Published in final edited form as:

Arch Phys Med Rehabil. 2008 September ; 89(9): 1693. doi:10.1016/j.apmr.2008.02.022.

Estimating minimal clinically important differences of upper extremity measures early after stroke

Catherine E. Lang, PT, PhD^{1,2,3}, Dorothy F. Edwards, PhD^{2,3,4}, Rebecca L. Birkenmeier, MS, OTR/L², and Alexander W. Dromerick, MD⁵

¹ Program in Physical Therapy, Washington University School of Medicine, St. Louis, MO

² Program in Occupational Therapy, Washington University School of Medicine, St. Louis, MO

³ Department of Neurology, Washington University School of Medicine, St. Louis, MO

⁴ Department of Kinesiology, University of Wisconsin, Madison WI

⁵ National Rehabilitation Hospital, Washington DC

Abstract

Objective—To estimate minimal clinically important difference values of several upper extremity measures early after stroke.

Design—Data in this report were collected during the VECTORS trial, an acute, single-blind randomized controlled trial of Constraint Induced Movement Therapy. Subjects were tested at the pre-randomization baseline assessment (average of 9.5 days post stroke), and the first post-treatment assessment (25.9 days post stroke). At each time point, the affected upper extremity was evaluated with a battery of 6 tests. At the second assessment, subjects were also asked to provide a global rating of perceived changes in their affected upper extremity. Anchor-based minimal clinically important difference values were calculated separately for the affected dominant upper extremities and the affected non-dominant upper extremities for each of the 6 tests.

Setting—Inpatient rehabilitation hospital.

Participants—Fifty-two people with hemiparesis post stroke.

Interventions—Not applicable.

Main Outcome Measures—Estimated minimal clinically important difference values for grip strength, composite upper extremity strength, Action Research Arm Test (ARAT), Wolf Motor Function Test (WMFT), Motor Activity Log (MAL), and duration of upper extremity use as measured with accelerometry.

Results—Minimal clinically important difference values for grip strength were 5.0 and 6.2 kg for the affected dominant and non-dominant sides respectively. Minimal clinically important difference values for the ARAT were 12 and 17 points, for the WMFT Function score were 1.0 and 1.2 points, and for the MAL How well score were 1.0 and 1.1 points for the two sides respectively. Minimal clinically important difference values were indeterminate for the dominant (composite strength), the non-dominant (WMFT Time score) or for both affected sides (duration of use) for the other measures.

Corresponding Author: Catherine E. Lang, PT, PhD, Program in Physical Therapy, Washington University School of Medicine, Campus Box 8502, 4444 Forest Park, St. Louis, MO 63108, Fax: 314.286.1410, Office: 314.286.1945, langc@wustl.edu.

We certify that no party having a direct interest in the results of the research supporting this article has or will confer a benefit on us or on any organization with which we are associated AND, if applicable, we certify that all financial and material support for this research (eg, NIH or NHS grants) and work are clearly identified in the title page of the manuscript.

Conclusions—Our data provide some of the first estimates of minimal clinically important difference values for upper extremity standardized measures early after stroke. Future studies with larger sample sizes are needed to refine these estimates and to determine if minimal clinically important difference values are modified by time post stroke.

Keywords

outcome assessment; minimal clinically important difference; upper extremity; rehabilitation; stroke

INTRODUCTION

Measurement in rehabilitation is an ongoing concern. For post stroke rehabilitation, there are now a respectable number of standardized clinical measures that assess upper extremity deficits in people with hemiparesis.^{1, 2} These clinical measures assess upper extremity deficits post stroke at the impairment level (e.g. grip strength, Fugl-Meyer), at the activity limitation level (e.g. Action Research Arm Test, Wolf Motor Function Test), at the participation restriction level (Motor Activity Log), or at multiple levels simultaneously (e.g. Stroke Impact Scale). Over the past few decades, the body of literature examining psychometric properties of upper extremity measures has grown tremendously. For many measures, reliability and validity have been sufficiently established for recommended use in daily clinical practice and in research. A newer focus in the studies of clinical measurement post stroke is the concept of the minimal clinically important difference, or MCID.¹

The MCID has been defined as “the smallest difference in score in the domain of interest which patients perceive as beneficial and which would mandate, in the absence of troublesome side effects and excessive cost, a change in the patient’s management”.³ MCID values are therefore important in interpreting the clinical relevance of observed changes, at both the individual and group levels. MCID values can be derived a number of ways, where the various methods are often categorized as either distribution-based or anchor-based.⁴ The distribution-based methods estimate MCID values based on the statistical characteristics of change scores within a sample. A limitation of the distribution-based methods is that the derived values do not typically indicate the importance of the score. The anchor-based methods estimate MCID values by comparing change scores with an “anchor”, usually either a patient or a clinician rating of change. A benefit of the anchor-based method is that the estimate is based on changes that are considered important to the patients or clinicians.

