FRANCO FRANCHIGNONI, MD[1] • STEFANO VERCELLI, PT, MSc[1] • ANDREA GIORDANO, PhD[2]
FRANCESCO SARTORIO, PT[1] • ELISABETTA BRAVINI, PT[3] • GIORGIO FERRIERO, MD[1]

Minimal Clinically Important Difference of the Disabilities of the Arm, Shoulder and Hand Outcome Measure (DASH) and Its Shortened Version (QuickDASH)

The Disabilities of the Arm, Shoulder and Hand (DASH) outcome measure and its shortened version (QuickDASH) are 2 region-specific measures of disability and symptoms in people with musculoskeletal disorders of the upper limb.[4,18] Both have been widely used and are available in different language versions.[6,34] A prerequisite for meaningful use of such patient-reported outcome measures is the quality of their clinimetric properties. Strengths and weaknesses of the DASH and Quick-DASH have been well investigated by both classical test theory[1,23] and Rasch analysis,[12,13] but many different approaches have been used to calculate the responsiveness of these measures, in particular the minimal clinically important difference (MCID), also known as the minimal important change, which focuses on within-person change over time.[3,21] The MCID represents the smallest improvement in score to reflect a change that is clinically meaningful for the patient.[17] The MCID threshold is very important in daily practice, where clinicians routinely compare, at the individual level, the current and previous values of outcome measures of interest.[39]

Distribution- and anchor-based methods are the 2 general approaches that have been used to interpret score changes. However, each method has its shortcomings. The major disadvantage of distribution-based approaches is that they do not provide a good indication of the importance of the observed change and thus cannot give the MCID.[10] Their main role lies in identifying the minimum detectable change (MDC), that is, the smallest change in score that can be detected beyond random error.[41] On the

⦿ **STUDY DESIGN:** Prospective, single-group observational design.

⦿ **OBJECTIVES:** To determine the minimal clinically important difference (MCID) for the Disabilities of the Arm, Shoulder and Hand (DASH) outcome measure and its shortened version (QuickDASH) in patients with upper-limb musculoskeletal disorders, using a triangulation of distribution- and anchor-based approaches.

⦿ **BACKGROUND:** Meaningful threshold change values of outcome tools are crucial for the clinical decision-making process.

⦿ **METHODS:** The DASH and QuickDASH were administered to 255 patients (mean ± SD age, 49 ± 15 years; 156 women) before and after a physical therapy program. The external anchor administered after the program was a 7-point global rating of change scale.

⦿ **RESULTS:** The test-retest reliability of the DASH and QuickDASH was high (intraclass correlation coefficient model 2,1 = 0.93 and 0.91, respectively; n = 30). The minimum detectable change at the 90% confidence level was 10.81 points for the DASH and 12.85 points for the QuickDASH. After triangulation of these results with those of the mean-change approach and receiver-operating-characteristic-curve analysis, the following MCID values were selected: 10.83 points for the DASH (sensitivity, 82%; specificity, 74%) and 15.91 points for the QuickDASH (sensitivity, 79%; specificity, 75%). After treatment, the MCID threshold was reached/surpassed by 61% of subjects using the DASH and 57% using the QuickDASH.

⦿ **CONCLUSION:** The MCID values from this study for the DASH (10.83 points) and the QuickDASH (15.91 points) could represent the lower boundary for a range of MCID values (reasonably useful for different populations and contextual characteristics). The upper boundary may be represented by the 15 points for the DASH and 20 points for the QuickDASH proposed by the DASH website. *J Orthop Sports Phys Ther 2014;44(1):30-39. Epub 30 October 2013. doi:10.2519/jospt.2014.4893*

⦿ **KEY WORDS:** *disability evaluation, musculoskeletal diseases, outcome assessment, psychometrics, rehabilitation, upper extremity*

other hand, the accuracy of results of anchor-based methods, which can provide the MCID, depends among other things on the choice of the anchor,[2] the definition of "minimal importance" on the anchor, and baseline values, type of population, and contextual characteristics.[11]

Because it is common for these different methods to yield different threshold values, recent papers recommend that the MCID be based primarily on anchor-based procedures,[32] be higher than MDC values (the boundary of variability typically found in stable patients),[32,41] and not be based on 1 study or 1 method only.[40] However, the studies calculating MCID through anchor-based approaches are limited for both the DASH[2,3,9,24,25,35,36] and QuickDASH.[24,27,30,36] In addition, it appears that the best choice to determine MCID is to select a small range of threshold estimates after comparing and interpreting the information conveyed by multiple reference standards, calculated on the same sample.[32,39,40] To date, such an approach has only recently been applied for the DASH in a study[3] dealing with soft tissue shoulder disorders. Therefore, further investigation of these outcome measures is needed to establish the stability of prior reports and to investigate MCID values in other cohorts, so as to provide useful insights about the magnitude of change in score that could be considered a clinically meaningful improvement.

The main aim of this study was to use both distribution- and anchor-based methods to triangulate on MCID values for the DASH and QuickDASH in a large sample of patients with upper-limb musculoskeletal disorders who were undergoing physical therapy, to enhance confidence in interpreting their change scores for clinical decision making.

METHODS

Subjects

THIS PROSPECTIVE OBSERVATIONAL study was carried out at the Salvatore Maugeri Foundation Sci-

TABLE 1	DESCRIPTIVE STATISTICS OF THE COHORT (N = 255)
Descriptor	**Data**
Age, y*	49 ± 15 (18-84)
Gender, n (%)	
Female	156 (61%)
Male	99 (39%)
Duration of symptoms, n (%)	
12 wk or less	214 (84%)
Greater than 12 wk	41 (16%)
Occupational status: in the labor force, n (%)†	217 (85%)
Treatment sessions, n‡	10 (8-12)
Duration of treatment, d*	22 ± 4 (15-32)

Values are mean ± SD (range).
†*Including students over 18 years of age and homemakers.*
‡*Value is median (25th-75th percentile).*

entific Institute of Veruno, Italy in 266 consecutive adult inpatients and outpatients referred for an intensive upper-limb physical therapy program. To be included, participants had to be adults over 18 years of age and to be suffering from upper-limb musculoskeletal disorders. Those who had severe cognitive or communication impairments, pain in the upper limb arising from other regions (eg, the neck), a diagnosis of central nervous system or psychiatric disorders, and a history of tumor malignancy were excluded from the study. Prior to taking part in the study, all participants signed an informed consent form approved by the Ethics Committee of the hospital. After excluding data from incomplete surveys (n = 4) and patients lost to follow-up (n = 7), 255 patients remained as the final study population. **TABLE 1** provides descriptive statistics for the cohort.

