JUST A. VAN DER LINDE, MD, PhD [1] • **DERK A. VAN KAMPEN,** MD, PhD[2] • **LOES W.A.H. VAN BEERS,** MSc[1]
DEREK F.P. VAN DEURZEN, MD[1] • **DANIËL B.F. SARIS,** MD, PhD[3,4,5] • **CAROLINE B. TERWEE,** PhD[6]

The Responsiveness and Minimal Important Change of the Western Ontario Shoulder Instability Index and Oxford Shoulder Instability Score

Patient-reported outcome measures (PROMs) are widely used to assess functional limitations.[4,23,33] Two PROMs that are specifically designed for the evaluation of functional limitations due to shoulder instability are the Western Ontario Shoulder Instability Index (WOSI) and the Oxford Shoulder Instability Score (OSIS).[7,26]

STUDY DESIGN: Prospective cohort study.

BACKGROUND: Patient-reported outcome measurements (PROMs) are widely used to evaluate functional limitations. Considering PROMs for shoulder instability, information is lacking with regard to what constitutes a relevant change from baseline scores.

OBJECTIVES: To evaluate the responsiveness of the Western Ontario Shoulder Instability Index (WOSI) and the Oxford Shoulder Instability Score (OSIS) and estimate their minimal important change (MIC).

METHODS: One hundred five consecutive patients with shoulder instability completed 5 PROMs at baseline and at 6-month follow-up. The PROMs included the WOSI and OSIS, the Simple Shoulder Test, the Oxford Shoulder Score, and the Disabilities of the Arm, Shoulder and Hand assessment. Patients also rated their functional change on an anchor question at follow-up. Responsiveness was evaluated by testing 9 hypotheses regarding predefined correlations between the changes in PROM scores, by calculating the area under the receiver operating characteristic curve and by calculating the standardized response mean and effect size statistics. The MIC was determined by identifying the optimal cutoff on the receiver operating characteristic curve.

RESULTS: Seven out of 9 hypotheses (78%) were confirmed; as expected, a high correlation (0.77) was found between change scores of the WOSI and OSIS, whereas the correlations of the change scores of the WOSI and OSIS with those of general shoulder PROMs were slightly lower (0.61-0.75). The area under the curve was 0.83 (95% confidence interval: 0.75, 0.91) for the OSIS and 0.82 (95% confidence interval: 0.74, 0.90) for the WOSI. The MIC was about 6 points for the OSIS and about 14 points for the WOSI.

CONCLUSION: Both the WOSI and OSIS are able to measure change in shoulder function in patients with shoulder instability. The estimated MIC is 6 points for the OSIS (on a scale from 0 to 48) and 14 points for the WOSI (on a scale from 0 to 100). *J Orthop Sports Phys Ther 2017;47(6):402-410. doi:10.2519/jospt.2017.6548*

KEY WORDS: *clinimetrics, functional outcome measures, outcome assessment, shoulder*

Both are widely used and have shown good measurement properties for use in patients with shoulder instability.[28,31,41] The WOSI has been translated and found to have good measurement properties in Italian, German, Swedish, and Japanese patients.[5,17,19,20,26,35,43] Both the WOSI and the OSIS have also been translated and found to have good measurement properties in the Dutch-speaking population.[42,43]

Responsiveness and the minimal important change (MIC) are 2 important properties of tools to assess treatment outcome and are closely related to each other. Responsiveness is defined by the COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN) group as the ability of an instrument to detect a change over time, and should be high in an instrument that is used to evaluate (functional) changes.[30] The MIC is not a measurement property but an important measure of the interpretability of the instrument's scores, and is defined as the smallest change in score that patients perceive to be important.[15] The MIC is often defined as the change in score that can best discriminate between patients who have changed and patients who have not changed, based on an external criterion.[15]