It is critical to appreciate that there is no single “true” MCID value for a given measure. MCID values are dynamic and context-specific.^{5, 6} In the area of stroke rehabilitation, factors that may affect MCID values include time since stroke, magnitude of initial deficits, and patient expectations of recovery. Published estimates of MCID values are available for two global activity level measures, the Functional Independence Measure and the Barthel Index^{7–9}, but we could find no data-based estimates of MCID values for measures that assess upper extremity deficits in people with hemiparesis post stroke. It is therefore time to begin to develop estimates of MCID values in these measures, particularly within the first few weeks and months after stroke, when most rehabilitation services are provided and most rehabilitation monies are spent.^{10, 11}

The purpose of this report was to estimate MCID values of several upper extremity measures early after stroke. We chose to estimate MCID values using a patient anchor-based method instead of a distribution based method^{6, 12} because the anchor-based method directly reflects the point of view of the patients.^{3, 13} We consider the patients’ view the gold-standard for judging important changes, particularly for measures that are trying to quantify constructs such as functional abilities. We performed separate analyses for affected dominant and non-

dominant upper extremities because estimates of MCID values could be influenced by whether the dominant versus non-dominant upper extremity was affected by the stroke. A secondary purpose of this report was to see if the magnitude of the MCID estimates was similar across scales, i.e. a similar proportion of the scale. This was worth evaluating, because if estimates of MCID were similar, then clinicians might be able to use a proportional rule to estimate MCID values on other, conceptually similar measures.¹³

METHODS

Subjects

Fifty-two subjects with hemiparesis post stroke were enrolled in this study (table 1). Hemiparetic subjects were participating in VECTORS (Very Early Constraint-induced Therapy for Recovery of Stroke), a single-center randomized control trial investigating early motor recovery of the UE following stroke conducted at the Washington University School of Medicine. Subjects were recruited via the Cognitive Rehabilitation Research Group Stroke Registry from the acute neurology service of Barnes Jewish Hospital and from the rehabilitation service of the Rehabilitation Institute of Saint Louis. In all, 1850 patients with stroke were screened to achieve the enrolled 52 subjects. Subjects were included in the trial if they had: 1) an ischemic or hemorrhagic stroke within 28 days of admission to inpatient rehabilitation; 2) persistent hemiparesis as indicated by a score of 1 – 3 on the motor arm item of the National Institutes of Health Stroke Scale (NIHSS); 3) the presence of some upper extremity voluntary activity as indicated by the ability to move proximal and/or distal joints against gravity; 4) evidence of preserved cognitive function, as indicated by scores of 0 or 1 on the consciousness items of the NIHSS and a score of greater than 19 on the Short Blessed Memory Orientation and Concentration test¹⁴; 5) the ability to follow 2 step commands, as indicated by a score of 0 on the item 1c of the NIHSS; and 6) no upper extremity injury or conditions that limited use prior to the stroke. Subjects were excluded from the trial if they: 1) could not give informed consent; 2) had clinically significant fluctuations in mental status in the 72 hours prior to enrollment; 3) had hemispatial neglect as assessed by the Star Cancellation Test; and/or 4) were not expected to survive 1 year due to other illnesses (e.g. cardiac disease, malignancy). Subject characteristics are provided in table 1; data from this cohort have been reported previously.^{15, 16} The Human Studies Committee of Washington University School of Medicine approved the protocol for this study. Written informed consent was obtained from all subjects prior to testing.

Protocol

The data reported in this study are from the first two VECTORS evaluations: the pre-randomization baseline assessment (study day 0, an average of 9.5 days post stroke), and the first post-treatment assessment, (study day 14, an average of 25.9 days post stroke). Consistent with other reports investigating MCIDs, data from all subjects were pooled, regardless of group assignment in the trial.^{3, 17–19} All evaluations were performed by trained study personnel who were blinded to group assignment. At each evaluation session, the affected (contralateral to the lesion) upper extremity was evaluated using a battery of tests, as described below. At the second evaluation session, subjects were also asked to provide a global rating of perceived changes in their affected upper extremity by comparing “how well your arm is doing” using the following 7-point Likert scale.

- Score 1 = Much better
- Score 2 = A little better, meaningful
- Score 3 = A little better, not meaningful
- Score 4 = About the same

- Score 5 = A little worse, not meaningful
- Score 6 = A little worse, meaningful
- Score 7 = Much worse

We chose a 7 point scale^{19, 20} to evaluate our subjects' ratings of change instead of a larger 15 point scale³ because a portion of the sample were not literate and may have had difficulty understanding the finer gradations of the 15 point scale. The scores on this item were then used as the anchors for grouping subjects during the calculation of MCID estimates.

Upper extremity evaluation

The affected upper extremity was evaluated using a battery of tests that included measurement at the level of impairments, activity limitations, and participation restrictions. Impairment measures included in this report are grip strength and composite upper extremity strength. Activity measures included in this report are the Action Research Arm Test and the Wolf Motor Function Test. Participation measures included in this report are the Motor Activity Log and the duration of arm use measured with accelerometers.

Grip Strength—Affected hand grip strength is a common impairment reliably assessed by a variety of stroke professionals²¹, and has been proposed as a surrogate measure for upper extremity outcomes.²² Here, grip strength was assessed during the Wolf Motor Function Test (see below) via a dynamometer measurement of the maximum amount of force produced during a palmar grip.²³ Subjects were seated using with the upper extremity in 0 degrees of shoulder flexion and 90 degrees of elbow flexion. A Jamar grip dynamometer (Sammons Preston Rolyan, Bolingbrook IL) was used with the handle position set at 3 for all measurements for all subjects. At each session, three measurements were taken and then the average value was used in the analyses. Grip strength values were expressed in kilograms.