Assessment

Disabilities of the Arm, Shoulder and Hand The DASH[18] is a region-specific measure of disability and symptoms in people with any or multiple musculoskeletal disorders of the upper limb. The items inquire about (1) the degree of difficulty during the preceding week in performing various physical activities because of problems in a shoulder, arm, or hand (21 items); (2) the severity of each of the symptoms of pain, activity-related pain, tingling, weakness, and stiffness (5 items); and (3) the problem's effect on social activities, work, and sleep, and its psychological impact (4 items). There are 5 response options for each item, from 1 (no difficulty to perform, no symptom, or no impact) to 5 (unable to do, very severe symptom, or high impact). The responses to the 30 items are summed to form a raw score that is then converted to a 0-to-100 scale with the following formula: $[(\text{sum of score}/n) - 1] \times 25$, where n is the number of completed responses. A higher score reflects greater disability. If less than 10% of the items are left blank by the respondent, then the mean value of the responses to the completed items is substituted for each missing item. The full-length Italian version of the DASH, as published on the DASH website (http://www.dash.iwh.on.ca), was used in this study, excluding the optional modules for work and sports/performing arts. **Shortened Version of the DASH** The QuickDASH[4] is a shortened version of the DASH questionnaire that uses 11 items to measure the degree of difficulty in performing various physical activities due to a shoulder, arm, or hand problem (6 items); the severity of pain and tingling (2 items); and the problem's ef-

fect on social activities, work, and sleep (3 items). It uses the same 5 response options of the DASH. In this study, the QuickDASH item responses were extracted from the subjects' responses to the full-length questionnaire. If at least 10 of the 11 items were completed, the responses to the items were summed to form a raw score, then converted to a 0-to-100 scale with the same formula used to calculate the DASH score, with higher scores reflecting greater disability. The 2 optional scales of the QuickDASH (work and sport/music) were not part of this study.

Global Rating of Change Scale The global rating of change scale (GRCS) is a rating scale designed to quantify a patient's improvement/deterioration over time, usually to determine the effect of an intervention or to chart the clinical course of a condition. At the time of the final assessment (after the rehabilitation treatment), patients were asked to independently rate the overall change in their upper-limb condition from when they began treatment, using a 7-point scale ranging from –3 (a great deal worse) to +3 (a great deal better), with 0 indicating "unchanged."[21] The self-assessment value was used as an external anchor, with patients who rated their improvement from 0 to +1 (a little better) being considered as not improved or minimally improved and those who rated their improvement as +2 (somewhat better) or higher being considered as moderately to largely improved.[41]

Procedure

All patients completed the full-length DASH questionnaire (written format) at the first and last sessions of a tailored, comprehensive physical therapy program that included, when indicated, passive or active mobilization, stretching, strength training, and functional exercises. Sessions were planned twice or 3 times weekly. At the end of the intensive physical therapy program, which lasted 2 to 5 weeks, patients also completed the GRCS. Three physical therapists (S.V.,

F.S., E.B.) distributed all questionnaires and collected answer sheets. Test-retest reliability was analyzed in a subset of 30 consecutive patients (mean ± SD age, 50 ± 13 years; 10 women) who were assessed twice, 1 to 3 days prior to treatment and at the start of the treatment.

Statistical Analysis

Descriptive Statistics, Validity, and Reliability Descriptive statistics were calculated for the DASH, QuickDASH, and GRCS. All statistical analyses were performed using Stata/IC Version 10.1 (StataCorp LP, College Station, TX).

Convergent validity was assessed by calculating the Pearson correlation coefficient (r) of the total scores of the DASH and QuickDASH at the first evaluation and follow-up and the change in score from first evaluation to follow-up. In addition, because the GRCS was the reference standard (ie, the external criterion against which we judged if a real improvement in patients had occurred), it assessed the same construct measured by the tools under longitudinal investigation.[42] Thus, the correlation between the GRCS and the change (after versus before rehabilitation) in the 2 questionnaires was calculated. For all these correlations, an at least fair association ($r>0.30$) between measures was expected.[32]

Test-retest reliability was calculated for both scales using an intraclass correlation coefficient ($ICC_{2,1}$). We determined the sample size for test-retest reliability, expecting to obtain ICC values of at least 0.90 with a confidence interval (CI) of 0.20.[5]

Responsiveness The 2 main methods of evaluating the ability of a measure to detect changes are a distribution-based method and an anchor-based method. Distribution-based methods rely on statistical characteristics of the sample and analyze the ability to detect change in general. Anchor-based methods require an external patient-based criterion (anchor) to determine if changes in outcome scores are clinically meaningful.[8] We used both approaches to have a wider range of

results from which to draw inferences about the MCID of both scales, aware of the large variation and lack of convergence that these different methods can have.[40]

For the distribution-based methods, we calculated the standard error of measurement (SEM), which links the reliability of a measurement tool to the standard deviation of the population.[10] The SEM was calculated from the analysis of variance used to produce the test-retest ICC.[38] Starting from the SEM, we calculated the MDC, which represents the smallest change in score likely to reflect true change, free of measurement error.[37] The calculation is the result of the multiplication of the SEM by the z-value by the square root of 2. The 90% confidence level (MDC_{90}) was established, corresponding to a z-value of 1.65. The meaning of this statistic is that if a patient has a change score equal to or above the MDC_{90} threshold, it is possible to state with 90% confidence that this change is real and not due to measurement error.

Regarding anchor-based methods, the mean-change and receiver-operating-characteristic (ROC) curve approaches were followed, and GRCS assessment was used as the external reference in evaluating responsiveness. For the mean-change approach, we calculated the mean change score in the different subgroups of patients who respectively reported themselves on the anchor as not improved (GRCS, 0), minimally improved (GRCS, +1), moderately improved (GRCS, +2), or largely improved (GRCS, +3). We used the mean change in those minimally improved for triangulating the MCID values.[40]

For the ROC curve approach, we determined the optimal cutoff score and the area under the curve (AUC), considering the subjects improved according to a GRCS of +2 or greater. A ROC curve plots sensitivity (y-axis) against 1 – specificity (x-axis). In this context, sensitivity was calculated as the number of patients correctly identified as improved based on the cutoff value divided by all patients