[1]Department of Orthopaedic Surgery and Traumatology, Onze Lieve Vrouwe Gasthuis, Amsterdam, the Netherlands. [2]Department of Orthopaedic Surgery and Traumatology, Waterlandziekenhuis, Purmerend, the Netherlands. [3]Department of Orthopaedic Surgery and Traumatology, Universitair Medisch Centrum Utrecht, Utrecht, the Netherlands. [4]Reconstructive Medicine, University of Twente, Enschede, the Netherlands. [5]Department of Orthopedic Surgery, Mayo Clinic, Rochester, MN. [6]Department of Epidemiology and Biostatistics; EMGO Institute for Health and Care Research, VU University Medical Center, Amsterdam, the Netherlands. Institutional Review Board and ethical approval was obtained from the Onze Lieve Vrouwe Gasthuis (Amsterdam, the Netherlands) ethical committee (study number 12.097). The authors certify that they have no affiliations with or financial involvement in any organization or entity with a direct financial interest in the subject matter or materials discussed in the article. Address correspondence to Dr Just A. van der Linde, Orthocentre, Kareena Private Hospital, 86 Kareena Road, Carringbah, 2228, NSW, Australia. E-mail: javdlinde@gmail.com

It can thus be regarded as a threshold; exceeding change scores is likely to reflect a relevant change for the patient. This will allow better interpretation of the results of treatment.[24]

The first aim of this study was to evaluate and compare the responsiveness of the WOSI and OSIS. The second aim of this study was to assess the MICs of the WOSI and OSIS in terms of change scores that would be relevant for patients.

METHODS

Patient Population and Design

A cohort of patients was prospectively recruited between 2009 and 2014. Inclusion criteria were (1) aged 18 years or older and (2) diagnosed with shoulder instability by one of the doctors in the outpatient clinic, using patient history and physical examination, including a positive apprehension test. Both patients with primary and recurrent shoulder instability were included. Exclusion criteria were (1) large fractures (other than Hill-Sachs or bony Bankart lesions), (2) a frozen shoulder, and (3) the inability to understand the Dutch language. Ethical committee approval was obtained from the local ethical committee, and written informed consent was obtained from all participants.

Patients were asked to complete the WOSI and OSIS twice, alongside the Simple Shoulder Test (SST), the Oxford Shoulder Score (OSS), and the Disabilities of the Arm, Shoulder and Hand assessment (DASH), at baseline and after 6 months (follow-up). Baseline consultation occurred before the treatment started; treatment was either nonoperative or surgical. Because a change in shoulder function and not an intervention was assessed, both treatment modalities and any type of surgery were allowed. Patients who received surgical treatment completed follow-up approximately 6 months after surgery, when the postoperative pain was resolved and their rehabilitation completed.

Both PROMs were completed at baseline and follow-up at home, using either internet or paper-based administration. Patients used the same format across both instruments and time points. In the computer version, completion of an item was required to move to the next item. On paper, missing answers were verified by telephone. During completion, patients were able to withdraw from the study when they did not want to answer a question.

Patient-Reported Outcome Measurements

All PROMs are translated, and their validity is supported for the Dutch-speaking population.[2,42,44,45]

Western Ontario Shoulder Instability Index The WOSI was developed in Canada to be used as an outcome measure in clinical trials that evaluate treatment for patients with shoulder instability.[26] The instrument consists of 21 items that reflect 4 domains: physical symptoms (10 items), sport/recreation/work function (4 items), lifestyle function (4 items), and emotional function (3 items). Questions include, for example, "How much do you need to protect your arm during activities?" and "How much frustration do you feel because of your shoulder?" Responses are given on a 100-mm visual analog scale that ranges from no complaints (0 mm) to severe complaints (100 mm). Items are summarized for a total score ranging from 0 to 2100, where 0 indicates no limitations and 2100 indicates extreme limitations. The score can also be expressed as a percentage of normal shoulder function, where a score of 2100 reflects 0% and 0 reflects 100% of normal function.[25] Validity was originally tested by comparing WOSI scores to scores on the DASH and the University of California at Los Angeles shoulder rating scale, yielding correlations of 0.77 and 0.65, respectively. Other studies also found high correlations with the DASH (0.79 and 0.81).[6,43]

The WOSI's reliability for use in the Dutch language was supported by good internal consistency (Cronbach α = .96), an excellent test-retest reliability (intraclass correlation coefficient [ICC] = 0.92; 95% confidence interval [CI]: 0.88, 0.95), and no floor or ceiling effects. The smallest detectable change (SDC) was found to be 23.0 points (on a scale from 0 to 100).[43]