Composite upper extremity strength—Strength is one of the most often assessed impairments post stroke. Here, strength of the shoulder, elbow, and wrist flexor and extensors were measured using a hand-held dynamometer (MICROFET2, Hogan Health Industries, Draper UT) following a standard protocol²⁴ except that subjects were seated during testing. Warm-up or practice trials were not performed prior to testing. Maximal voluntary isometric strength values were recorded in lbs. for each muscle group tested. Subjects unable to produce force against the dynamometer were given a score of 0 lbs. for that particular muscle group. The strength of each muscle group was expressed as the ratio of affected side to unaffected side maximal isometric force. Ratio values from each muscle group were averaged to form a single, composite score for the affected upper extremity for each subject.^{25, 26}

Action Research Arm Test (ARAT)—The ARAT assesses activity limitations of the upper extremity.²⁷ It includes 19 items divided into four subscales: grasp, grip, pinch, and gross movement. Reliability (interrater 0.99, test-retest 0.98) and validity of the ARAT have been well established.^{15, 27–31} Performance on the ARAT is strongly correlated to performance on the upper extremity motor portion of the Fugl-Meyer scale and to performance on the Box and Block test.^{31, 32} Item scores on the ARAT are summed to create subtest and full-scale scores with a maximum score of 57, indicating normal performance.

Wolf Motor Function Test (WMFT)—The WMFT is a 17-item measure used to assess activity limitations of the upper extremity. It is comprised of 2 strength items and 15 timed task performance items. The task performance items begin with the measurement of simple proximal movements and progress to more complex distal and whole limb movements. The WMFT yields two scores: 1) a functional ability score quantifying quality of performance, and 2) a timed score quantifying speed of performance in seconds. The test has published reliability

and validity.^{33–36} In the VECTORS study, the key use task was not collected because of difficulties with instrumentation, and the results reported do not include this item.

Motor Activity Log (MAL)—The MAL is a measure of self-perceived upper extremity participation restrictions.^{37–39} It uses a semi-structured interview to assess how much and how well patients use their affected arm for activities of daily living (ADLs) over a specified period of time. Thirty specific ADL tasks are evaluated using a 6-point amount of use scale (how much) and a 6-point quality of movement scale (how well). The tasks include activities such as buttoning a shirt, brushing teeth, and using a key. Adequate inter-rater reliability (> 0.91) and internal consistency ($\alpha > 0.81$) have been reported.³⁹ The quality of movement scale, or how well the arm functioned in the ADL tasks, was collected at both study day 0 and day 14 assessments and is therefore included in this report. The amount of use scale, or how much the arm was used, was not administered at study day 0, making it impossible to calculate change scores and include in this report.

Duration of use—The duration of use in a 24 hour period is an upper extremity-specific measure of participation. We have previously reported on duration of upper extremity use in this cohort.⁴⁰ Briefly, duration of use was captured using uni-axial accelerometers (model 7164-2.4 Activity Monitors, MTI Health Services, Fort Walton Beach FL). Accelerometers were placed on the distal arm just above the wrist with the axis parallel to the length of the arm. Subjects wore the accelerometers at all times during a 24 hour data collection period except when the devices would be exposed to water (e.g. personal hygiene). Data were collected in 2-second epochs over the 24 hours. Using an established methodology that provides a valid (r values 0.93–0.99) and reliable (test-retest 0.90) measure of the duration of upper extremity use^{41, 42}, data from each 2 second epoch were used to classify the upper extremity as either moving or not moving during that 2 second period. The sum of the epochs when the upper extremity moved then represented the duration of upper extremity movement over the 24 hour period. For ease of communication, this summed variable was converted from seconds to hours. It should be noted that accelerometer measures of upper extremity use have some limitations since they provide information about whether or not the upper extremity was moving, but not information about what it was moving to do (e.g. functional task versus arm swing during gait). Thus, the duration of use obtained via accelerometry may be considered small overestimations of the time spent using the upper extremity for functional activities.⁴⁰

Analyses

SPSS for Windows Version 13.0 was used for all statistical analyses. Individual change scores were calculated for each measure by subtracting scores at study day 0 from scores at study day 14. Change scores for each measure were normally distributed as evaluated by Kolmogorov-Smirnov tests.

The magnitude of change that is considered meaningful may be influenced by whether the dominant versus non-dominant upper extremity was affected by the stroke. Data from affected dominant upper extremities were therefore analyzed separately from data from affected non-dominant upper extremities. To estimate the MCIDs on each measure, mean change scores for each perceived change rating were computed. Consistent with other recent reports using anchor-based methods in this population^{8, 9}, the mean change score for the smallest meaningful change (here = score of 2, “a little better, meaningful”) was taken as the MCID. In addition to expressing the MCID as a change in raw scores on each measure, we also calculated the MCID as a percentage of the total scale (when possible) and as a single-population effect size⁴³, where the change score is divided by the standard deviation of the study day 0 score. Expressing the MCIDs in these alternative formats allows for comparison of MCIDs across the measures evaluated.

RESULTS

Fifty-two subjects with hemiparesis post stroke were included in this report (table 1). Subjects were an average of 64 ± 14 years old and had been largely independent prior to the stroke. The first evaluations occurred an average of 9.5 days post stroke (study day 0) and the second evaluations occurred an average of 25.9 days post stroke (study day 14). The sample had a greater proportion of females than males and was 58% non-white. Seventy-nine percent of subjects had ischemic strokes and 56% of subjects were affected on their non-dominant side.