TABLE 2	Diagnoses of Upper-Limb Dysfunction Within the Study Population	
Region/Diagnosis Group	**n**	**Percent of Total**
Shoulder	149	58.5
Biceps or rotator cuff surgery	51	
Biceps or rotator cuff tendinitis	32	
Shoulder impingement	21	
Shoulder instability	18	
Shoulder fracture	15	
Shoulder arthritis	10	
Shoulder prosthesis	2	
Elbow	23	9.0
Elbow fracture	13	
Elbow tendinitis	8	
Nerve entrapment (surgery)	2	
Wrist	23	9.0
Wrist fracture	23	
Hand	47	18.4
Flexor/extensor tendon injury	21	
Carpal tunnel syndrome	15	
Carpal/finger fracture	8	
Dupuytren's disease (surgery)	3	
Multijoint	13	5.1
Multiple fracture	5	
Complex regional pain syndrome	4	
Nerve injury	3	
Tendinitis (repetitive strain injury)	1	
Total	255	100.0

identified as having had a meaningful change (GRCS, +2 or greater), whereas specificity refers to the number of patients who were correctly identified as not improved based on the cutoff value divided by all patients who truly did not have a meaningful change (GRCS, less than +2). The optimal cutoff was chosen as the point that jointly maximized sensitivity and specificity (was associated with the least amount of misclassification). The AUC can be interpreted as the probability of correctly identifying a patient who has improved in randomly selected pairs of patients who have and have not shown an improvement. The greater the AUC, the greater a measure's ability to distinguish patients who have improved from those who have not improved. As a general rule, AUC values between 0.7 and 0.8 are judged as acceptable, and an AUC value greater than 0.8 is considered to have good to excellent discrimination.[44] In accordance with Turner et al,[42] our ROC analysis used the entire cohort, rather than just those subjects with ratings adjacent to the dichotomization point, to increase precision and obtain more logical estimates of the MCID. To obtain CIs for the ROC-derived parameters, we drew 500 bootstrap samples and calculated both the cutoff value and the AUC in each bootstrap replication. The mean of the 500 bootstrap AUC values was taken as the best estimate, with the 95% CI calculated as $1.96 \times$ SD (as an estimate of the standard error) of the bootstrap values.[40] Once the best cutoff value was estimated, we used the sensitivity and specificity at this value in each of the 500 ROCs obtained by bootstrapping to compute mean and 95% CI for

these parameters.

The MCID was set at the best triangulation of the results coming from both anchor-based (mean change and the ROC curve) and distribution-based (the MDC_{90} threshold) methods, considering that the MCID should be based primarily on anchor-based procedures[32] and be higher than the MDC value.[9,41] In this regard, the MDC should be interpreted as another piece in the puzzle toward establishing the MCID, by benchmarking it to the boundaries of error. According to Turner et al,[41] "if the two anchor-based methods calculated on the same population yield different MCID values, then the knowledge that one value is below the MDC could aid in the decision to select the other." In addition, the ROC-curve approach was preferred as the first choice, as it successfully addresses most limitations of the mean-change approach.[32,40,42] Furthermore, our calculation of the 95% CIs gave a useful indication of the sampling variation.[10]

In addition, patients who reached (or surpassed) MCID thresholds were considered positive responders. Then, we calculated how many positive responders showed a final score of 16 points or less, which was lower than both the fifth percentile of our dysfunctional population at baseline (16.42 points for the DASH, 19.77 points for the QuickDASH) and the 80th percentile of the DASH and QuickDASH norms for the general population.[19,20]

RESULTS

Descriptive Statistics and Validity of the Measures

TABLE 2 SHOWS THE DISTRIBUTIONS of joints involved and diagnostic groups related to upper-limb dysfunction within the sample. No significant difference in baseline scores was observed between the acute (12 weeks or less) and chronic (greater than 12 weeks) conditions in either scale. **FIGURES 1** and **2** show the score distribution of the 2 scales before and after treatment.
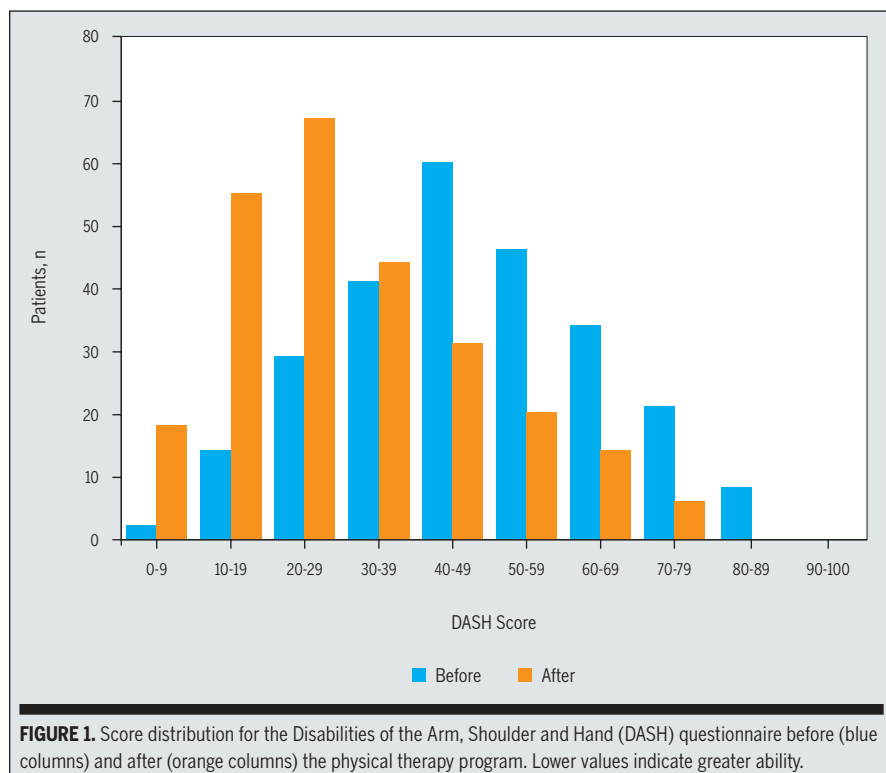
FIGURE 1. Score distribution for the Disabilities of the Arm, Shoulder and Hand (DASH) questionnaire before (blue columns) and after (orange columns) the physical therapy program. Lower values indicate greater ability.
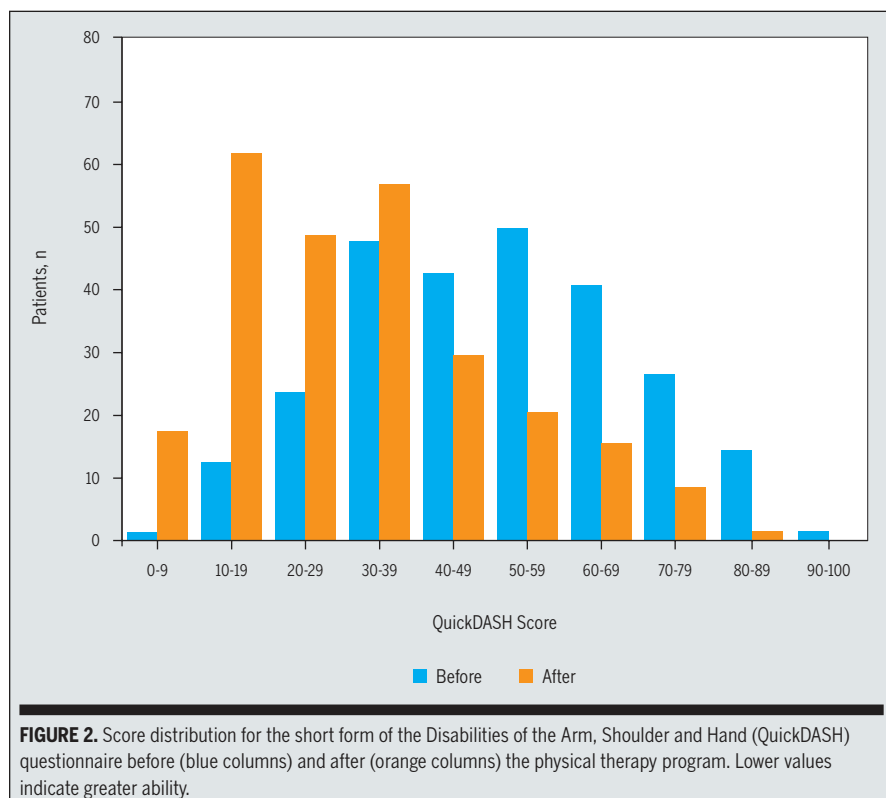

