol*. 2004;57:1153-1160.
38. Yost KJ, Sorensen MV, Hahn EA, Glendenning GA, Gnanasakthy A, Cella D. Using multiple anchor- and distribution-based estimates to evaluate clinically meaningful change on the Functional Assessment of Cancer Therapy-Biologic Response Modifiers (FACT-BRM) instrument. *Value Health*. 2005;8:117-127.
39. Hsieh YW, Wang CH, Wu SC, Chen PC, Sheu CF, Hsieh CL. Establishing the minimal clinically important difference of the Barthel Index in stroke patients. *Neurorehabil Neural Repair*. 2007;21:233-238.
40. Cella D, Bullinger M, Scott C, et al. Group vs individual approaches to understanding the clinical significance of differences or changes in quality of life. *Mayo Clin Proc*. 2002;77:384-392.
41. Dawson J, Doll H, Coffey J, et al. Responsiveness and minimally important change for the Manchester-Oxford foot questionnaire (MOXFQ) compared with AOFAS and SF-36 assessments following surgery for hallux valgus. *Osteoarthritis Cartilage*. 2007;15:918-931.
42. Fritz JM, Irrgang JJ. A comparison of a modified Oswestry Low Back Pain Disability Questionnaire and the Quebec Back Pain Disability Scale. *Phys Ther*. 2001;81:776-788.
43. Resnik L, Dobrykowski E. Outcomes measurement for patients with low back pain. *Orthop Nurs*. 2005;24:14-24.
44. Stratford PW, Binkley J, Solomon P, Finch E, Gill C, Moreland J. Defining the minimum level of detectable change for the Roland-Morris questionnaire. *Phys Ther*. 1996;76:359-365.
45. Testa MA. Interpretation of quality-of-life outcomes: issues that affect magnitude and meaning. *Med Care*. 2000;38:II166-174.