Scores on each upper extremity measure from the study day 0 and day 14 time points are provided in table 2. We considered our sample to be moderately affected at the day 0 time point and mild-to-moderately affected at the day 14 time point, although there was a fair amount of variation in performance which can be seen in the standard deviations of each measure. As expected in an inpatient rehabilitation setting early after stroke, the sample improved on all measures from study day 0 to study day 14. This can be seen in the mean change scores provided in the last column of table 2. Note that improvement on the WMFT Time score is indicated by a negative number, i.e. shorter time to complete the test items. For all other measures, improvement is indicated by a positive number.

Distribution of subjects' perceived change ratings

Overall, the distribution of perceived change ratings was positively skewed. Twenty-nine subjects (56%) rated their upper extremity a 1, "much better", 12 subjects (23%) rated their upper extremity a 2, "a little better – meaningful", 4 subjects (8%) rated their upper extremity a 3, "a little better – not meaningful", and 7 subjects (13%) rated their upper extremity a 4, "about the same". No subject rated their upper extremity as having become worse (scores of 5–7). Figure 1 shows the distribution of perceived change ratings based on whether the dominant or the non-dominant upper extremity was affected by the stroke. Dominant and non-dominant affected sides were equally represented in the groups of subjects rating their upper extremities as a 1 or 2. More subjects who rated themselves a 3 had their dominant arm affected, and all subjects who rated themselves a 4 had their non-dominant arm affected.

Mean change scores for each rating and estimated MCIDs

Figures 2–4 show mean change scores for each measure plotted as a function of perceived change ratings and affected side. Subjects who rated their affected upper extremities as a 3 or 4 were grouped together in each graph because both ratings represent less than a meaningful change. For the impairment level measures, grip strength change scores (figure 2A) had a generally linear relationship with the perceived change rating for both the dominant and non-dominant groups. Composite upper extremity strength change scores (figure 2B) for the non-dominant group (open symbols) had a linear relationship with the perceived change ratings. This was not the case however for the group affected on the dominant side (filled symbols). Subjects who rated their affected dominant upper extremity as having not changed or not meaningfully changed (score of 3 or 4) had somewhat higher change scores than those who rated their affected dominant upper extremity as having a small, meaningful change (score of 2).

For the activity level measures, ARAT scores (figure 3A) had a linear relationship with the perceived change rating for both the dominant and non-dominant groups. The WMFT Time change scores (figure 3B) had a linear relationship when the dominant side was affected but had little relationship with perceived change ratings when the non-dominant side was affected. For this measure, subjects who rated their non-dominant affected upper extremity as being much better (score of 1) had change scores that were similar to change scores of the subjects who rated their non-dominant affected upper extremity as changing less (scores 2, 3, and 4).

The WMFT Function change scores (figure 3C) had a better relationship with the perceived change ratings than the WMFT Time scores.

For the participation level measures, the MAL How well change scores (figure 4A) had a linear relationship with perceived change ratings, with the dominant and non-dominant groups being very similar. The duration of upper extremity use measured with accelerometers (figure 4B) had no clear relationships with perceived ratings of change for either the dominant or the non-dominant groups. Regardless of which side was affected the largest changes in use were found in those people who considered their affected upper extremity as having not meaningfully changed or not changed at all (score of 3 or 4).

We used the mean change score for the smallest meaningful change (here = score of 2, “a little better, meaningful”) as the estimate of the MCID in our sample of people with acute hemiparesis. Table 3 shows estimated values of MCID for each measure. For some measures, the mean change score for subjects who rated themselves a 2 was smaller than the mean change score for those who rated themselves a 3 or 4 (see figures 2B, 3B, and 4B). The MCID on these measures was classified as unknown. When expressed as a percentage of the total scale, estimated MCID values ranged from 16–21% in the group that was affected on the dominant side, and 18–30% in the group that was affected on the non-dominant side. When expressed as an effect size, estimated MCID values ranged from 0.48–2.0 and 0.59–2.2 in the dominant and non-dominant groups, respectively.

DISCUSSION

Our purpose was to begin to examine how much change constitutes a clinically meaningful difference on upper extremity assessments early after stroke. Using subjects’ global ratings of upper extremity change, we attempted to estimate MCID values for 2 impairment level measures, 3 activity level measures, and 2 participation level measures during the first month post stroke (table 3). Estimates of MCID ranged from 16–30% of the total scales. For a few measures, we were unable to estimate MCID values because there appeared to be little relationship between change scores and perceived change ratings. Our data from this sample of people with acute hemiparesis post stroke provide some of the first estimates of MCID values for upper extremity standardized measures. As such, they are an important contribution to the rehabilitation literature because clinicians need to evaluate meaningful change in individual patients and researchers need to evaluate meaningful change across patient groups.

Cautions before interpreting our estimates

Our estimates are from a sample of people with acute hemiparesis post stroke, studied during their inpatient rehabilitation stay. It is not known how these estimates might generalize to people with more chronic hemiparesis or to people in other clinical settings. In the first month post stroke, substantial improvements in neuromotor capabilities and in function often occur^{44,45}, particularly for those people selected to go to inpatient rehabilitation facilities in the United States. It may be that people with acute hemiparesis require a greater change in movement capabilities to consider the change clinically meaningful because: 1) a large portion of the recovery happens during this time period, and 2) patients have strong expectations for recovery early after their stroke. Thus, our MCID values might be higher than MCID values at later time points post stroke, when people with hemiparesis may have a greater awareness of how smaller changes may be functionally beneficial and may have lower expectations for full recovery.