FIGURE 2. Score distribution for the short form of the Disabilities of the Arm, Shoulder and Hand (QuickDASH) questionnaire before (blue columns) and after (orange columns) the physical therapy program. Lower values indicate greater ability.

Scores for the DASH and QuickDASH (before and after treatment), as well as GRCS values, are presented in **TABLE 3**. Mean score changes for the DASH and QuickDASH questionnaires according to each GRCS grade are shown in **TABLE 4**. The scores of the DASH and QuickDASH were highly correlated at both baseline (r = 0.96; 95% CI: 0.94, 0.97; P<.001) and follow-up (r = 0.97; 95% CI: 0.96, 0.97; P<.001). The correlation between score changes of the DASH and Quick-DASH over the course of the rehabilitation program was r = 0.92 (95% CI: 0.90, 0.94; P<.001). The correlations between the GRCS and the score changes of both scales were r = 0.72 (95% CI: 0.66, 0.78) for the DASH and r = 0.71 (95% CI: 0.64, 0.76) for the QuickDASH (P<.001 for both).

The test-retest reliability of the DASH and QuickDASH was high for both scales ($ICC_{2,1}$ = 0.93; 95% CI: 0.81, 0.97 and 0.91; 95% CI: 0.81, 0.96, respectively).
Responsiveness: Distribution-Based Methods For the DASH, the SEM was 4.63 and the MDC_{90} corresponded to 10.81 points. For the QuickDASH, the SEM was 5.51, with an MDC_{90} of 12.85 points.
Responsiveness: Anchor-Based Methods The mean changes for the DASH and QuickDASH are reported in **TABLE 4**. In particular, those patients who were rated as having had small (GRCS, +1) or moderate (GRCS, +2) improvement had a mean change of 11.4 (95% CI: 9.1, 13.6) and 19.8 (95% CI: 17.8, 21.7) points for the DASH and of 13.2 (95% CI: 10.6, 15.6) and 21.9 (95% CI: 19.6, 24.1) points for the QuickDASH.

Splitting data according to the presence of a moderate versus large GRCS improvement (GRCS less than +2 versus GRCS of +2 or greater), the AUC of the DASH was 0.87 (95% CI: 0.84, 0.91), and that of the QuickDASH was 0.86 (95% CI: 0.82, 0.89) (**FIGURE 3**). The cut-off scores that best identified meaningful improvement in clinical status (as measured by GRCS values of +2 or greater) were 9.17 (95% CI: 7.50, 10.83) for the

DASH and 15.91 (95% CI: 9.1, 20.5) for the QuickDASH.

For our triangulation, we took into account the following data: (*a*) an MDC_{90} of 10.81 points for the DASH and of 12.85 points for the QuickDASH, (*b*) a mean change for small improvement of 11.4 points for the DASH and of 13.2 points for the QuickDASH, and (*c*) a cutoff score that best identified meaningful improvement in clinical status of 9.17 (95% CI: 7.50, 10.83) for the DASH and 15.91 (95% CI: 9.1, 20.5) for the QuickDASH.

Analyzing the overall results led us to select as the best triangulation the following values:

- For the DASH, an MCID of 10.83 points (sensitivity, 82%; specificity, 74%; correctly classified, 79%), slightly higher than the best cutoff score (9.17 points) and the first available DASH measure higher than the MDC_{90} (10.81 points)
- For the QuickDASH, an MCID of 15.91 points (sensitivity, 79%; specificity, 75%; correctly classified, 78%), representing the best cutoff score, slightly higher than the MDC_{90} (12.85 points)

In our cohort, the percentage of positive responders, that is, those who reached or surpassed the MCID threshold (after a median number of 10 physical therapy sessions), was 61% using the DASH and 57% using the QuickDASH. Moreover, 128 of the 151 subjects (85%) who had a moderate to large improvement (GRCS of +2 or greater) showed a change after physical therapy equal to or higher than the MCID (10.83 points) on the DASH, whereas 119 (79%) showed a change equal to or higher than the MCID (15.91 points) on the QuickDASH. Among them, 45 subjects for the DASH and 46 for the QuickDASH (coefficient of agreement, 87%) attained a "good" final state (final score of 16 points or less). In addition, among the 43 patients who reported a small improvement (GRCS, +1), 22 (51%) showed a change of 10.83 points or greater on the DASH, and 21 (49%) showed a change of 15.91 points or great-

er on the QuickDASH. In the remaining GRCS categories (GRCS of 0 or less), no subjects showed a positive change equal to or higher than the respective MCIDs.

DISCUSSION

Assessing patient progress is an integral part of clinical practice, and meaningful threshold change values of outcome tools are essential for decision making regarding a patient's status and to facilitate the communication of results in a concise and comprehensible fashion. This study used both distribution- and anchor-based approaches to define clinically meaningful MCID values for the DASH and QuickDASH. Triangulation of our results considered that the MCID should be based primarily on anchor-based procedures (and in the first instance on the ROC curve)[32] and

be higher than the MDC value.[9,41] These MCID thresholds represent the smallest improvement considered worthwhile by a patient, and thus increase the interpretability of score changes at the individual level observed in the clinical setting.