The majority of people (56%) in our sample considered themselves to be “much better” at the second evaluation. This created a skewed distribution of perceived change scores (figure 1), with a smaller portion of people who considered themselves to have changed a little or not at

all. Skewed distributions of perceived change scores are common in studies investigating MCID values.^{3, 13, 20, 46, 47} Given the substantial recovery that occurs in this early time period post stroke, it is nearly impossible to obtain a non-skewed distribution of perceived change scores.^{8, 9} We consider that the primary effect of the skewed distribution was that this left us with a smaller number of people ($n = 12$) from which to estimate the MCID values.

The number of subjects used here to calculate MCID values is similar to numbers in other anchor-based reports in the rehabilitation literature.^{9, 46, 47} Because of the small sample, the variability in change scores was large (see standard error bars in figures 2–4). Distributions of change scores in our data set overlap for different perceived change ratings. This means that we had individuals with change scores greater than the calculated MCID values who considered themselves “about the same” and we had individuals with change scores less than the calculated MCID values who considered themselves “much better”. Our MCID values should be interpreted with caution, particularly when making judgments about individual patients.

Interpreting our estimates

Based on our data, it appears that change scores of 16–30% on these upper extremity measures may be needed for a person to consider an improvement to be meaningful. In comparison, change scores of 5–10% have been found to be clinically meaningful on a number of health related quality of life measures in a variety of patient populations.^{3, 18, 19, 47} It is possible that our estimates are proportionally higher because health related quality of life measures are designed to assess changes that are inherently meaningful to the patient, whereas upper extremity measures are designed to assess changes in upper extremity motor capacity, some aspects of which may be less inherently meaningful to the patient. Thus, a greater proportional change may need to occur before the change is considered meaningful on focal upper extremity measures versus more global quality of life measures.

There are only a small number of MCID values available for comparison in the neuro-rehabilitation literature. We have been unable to find any estimates of MCID values for impairment level measures such as grip strength. For the activity level measures, a change of 6 points on the ARAT was selected as clinically meaningful in a population of people with chronic hemiparesis because it represented approximately 10% of 57 point scale.⁴⁸ Our data suggest that in an acute hemiparetic population, the patient-perceived MCID is 12 points for dominant side and 17 for non-dominant side, representing 21% and 30% of the scale, respectively. On the WMFT time score, an improvement of 19 seconds on the affected dominant side (16% of the 120 second limit) was considered a meaningful change in our sample, but we were unable to obtain an estimate for the affected non-dominant side. Interestingly, a recent report in people with subacute hemiparesis indicated that it was follow-up time scores on the WMFT and not change scores that predicted perceived upper extremity recovery.³⁸ This could imply that people may place more importance on their current level of function than on how much they have progressed when determining benefits of treatments. For the participation level measures, a change of 0.5 points (10%) on the MAL was selected as clinically meaningful in a population of people with chronic hemiparesis, again because it represented approximately 10% of the scale.⁴⁸ Our data suggest that in an acute hemiparetic population, the patient-perceived MCID on the MAL is at least 1 point (17–18% of the scale). Interestingly, our MCID values were proportionally similar to MCID values on 2 global activity measures in the stroke population, where the MCID of the Barthel Index is 1.85 on the 20 point scale (19%; acute population)⁹, and the MCID of the motor portion of the Functional Independence Measure is 17 on the 105 point scale (16%; subacute population).⁸

We determined MCID values separately for the affected dominant and non-dominant sides. Our estimated MCID values were a little smaller when the dominant side was affected compared to when the non-dominant side was affected (table 3). Although people typically use

the dominant and non-dominant upper extremities a similar amount during daily life⁴⁰, the two sides are used somewhat differently. The dominant upper extremity is used for more skilled manipulation (e.g. holding and writing with a pen), while the non-dominant upper extremity is used in a more supportive role (e.g. holding the paper still during writing). A limited ability to use the dominant side after stroke may be more burdensome than a limited ability to use the non-dominant side. It then follows that a smaller change in the ability to use the affected dominant side post stroke may be perceived more quickly and considered more meaningful than a similar size change in the ability to use the affected non-dominant side. Because the distributions of our dominant and non-dominant affected estimates overlap (see size of standard error bars in figures 2–4), our data can only provide preliminary support for this idea.

We were unable to estimate MCID values in instances where there was no apparent relationship between the perceived change score and the amount of change. This was the case for composite upper extremity strength when the dominant side was affected, for the WMFT time score when the non-dominant side was affected, and for the duration of use when either side was affected (figures 2–4 and table 3). One explanation for our inability to detect a relationship between the perceived change score and the amount of change may be our small sample size. Thus, the lack of relationship may be due to the large variability in our sample. This variability may be masking a real relationship that could be uncovered with larger sample sizes, from which it may then be possible to obtain estimates of MCID values. We consider this a real possibility for composite upper extremity strength measure and for the WMFT time score, where a relationship existed when one side was affected but not when the other side was affected. An alternative explanation for our inability to detect a relationship may be that change scores on these measures are not related to patient-perceived change. For example, it may have little meaning to the patient that their duration of upper extremity use improves by 1.5 hours. If this alternative explanation is true, then change scores on these particular measures may be less useful for assessing outcomes in individual patients or in research studies.

Conclusions

Our data provide some of the first estimates of what constitutes a clinically meaningful change on standardized upper extremity measures early after stroke. Proportional changes of 16–30% were considered important to the people in our study. Care should be taken when attempting to generalize these results to people with chronic stroke or to people in other rehabilitation settings. Future studies with larger sample sizes are needed to refine our estimates and to determine how MCID values are affected by time since stroke, by initial stroke severity, and/or by clinical setting.

Acknowledgments

This work was supported by NIH NS41261, HD047669, and the James S. McDonnell Foundation 21002032. We thank Joanne Wagner and Lily Hu for their assistance with data collection and the therapists who assisted with recruitment and scheduling during this project.

References

1. Barak S, Duncan PW. Issues in selecting outcome measures to assess functional recovery after stroke. *NeuroRx* 2006;3:505–24. [PubMed: 17012065]
2. Finch, E.; Brooks, D.; Stratford, PW.; Mayo, NE. Physical rehabilitation outcome measures. Hamilton, ON: BC Decker Inc; 2002.
3. Jaeschke R, Singer J, Guyatt GH. Measurement of health status. Ascertaining the minimal clinically important difference. *Control Clin Trials* 1989;10:407–15. [PubMed: 2691207]

4. de Vet HC, Terwee CB, Ostelo RW, Beckerman H, Knol DL, Bouter LM. Minimal changes in health status questionnaires: Distinction between minimally detectable change and minimally important change. *Health Qual Life Outcomes* 2006;4:54. [PubMed: 16925807]
5. 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;14:109–14. [PubMed: 11845014]
6. Hays RD, Woolley JM. The concept of clinically meaningful difference in health-related quality-of-life research. How meaningful is it? *Pharmacoeconomics* 2000;18:419–23. [PubMed: 11151395]
7. Wallace D, Duncan PW, Lai SM. Comparison of the responsiveness of the Barthel Index and the motor component of the Functional Independence Measure in stroke: The impact of using different methods for measuring responsiveness. *J Clin Epidemiol* 2002;55:922–28. [PubMed: 12393081]
8. Beninato M, Gill-Body KM, Salles S, Stark PC, Black-Schaffer RM, Stein J. Determination of the minimal clinically important difference in the FIM instrument in patients with stroke. *Arch Phys Med Rehabil* 2006;87:32–9. [PubMed: 16401435]
9. 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;21:233–8. [PubMed: 17351082]
10. Caro JJ, Huybrechts KF, Duchesne I. Management patterns and costs of acute ischemic stroke: An international study. For the stroke economic analysis group. *Stroke* 2000;31:582–90. [PubMed: 10700489]
11. Spieler JF, Lanoe JL, Amarenco P. Costs of stroke care according to handicap levels and stroke subtypes. *Cerebrovasc Dis* 2004;17:134–42. [PubMed: 14707413]
12. Wyrwich KW, Wolinsky FD. Identifying meaningful intra-individual change standards for health-related quality of life measures. *J Eval Clin Pract* 2000;6:39–49. [PubMed: 10807023]
13. Ringash J, O'Sullivan B, Bezjak A, Redelmeier DA. Interpreting clinically significant changes in patient-reported outcomes. *Cancer* 2007;110:196–202. [PubMed: 17546575]
14. Katzman R, Brown T, Fuld P, Peck A, Schechter R, Schimmel H. Validation of a short orientation-memory-concentration test of cognitive impairment. *Am J Psychiatry* 1983;140:734–9. [PubMed: 6846631]
15. Lang CE, Wagner JM, Dromerick AW, Edwards DF. Measurement of upper-extremity function early after stroke: Properties of the action research arm test. *Arch Phys Med Rehabil* 2006;87:1605–10. [PubMed: 17141640]
16. Dromerick, AW.; Lang, CE.; Powers, WJ.; Videen, TO.; Wagner, JM.; Sahrman, SA.; Miller, JP.; Birkenmeier, R.; Hahn, M.; Wolf, S.; Edwards, DF. Very early constraint therapy for recovery from stroke (vectors): Phase II trial results. *International Stroke Conference*; 2007.
17. Coeytaux RR, Kaufman JS, Chao R, Mann JD, Devellis RF. Four methods of estimating the minimal important difference score were compared to establish a clinically significant change in headache impact test. *J Clin Epidemiol* 2006;59:374–80. [PubMed: 16549259]
18. Wyrwich KW, Nienaber NA, Tierney WM, Wolinsky FD. Linking clinical relevance and statistical significance in evaluating intra-individual changes in health-related quality of life. *Med Care* 1999;37:469–78. [PubMed: 10335749]
19. Osoba D, Rodrigues G, Myles J, Zee B, Pater J. Interpreting the significance of changes in health-related quality-of-life scores. *J Clin Oncol* 1998;16:139–44. [PubMed: 9440735]
20. Lauridsen HH, Hartvigsen J, Manniche C, Korsholm L, Grunnet-Nilsson N. Responsiveness and minimal clinically important difference for pain and disability instruments in low back pain patients. *BMC Musculoskelet Disord* 2006;7:82. [PubMed: 17064410]
21. Bertrand AM, Mercier C, Bourbonnais D, Desrosiers J, Gravel D. Reliability of maximal static strength measurements of the arms in subjects with hemiparesis. *Clin Rehabil* 2007;21:248–57. [PubMed: 17329282]
22. Boissy P, Bourbonnais D, Carlotti MM, Gravel D, Arsenault BA. Maximal grip force in chronic stroke subjects and its relationship to global upper extremity function. *Clin Rehabil* 1999;13:354–62. [PubMed: 10460123]
23. Schmidt RT, Toews JV. Grip strength as measured by the Jamar dynamometer. *Arch Phys Med Rehabil* 1970;51:321–7. [PubMed: 5423802]