The high correlation between the DASH questionnaire and its shortened version, and between their changes after a rehabilitation intervention, supports the convergent validity of the 2 measures. In this study, the QuickDASH scores were extracted from the full-length DASH responses, and it is not known if patients' responses to the 11 items would have differed if only the QuickDASH had been administered; however, this method has been shown to be sound in a previous study.[16] In addition, the ability of the participants to estimate the change in their upper-limb disability during the treatment period was confirmed by the

TABLE 3	Scores of the DASH, QuickDASH, and GRCS

	Mean ± SD	Median (25th-75th Percentile)
Admission DASH	46 ± 18	46 (35-58)
Discharge DASH	30 ± 17	27.5 (17.5-40)
Admission QuickDASH	49 ± 18	50 (36-61)
Discharge QuickDASH	31 ± 18	30 (18-43)
GRCS	1.45 ± 1.09	1 (1-2)

Abbreviations: DASH, Disabilities of the Arm, Shoulder and Hand questionnaire; GRCS, global rating of change scale; QuickDASH, short form of the Disabilities of the Arm, Shoulder and Hand questionnaire.

TABLE 4	Mean Score Changes for DASH and QuickDASH Questionnaires, According to Each GRCS Grade

GRCS Score*	Subjects, n (%)	DASH Mean Change	QuickDASH Mean Change
Less than 0	7 (3%)	−11.3	−8.2
0	54 (21%)	3.6	4.3
+1	43 (17%)	11.4	13.2
+2	117 (46%)	19.8	21.9
+3	34 (13%)	35.3	39.1

Abbreviations: DASH, Disabilities of the Arm, Shoulder and Hand questionnaire; GRCS, global rating of change scale; QuickDASH, short form of the Disabilities of the Arm, Shoulder and Hand questionnaire.
GRCS less than 0 is worsened, 0 is unchanged, +1 is a little better, +2 is somewhat better, and +3 is a great deal better.
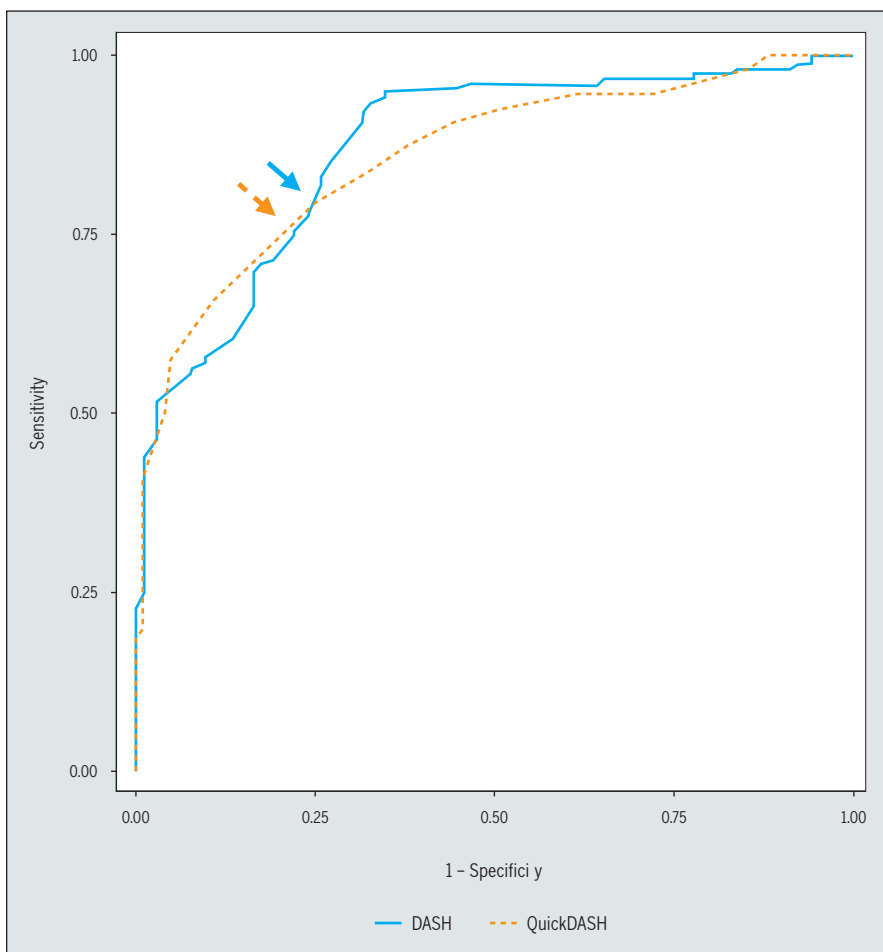
FIGURE 3. Comparison between the receiver-operating-characteristic curves of the DASH (solid line) and QuickDASH (dashed line), showing their overall accuracy in identifying an improvement according to a global rating of change scale (less than +2 versus +2 or greater). Arrows show the MCID value for each scale. Abbreviations: DASH, Disabilities of the Arm, Shoulder and Hand questionnaire; MCID, minimal clinically important difference; QuickDASH, short form of the Disabilities of the Arm, Shoulder and Hand questionnaire.

good correlation of their GRCS assessment with change in both the DASH and QuickDASH.

The responsiveness indices calculated with distribution-based methods showed values very close to those reported in the literature. For the DASH, our MDC_{90} was 10.8 points, compared to values ranging from 6.5 to 14.5 in the literature,[1,2,9,24,34,35] whereas the MDC_{95} reported on the DASH website and in the DASH manual ranges from 7.9 to 17.2 points.[24,40] In particular, our value is very close to that reported by DASH developers (MDC_{90} = 11 points).[2] Likewise, our MDC_{90} for the QuickDASH was 12.85 points, compared to values ranging from 11 to 15.7 points

in the literature,[1,14,23,27,30,31] whereas the MDC_{95} reported on the DASH website ranges from 16 to 20 points.[20] It is important to recognize that (a) the MDC values reflect the specific ICC and standard deviation values in each particular study, and therefore different groups of patients would be expected to generate different values; (b) changes below the MDC threshold could be interpreted at an individual level as random fluctuations in score rather than actual change[39]; and (c) the lower the reliability coefficient, the greater the SEM and the higher the MDC will be. Overall, these MDCs fall within the range of values reported by the DASH manual[24] and provide supportive

information for MCID estimates. Their variability in the literature is mainly due to differences in experimental conditions (eg, observers, scoring procedures, population under study).[24,41]

As for anchor-based methods, the first issue related to the appropriateness of cutoff values is the selection of the anchor. We used a 7-point GRCS (3 positives, 3 negatives, and an "unchanged" category) and considered patients with GRCS values of +2 or greater (+2, somewhat better; +3, a great deal better) as significantly improved and the others as not significantly improved.[28,42] In the literature, there is no agreement on the type of GRCS to use, the threshold at which to dichotomize the GRCS, or which groups to include in the analysis.[42] For example, the choice to include in the "changed" group those who reported only slight improvement (GRCS of +1, a little better) would inevitably lead to a lower cutoff point in the ROC analysis.[9] In addition, different criteria have been used to calculate and select both ROC cutoff values and MCIDs in DASH and QuickDASH studies.[2,3,9,25,27,36] Finally, the alternate use of raw scores[25] and 0-to-100 converted scores (rounded or not) represents a further potential source of confusion. Overall, any direct comparison of MCIDs is difficult, due to methodological issues that include type of anchor, calculation procedures, decision rules, and so on.[24,40] An additional problem is that some of the proposed MCIDs are lower than their respective MDC values.[9,27]