24. Andrews AW, Thomas MW, Bohannon RW. Normative values for isometric muscle force measurements obtained with hand-held dynamometers. *Phys Ther* 1996;76:248–59. [PubMed: 8602410]
25. Wagner JM, Lang CE, Sahrman SA, Edwards DF, Dromerick AW. Sensorimotor impairments and reaching performance in subjects with poststroke hemiparesis during the first few months of recovery. *Phys Ther* 2007;87:751–65. [PubMed: 17442839]
26. Wagner JM, Lang CE, Sahrman SA, Hu Q, Bastian AJ, Edwards DF, Dromerick AW. Relationships between sensorimotor impairments and reaching deficits in acute hemiparesis. *Neurorehabil Neural Repair* 2006;20:406–16. [PubMed: 16885427]
27. Lyle RC. A performance test for assessment of upper limb function in physical rehabilitation treatment and research. *Int J Rehabil Res* 1981;4:483–92. [PubMed: 7333761]
28. van der Lee JH, De Groot V, Beckerman H, Wagenaar RC, Lankhorst GJ, Bouter LM. The intra- and interrater reliability of the Action Research Arm Test: A practical test of upper extremity function in patients with stroke. *Arch Phys Med Rehabil* 2001;82:14–9. [PubMed: 11239280]
29. 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. [PubMed: 11482350]
30. Hsieh CL, Hsueh IP, Chiang FM, Lin PH. Inter-rater reliability and validity of the Action Research Arm Test in stroke patients. *Age Ageing* 1998;27:107–13. [PubMed: 16296669]
31. De Weerd W, Harrison MA. Measuring recovery of arm-hand function in stroke patients: A comparison of the Brunnstrom-Fugl-Meyer test and the Action Research Arm Test. *Physiother Can* 1985;37:65–70.
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;19:404–11. [PubMed: 15929509]
33. 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;32:1635–9. [PubMed: 11441212]
34. 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;104:125–32. [PubMed: 2707361]
35. Taub E, Miller NE, Novack TA, Cook EW 3rd, Fleming WC, Nepomuceno CS, Connell JS, Crago JE. Technique to improve chronic motor deficit after stroke. *Arch Phys Med Rehabil* 1993;74:347–54. [PubMed: 8466415]
36. 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;82:750–5. [PubMed: 11387578]
37. van der Lee JH, Beckerman H, Knol DL, de Vet HC, Bouter LM. Clinimetric properties of the Motor Activity Log for the assessment of arm use in hemiparetic patients. *Stroke* 2004;35:1410–4. [PubMed: 15087552]
38. Fritz SL, George SZ, Wolf SL, Light KE. Participant perception of recovery as criterion to establish importance of improvement for constraint-induced movement therapy outcome measures: A preliminary study. *Phys Ther* 2007;87:170–8. [PubMed: 17244694]
39. Uswatte G, Taub E, Morris D, Vignolo M, McCulloch K. Reliability and validity of the upper-extremity Motor Activity Log-14 for measuring real-world arm use. *Stroke* 2005;36:2493–6. [PubMed: 16224078]
40. Lang CE, Wagner JM, Edwards DF, Dromerick AW. Upper extremity use in people with hemiparesis in the first few weeks after stroke. *J Neurol Phys Ther* 2007;31:56–63. [PubMed: 17558358]
41. Uswatte G, Miltner WH, Foo B, Varma M, Moran S, Taub E. Objective measurement of functional upper-extremity movement using accelerometer recordings transformed with a threshold filter. *Stroke* 2000;31:662–7. [PubMed: 10700501]

42. Uswatte G, Foo WL, Olmstead H, Lopez K, Holand A, Simms LB. Ambulatory monitoring of arm movement using accelerometry: An objective measure of upper-extremity rehabilitation in persons with chronic stroke. *Arch Phys Med Rehabil* 2005;86:1498–501. [PubMed: 16003690]
43. Kazis LE, Anderson JJ, Meenan RF. Effect sizes for interpreting changes in health status. *Med Care* 1989;27:S178–89. [PubMed: 2646488]
44. Duncan PW, Lai SM, Keighley J. Defining post-stroke recovery: Implications for design and interpretation of drug trials. *Neuropharmacology* 2000;39:835–41. [PubMed: 10699448]
45. Jorgensen HS, Nakayama H, Raaschou HO, Vive-Larsen J, Stoier M, Olsen TS. Outcome and time course of recovery in stroke. Part II: Time course of recovery. The Copenhagen Stroke Study. *Arch Phys Med Rehabil* 1995;76:406–12. [PubMed: 7741609]
46. Iyer LV, Haley SM, Watkins MP, Dumas HM. Establishing minimal clinically important differences for scores on the Pediatric Evaluation of Disability Inventory for inpatient rehabilitation. *Phys Ther* 2003;83:888–98. [PubMed: 14519060]
47. Juniper EF, Guyatt GH, Willan A, Griffith LE. Determining a minimal important change in a disease-specific quality of life questionnaire. *J Clin Epidemiol* 1994;47:81–7. [PubMed: 8283197]
48. van der Lee JH, Wagenaar RC, Lankhorst GJ, Vogelaar TW, Deville WL, Bouter LM. Forced use of the upper extremity in chronic stroke patients: Results from a single-blind randomized clinical trial. *Stroke* 1999;30:2369–75. [PubMed: 10548673]

Figure 1.