After triangulation of all our results, for the DASH, a change of 10.83 points was defined as the most acceptable MCID for moderate improvement, with good sensitivity (82%), specificity (74%), and classification accuracy (79%).[10] This MCID was inside the 95% CI for our ROC cutoff values, slightly superior to both our MDC_{90} (10.8 points) and the MDC_{95} (10.7 points) reported by Beaton et al,[3] and in line with the MCID (10 points; 95% CI: 5, 15) calculated in a sample of patients undergoing nonoperative treatment for forearm, wrist, and

hand disorders[36] and with mean change reported in patients declaring a small improvement in the present paper (11.4 points) and in those reported by Dawson et al[9] (about 10 points) and Beaton et al[3] (11.5-11.6 points).

Various authors have suggested that it would be better to define a range of MCID values rather than a fixed value,[10,11] and there are reasons to be skeptical about claims of a single MCID value.[24] Overall, due to our methodological procedure, with its main focus on the ROC-curve approach and an MCID value higher than MDC_{90} and not MDC_{95}, our threshold of 10.83 points could represent the lower boundary for a small range of reasonable MCIDs, in which the upper boundary could be represented by the 15 points proposed by the DASH website,[20] according to Beaton et al,[2] who just considered the AUC in ROC curves for score changes of −1, −5, −7, −10, −15, and −20. Also in this range is the MCID value (12.6 points) calculated by Schmitt and Di Fabio[35] in a sample of 78 patients suffering from a mixture of diagnoses involving the upper extremity. Lehman et al,[25] however, reported a slightly higher cutoff point in their ROC-curve analysis (16.7 points), because they privileged the cutoff value with higher specificity.

Similarly, for the QuickDASH, we identified a change of 15.91 points as the most adequate MCID. This is higher than its MDC_{90} (12.85 points) and corresponds to the best ROC cutoff value, with a trade-off between sensitivity (79%) and specificity (75%) very close to that of the DASH MCID and, again, good classification accuracy (78%). This MCID is in line with that calculated by Sorensen et al[36] (14 points; 95% CI: 9, 20) and with the MDC_{95} of 15.8 points in the original study by Beaton et al.[4,24] Moreover, it is higher than the 8-point MCID proposed by Mintken et al[27] for the QuickDASH; however, in that study the MDC_{90} was 11.2 points. Again, our threshold of 15.91 points could represent the lower boundary for a range of MCIDs having an upper boundary of 18 to 20 points, the limit

suggested by the DASH website.[40] In this range lies the MCID value of 19 points reported by Polson et al[30] in a small group of patients recruited from private physiotherapy practices.

The ranges of MCIDs we suggest here (DASH, 10.83 to 15 points; QuickDASH, 15.91 to 20 points) encompass most of the MCID values proposed for both scales, and are higher than their respective MDC values. Moreover, the 2 ranges represent a narrowing of the score range proposed for the DASH by its developers (8-17 points, with a mean of 13) and a confirmation of that for the QuickDASH, where the same developers considered the MDC_{95} as an interim proxy for their MCIDs.[20,24]

Taking into account the variation of MCID thresholds that can be found among populations, approaches, and methods used to calculate them,[3,40] the choice of the threshold within our range of MCIDs should be clinically driven. Using high thresholds for the DASH and QuickDASH, fewer people would be identified by the instrument as having shown a minimal clinically important change, but more people would be classified as not having shown a clinically important improvement[25]; this could reduce, for example, the risk of a too-early discharge in the clinical context. For an extensive discussion about these issues, we refer to the latest version of the DASH and QuickDASH user's manual.[24]

These MCID thresholds identify patients with a clinically important improvement, not those who have recovered. For this reason, to better understand the effects of treatment in some clinical settings, the construct of a return to "normal" functioning (that again could be linked to different indicators)[3] should also be taken into account, as we have done by calculating how many subjects both experienced an MCID and moved into a healthy functional range. This (conservative) result could be useful, where applicable, for interpreting clinical outcome studies and give us more confidence in clinical decision making.[26]

As an example, after a median number of 10 physical therapy sessions, about 60% of our patients—using either of the 2 scales—were positive responders, that is, achieved or surpassed the proposed MCID thresholds. Moreover, among the patients who reported a moderate to large improvement on the GRCS (+2 or greater), 85% showed a change equal to or higher than our MCIDs on the DASH, and 79% on the QuickDASH. Thirty-five to forty percent of subjects belonging to these last positive groups also arrived at a good final state, showing scores of 16 points or less, within a range consistent with that of healthy individuals.

Limitations

A number of potential limitations of this study should be mentioned. Caution is mandatory when interpreting and using these MCID values, particularly considering the intrinsic weaknesses of the GRCS.[7,22] The GRCS (and the MCID values derived from it) may suffer from subjective retrospective judgments of change (eg, due to recall bias or problematic patient ability to understand the context of improvement). Nonetheless, in our opinion, the problems of recall were unlikely to represent a significant issue, as treatment time was rather short (2-5 weeks) and featured periodic discussions of patients' health status.[22]

Moreover, as mentioned in the Introduction, when applying MCIDs in clinical research and practice, one must take into consideration that MCID values are dependent on selected characteristics of the sample (eg, age, disease group, baseline functional status, acuity of the medical condition, and potential for change), as well as on treatment and the time interval between testing.[24,43] Indeed, it has been recently emphasized that measurement error (and parameters derived from it) is often not constant across different levels of function and related scores,[33] and, because the interval-level scaling properties of the DASH and QuickDASH are not definitively proved, the MCID should be managed cautiously, particu-

larly when raw scores are near the range margins.[15]

The selection criteria of the present study could have posed a threat to the study's external validity. The sample was a cross-section of adults with upper-limb musculoskeletal disorders of different origin and severity (**TABLE 2**), recruited with a consecutive sampling method in a single rehabilitation facility. However, the distribution of baseline scores of the sample was consistent with other large studies on the DASH.[2] In addition, our rules for defining positive responders were tailored to our sample and setting. Different populations, rules, indicators, or cutoffs could lead to different results.

Finally, we cannot exclude that some specific (linguistic, cultural, or technical) characteristics of the Italian version of the DASH and QuickDASH might have influenced results, although these versions were the ones suggested by the DASH website,[20] obtained by means of a thorough forward/backward-translation process,[29] and further checked through a pilot test and expert analysis, without finding any major problem.