Frequency of subjects' perceived change ratings. Subjects are grouped by whether the dominant or the non-dominant upper extremity was affected by the stroke. Perceived change ratings: 1 = Much better; 2 = A little better, meaningful; 3 = A little better, not meaningful; 4 = About the same; 5 = A little worse, not meaningful; 6 = A little worse, meaningful; 7 = Much worse.

Figure 2.

Mean change scores in impairment level measures by perceived change ratings and upper extremity affected. Error bars represent standard errors. A: Grip strength in the affected hand, measured in kgs as part of the Wolf Motor Function Test. B: Composite upper extremity strength on the affected side; values are expressed as ratios of the unaffected upper extremity. Perceived change ratings: 1 = Much better; 2 = A little better, meaningful; 3 = A little better, not meaningful; 4 = About the same.

Figure 3.

Mean change scores in activity level measures by perceived change ratings and upper extremity affected. Error bars represent standard errors. A: Action Research Arm Test. B: Wolf Motor Function Test Time score. C: Wolf Motor Function Test Function score. Perceived change ratings: 1 = Much better; 2 = A little better, meaningful; 3 = A little better, not meaningful; 4 = About the same.

Figure 4.

Mean change scores in participation level measures by perceived change ratings and upper extremity affected. Error bars represent standard errors. A: Motor Activity Log How well score. B: Duration of affected upper extremity use, measured by wrist accelerometers and expressed in hours. Perceived change ratings: 1 = Much better; 2 = A little better, meaningful; 3 = A little better, not meaningful; 4 = About the same.

Table 1

Subject characteristics (N = 52)

Variable	Mean $\pm$ SD
Age (years)	64 $\pm$ 14
Admission NIHSS Score	5.3 $\pm$ 1.8
Pre-Morbid Barthel Score	99.6 $\pm$ 2.2
Pre-Morbid Modified Rankin Index	0.3 $\pm$ 0.6
Time to day 0 evaluation (days since stroke)	9.5 $\pm$ 4.5
Time to day 14 evaluation (days since stroke)	25.9 $\pm$ 10.6
	N (%)
Gender	
Male	21 (40%)
Female	31 (60%)
Race	
Caucasian	22 (42%)
African American	29 (56%)
Other	1 (2%)
Stroke Type	
Ischemic	41 (79%)
Hemorrhagic	11 (21%)
Affected Side	
Dominant	23 (44%)
Non-Dominant	29 (56%)

Table 2

Means and change scores at the study day 0 and day 14 time points.

		Mean $\pm$ SD	Change $\pm$ SD*
Impairment measures			
Grip strength (kg)	D0	9.6 $\pm$ 10.5	6.9 $\pm$ 7.3
	D14	16.8 $\pm$ 12.5	
Composite strength (ratio of unaffected side)	D0	0.34 $\pm$ 0.28	0.22 $\pm$ 0.20
	D14	0.56 $\pm$ 0.33	
Activity measures			
ARAT	D0	22.5 $\pm$ 15.3	15.1 $\pm$ 11.4
	D14	38.1 $\pm$ 16.6	
WMFT Time (s)	D0	42.5 $\pm$ 39.8	-22.6 $\pm$ 28.8
	D14	18.4 $\pm$ 28.4	
WMFT Function	D0	2.4 $\pm$ 1.1	1.2 $\pm$ 0.8
	D14	3.6 $\pm$ 1.1	
Participation measures			
MAL How well	D0	0.5 $\pm$ 0.5	1.2 $\pm$ 0.9
	D14	1.7 $\pm$ 1.1	
Duration of use (hours)	D0	3.3 $\pm$ 1.8	1.2 $\pm$ 1.4
	D14	4.4 $\pm$ 2.1	

SD: standard deviation; ARAT: Action Research Arm Test; WMFT: Wolf Motor Function Test; MAL: Motor Activity Log; D0: study day 0, an average of 9.5 days post stroke; D14: study day 14, an average of 25.9 days post stroke.

* Note that improvement on the WMFT Time score is indicated by a negative number, i.e. shorter time to complete test items. For all other measures, improvement is indicated by a positive number.

Estimates of minimal clinically important difference (MCID) for each measure based on whether the dominant or the non-dominant upper extremity is affected. Values are expressed as raw scores, percentages of the total scale (where appropriate), and effect sizes.

Table 3

	MCID if dominant side affected			MCID if non-dominant side affected		
	Raw value	Percent of scale	Effect size	Raw value	Percent of scale	Effect size
Impairment measures						
Grip strength (kg)	5.0	--	0.48	6.2	--	0.59
Composite strength (ratio of unaffected side)	Unknown ^a	--	--	0.22	22% ^b	0.79
Activity measures						
ARAT	12	21%	0.78	17	30%	1.1
WMFT Time (s)	-19	16% ^c	0.48	Unknown ^d	--	--
WMFT Function	1.0	17%	0.91	1.2	20%	1.0
Participation measures						
MAL How well	1.0	17%	2.0	1.1	18%	2.2
Duration of use (hours)	Unknown ^d	--	--	Unknown ^d	--	--

^aWe were unable to estimate MCID values when there was no apparent relationship between the perceived change rating and the mean change on a given measure. See figures 2B, 3B, and 4B.

^bTo calculate percentage, total scale was assumed to range from 0.0 to 1.0.

^cTo calculate percentage, total scale was assumed to range from 0 to 120 seconds.