CONCLUSION

THIS STUDY PROPOSES MCID VALUES for the DASH and QuickDASH based on a comprehensive triangulation of distribution- and anchor-based approaches. Our results seem sound from both a psychometric and clinical point of view, and are in line with the previous literature. Our MCIDs for the DASH (10.83 points) and QuickDASH (15.91 points) could represent the lower boundary of a small range of MCIDs that could be proposed for different populations and contextual characteristics,[1,3,20] where the upper boundary would be represented by 15 points for the DASH and 20 points for the QuickDASH, as proposed by the DASH website.[20] On the other hand, we agree that there is a clear need for improvement and standardization of the MCID methodology.[40] With this premise, our results are likely to be

a useful step toward the ultimate goal of accurately classifying patients' response to treatment using the DASH and Quick-DASH. ⦿

■ KEY POINTS

FINDINGS: In a large sample of patients (n = 255) with upper-limb musculoskeletal disorders, we calculated the MCID for moderate improvement in the DASH (10.83 points) and QuickDASH (15.91 points) with a combination of distribution- and anchor-based (using the GRCS) approaches.

IMPLICATIONS: The combined use of multiple methods for defining the MCID is strongly suggested. At present, the studies that have applied anchor-based approaches seem to converge on a small range of MCID values for both scales. Our results increase confidence in interpreting change in the DASH and QuickDASH for use in clinical decision making.

CAUTION: Due to the variation of MCID thresholds among populations and methods, caution is needed when interpreting and using the published MCID values at the individual level, and there is a clear need for improvement and standardization of the MCID methodology.

REFERENCES

1. Angst F, Schwyzer HK, Aeschlimann A, Simmen BR, Goldhahn J. Measures of adult shoulder function: Disabilities of the Arm, Shoulder, and Hand Questionnaire (DASH) and its short version (QuickDASH), Shoulder Pain and Disability Index (SPADI), American Shoulder and Elbow Surgeons (ASES) Society standardized shoulder assessment form, Constant (Murley) Score (CS), Simple Shoulder Test (SST), Oxford Shoulder Score (OSS), Shoulder Disability Questionnaire (SDQ), and Western Ontario Shoulder Instability Index (WOSI). *Arthritis Care Res (Hoboken)*. 2011;63 suppl 11:S174-S188. http://dx.doi.org/10.1002/acr.20630

2. Beaton DE, Katz JN, Fossel AH, Wright JG, Tarasuk V, Bombardier C. Measuring the whole or the parts? Validity, reliability, and responsiveness of the Disabilities of the Arm, Shoulder and Hand outcome measure in different regions of the upper extremity. *J Hand Ther*. 2001;14:128-146.

3. Beaton DE, van Eerd D, Smith P, et al. Minimal

change is sensitive, less specific to recovery: a diagnostic testing approach to interpretability. *J Clin Epidemiol*. 2011;64:487-496. http://dx.doi.org/10.1016/j.jclinepi.2010.07.012

4. Beaton DE, Wright JG, Katz JN. Development of the QuickDASH: comparison of three item-reduction approaches. *J Bone Joint Surg Am*. 2005;87:1038-1046. http://dx.doi.org/10.2106/JBJS.D.02060

5. Bonett DG. Sample size requirements for estimating intraclass correlations with desired precision. *Stat Med*. 2002;21:1331-1335. http://dx.doi.org/10.1002/sim.1108

6. Bot SD, Terwee CB, van der Windt DA, Bouter LM, Dekker J, de Vet HC. Clinimetric evaluation of shoulder disability questionnaires: a systematic review of the literature. *Ann Rheum Dis*. 2004;63:335-341. http://dx.doi.org/10.1136/ard.2003.007724

7. Cook CE. Clinimetrics corner: the minimal clinically important change score (MCID): a necessary pretense. *J Man Manip Ther*. 2008;16:E82-E83.

8. Crosby RD, Kolotkin RL, Williams GR. Defining clinically meaningful change in health-related quality of life. *J Clin Epidemiol*. 2003;56:395-407. http://dx.doi.org/10.1016/S0895-4356(03)00044-1

9. Dawson J, Doll H, Boller I, et al. Comparative responsiveness and minimal change for the Oxford Elbow Score following surgery. *Qual Life Res*. 2008;17:1257-1267. http://dx.doi.org/10.1007/s11136-008-9409-3

10. de Vet HC, Terluin B, Knol DL, et al. Three ways to quantify uncertainty in individually applied "minimally important change" values. *J Clin Epidemiol*. 2010;63:37-45. http://dx.doi.org/10.1016/j.jclinepi.2009.03.011

11. 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. http://dx.doi.org/10.1186/1477-7525-4-54

12. Franchignoni F, Ferriero G, Giordano A, Sartorio F, Vercelli S, Brigatti E. Psychometric properties of QuickDASH – a classical test theory and Rasch analysis study. *Man Ther*. 2011;16:177-182. http://dx.doi.org/10.1016/j.math.2010.10.004

13. Franchignoni F, Giordano A, Sartorio F, Vercelli S, Pascariello B, Ferriero G. Suggestions for refinement of the Disabilities of the Arm, Shoulder and Hand Outcome Measure (DASH): a factor analysis and Rasch validation study. *Arch Phys Med Rehabil*. 2010;91:1370-1377. http://dx.doi.org/10.1016/j.apmr.2010.06.022

14. Gabel CP, Michener LA, Melloh M, Burkett B. Modification of the upper limb functional index to a three-point response improves clinimetric properties. *J Hand Ther*. 2010;23:41-51; quiz 52. http://dx.doi.org/10.1016/j.jht.2009.09.007

15. Grimby G, Tennant A, Tesio L. The use of raw scores from ordinal scales: time to end mal-

practice? *J Rehabil Med*. 2012;44:97-98. http://dx.doi.org/10.2340/16501977-0938

16. Gummesson C, Ward MM, Atroshi I. The shortened disabilities of the arm, shoulder and hand questionnaire (QuickDASH): validity and reliability based on responses within the full-length DASH. *BMC Musculoskelet Disord*. 2006;7:44. http://dx.doi.org/10.1186/1471-2474-7-44

17. Guyatt GH, Osoba D, Wu AW, Wyrwich KW, Norman GR. Methods to explain the clinical significance of health status measures. *Mayo Clin Proc*. 2002;77:371-383.

18. Hudak PL, Amadio PC, Bombardier C. Development of an upper extremity outcome measure: the DASH (disabilities of the arm, shoulder and hand) [corrected]. The Upper Extremity Collaborative Group (UECG). *Am J Ind Med*. 1996;29:602-608. http://dx.doi.org/10.1002/(SICI)1097-0274(199606)29:6<602::AID-AJIM4>3.0.CO;2-L

19. Hunsaker FG, Cioffi DA, Amadio PC, Wright JG, Caughlin B. The American Academy of Orthopaedic Surgeons outcomes instruments: normative values from the general population. *J Bone Joint Surg Am*. 2002;84-A:208-215.

20. Institute for Work and Health. Disabilities of the Arm, Shoulder and Hand outcome measure. Available at: http://www.dash.iwh.on.ca/. Accessed July 25, 2013.

21. Jaeschke R, Singer J, Guyatt GH. Measurement of health status. Ascertaining the minimal clinically important difference. *Control Clin Trials*. 1989;10:407-415.

22. Kamper SJ, Maher CG, Mackay G. Global rating of change scales: a review of strengths and weaknesses and considerations for design. *J Man Manip Ther*. 2009;17:163-170.

23. Kennedy CA, Beaton DE, Smith P, et al. Measurement properties of the QuickDASH (Disabilities of the Arm, Shoulder and Hand) outcome measure and cross-cultural adaptations of the QuickDASH: a systematic review. *Qual Life Res*. 2013;22:2509-2547. http://dx.doi.org/10.1007/s11136-013-0362-4

24. Kennedy CA, Beaton DE, Solway S, McConnell S, Bombardier C. *The DASH and QuickDASH Outcome Measure User's Manual*. 3rd ed. Toronto, Ontario, Canada: Institute for Work and Health; 2011.

25. Lehman LA, Sindhu BS, Shechtman O, Romero S, Velozo CA. A comparison of the ability of two upper extremity assessments to measure change in function. *J Hand Ther*. 2010;23:31-39; quiz 40. http://dx.doi.org/10.1016/j.jht.2009.09.006

26. Mann BJ, Gosens T, Lyman S. Quantifying clinically significant change: a brief review of methods and presentation of a hybrid approach. *Am J Sports Med*. 2012;40:2385-2393. http://dx.doi.org/10.1177/0363546512457346

27. Mintken PE, Glynn P, Cleland JA. Psychometric properties of the Shortened Disabilities of the Arm, Shoulder, and Hand Questionnaire (QuickDASH) and Numeric Pain Rating Scale in patients with shoulder pain. *J Shoulder Elbow Surg*. 2009;18:920-926. http://dx.doi.org/10.1016/j.jse.2008.12.015

28. Ostelo RW, de Vet HC, Knol DL, van den Brandt PA. 24-item Roland-Morris Disability Questionnaire was preferred out of six functional status questionnaires for post-lumbar disc surgery. *J Clin Epidemiol*. 2004;57:268-276. http://dx.doi.org/10.1016/j.jclinepi.2003.09.005

29. Padua R, Padua L, Ceccarelli E, et al. Italian version of the Disability of the Arm, Shoulder and Hand (DASH) questionnaire. Cross-cultural adaptation and validation. *J Hand Surg Br*. 2003;28:179-186.

30. Polson K, Reid D, McNair PJ, Larmer P. Responsiveness, minimal importance difference and minimal detectable change scores of the shortened disability arm shoulder hand (QuickDASH) questionnaire. *Man Ther*. 2010;15:404-407. http://dx.doi.org/10.1016/j.math.2010.03.008

31. Puga VO, Lopes AD, Shiwa SR, Alouche SR, Costa LO. Clinimetric testing supports the use of 5 questionnaires adapted into Brazilian Portuguese for patients with shoulder disorders. *J Orthop Sports Phys Ther*. 2013;43:404-413. http://dx.doi.org/10.2519/jospt.2013.4441

32. Revicki D, Hays RD, Cella D, Sloan J. Recommended methods for determining responsiveness and minimally important differences for patient-reported outcomes. *J Clin Epidemiol*. 2008;61:102-109. http://dx.doi.org/10.1016/j.jclinepi.2007.03.012

33. Romero S, Bishop MD, Velozo CA, Light K. Minimum detectable change of the Berg Balance Scale and Dynamic Gait Index in older persons at risk for falling. *J Geriatr Phys Ther*. 2011;34:131-137.

34. Roy JS, MacDermid JC, Woodhouse LJ. Measuring shoulder function: a systematic review of four questionnaires. *Arthritis Rheum*. 2009;61:623-632. http://dx.doi.org/10.1002/art.24396

35. Schmitt JS, Di Fabio RP. Reliable change and minimum important difference (MID) propor-

tions facilitated group responsiveness comparisons using individual threshold criteria. *J Clin Epidemiol*. 2004;57:1008-1018. http://dx.doi.org/10.1016/j.jclinepi.2004.02.007

36. Sorensen AA, Howard D, Tan WH, Ketchersid J, Calfee RP. Minimal clinically important differences of 3 patient-rated outcomes instruments. *J Hand Surg Am*. 2013;38:641-649. http://dx.doi.org/10.1016/j.jhsa.2012.12.032

37. Stratford PW, Binkley JM, Riddle DL. Health status measures: strategies and analytic methods for assessing change scores. *Phys Ther*. 1996;76:1109-1123.

38. Stratford PW, Goldsmith CH. Use of the standard error as a reliability index of interest: an applied example using elbow flexor strength data. *Phys Ther*. 1997;77:745-750.

39. Stratford PW, Riddle DL. When minimal detectable change exceeds a diagnostic test-based threshold change value for an outcome measure: resolving the conflict. *Phys Ther*. 2012;92:1338-1347. http://dx.doi.org/10.2522/ptj.20120002

40. Terwee CB, Roorda LD, Dekker J, et al. Mind the MIC: large variation among populations and methods. *J Clin Epidemiol*. 2010;63:524-534. http://dx.doi.org/10.1016/j.jclinepi.2009.08.010

41. Turner D, Schünemann HJ, Griffith LE, et al. The minimal detectable change cannot reliably replace the minimal important difference. *J Clin Epidemiol*. 2010;63:28-36. http://dx.doi.org/10.1016/j.jclinepi.2009.01.024

42. Turner D, Schünemann HJ, Griffith LE, et al. Using the entire cohort in the receiver operating characteristic analysis maximizes precision of the minimal important difference. *J Clin Epidemiol*. 2009;62:374-379. http://dx.doi.org/10.1016/j.jclinepi.2008.07.009

43. Wang YC, Hart DL, Stratford PW, Mioduski JE. Baseline dependency of minimal clinically important improvement. *Phys Ther*. 2011;91:675-688. http://dx.doi.org/10.2522/ptj.20100229

44. Wright AA, Cook CE, Baxter GD, Dockerty JD, Abbott JH. A comparison of 3 methodological approaches to defining major clinically important improvement of 4 performance measures in patients with hip osteoarthritis. *J Orthop Sports Phys Ther*. 2011;41:319-327. http://dx.doi.org/10.2519/jospt.2011.3515

@ MORE INFORMATION WWW.JOSPT.ORG