Oxford Shoulder Instability Score The OSIS was designed in the United Kingdom to assess the outcome of treatment for shoulder instability.[7] The instrument consists of 12 questions, each with 5 response categories. Examples of these questions are, "During the last 3 months, have you had any trouble (or worry) dressing because of your shoulder?" and "During the last 4 weeks, how much has the problem with your shoulder interfered with your sporting activities or hobbies?" In the original scoring system, answers were scored from 1 to 5 points, and summarized to a total score that ranged from 12 (least impaired) to 60 (most impaired). The scoring system was revised in 2009, in accordance with the revised scoring for the OSS, which originated from the same institute.[9] In the revised scoring system, answers are scored from 0 to 4, and the score is reversed; thus, the total score ranges from 0 (most impaired) to 48 (least impaired). We present our results in terms of the new scoring system.

Validity of the OSIS was originally tested by comparing OSIS scores to the Rowe and Constant scores, yielding correlations of 0.51 and 0.56, respectively. A high correlation was also observed with the DASH (0.82).[42]

The OSIS's reliability for use in the Dutch language was supported by good internal consistency (Cronbach α = .88) and excellent test-retest reliability (ICC = 0.87; 95% CI: 0.82, 0.91). The SDC was found to be 9.0 points (on a scale from 0 to48).[42]

Simple Shoulder Test The SST was designed in the United States to measure functional limitations in patients with common shoulder problems, including rotator cuff tears, degenerative osteoarthritis, and instability.[27,34] It contains 12 items (yes/no), with a total score that ranges from 0 to 12. Its validity was tested by comparing SST scores to scores on the

American Shoulder and Elbow Surgeons standardized shoulder assessment, yielding an observed correlation of 0.81. The SST's reliability was supported for use in the Dutch language by high internal consistency (Cronbach α = .78) and high test-retest reliability (ICC = 0.92; 95% CI: 0.95, 0.99).[44] The SDC was found to be 2.8 (on a scale from 0 to 12).[44]

Oxford Shoulder Score The OSS was developed in the United Kingdom for patients with general shoulder complaints.[8] The OSS contains 12 items (each question has 5 response categories), with a total score that ranges from 12 to 48. Its validity was originally tested by comparing OSS scores to scores on the Constant shoulder score and the Medical Outcomes Study 36-Item Short-Form Health Survey subscales, with observed correlations of −0.74 and −0.66, respectively. The OSS is considered to be responsive, and its reliability was supported for use in the Dutch language by high internal consistency (Cronbach α = .92) and high test-retest reliability (ICC = 0.98; 95% CI: 0.96, 0.99).[3,6] The SDC is 6.0 (on a scale from 0 to 48).[44]

Disabilities of the Arm, Shoulder and Hand Assessment The DASH was developed in Canada and the United States to measure physical functions and symptoms in patients with musculoskeletal disorders caused by any condition in any joint of the upper extremity.[22] It contains 30 items (each question has 5 response categories), with a total score that ranges from 100 to 0. The DASH was shown to be reliable, valid, and responsive for patients with shoulder disabilities across different countries.[1,12,21,40] Its reliability was supported for use in the Dutch language, for patients with disorders of the upper limb, by high internal consistency (Cronbach α = .95), and by test-retest reliability (Pearson correlation coefficient of 0.98).[45] The SDC was found to be 16.3 (on a scale from 100 to 0).[45]

External Anchor

The only difference between baseline and follow-up was the addition of an anchor question at follow-up, as a global rating of change.[16] This was formulated as, "How has your shoulder function changed compared to the first time you completed this questionnaire?" The item included 7 response categories that ranged from "completely recovered" to "worse than ever" (**TABLE 1**). The anchor question was asked prior to the PROMs to prevent the PROMs from influencing patients' judgment of functional change.

Statistical Analysis

Statistical analysis was performed using SPSS Version 18 (IBM Corporation, Armonk, NY). Change scores were calculated by subtracting the score at follow-up (6 months) from the score at baseline.

Responsiveness was assessed by 3 methods, based on the COSMIN recommendations.[29] First, hypotheses were tested about relationships with other instruments (also called *longitudinal validity*). The hypotheses were defined before data collection and based on clinical experience, knowledge from the literature about several PROMs, and a consensus among the study investigators. Seven hypotheses were predefined, and 2 hypotheses were added after a supplemental analysis. A high correlation (0.7 or greater) was expected between PROMs that are specifically designed to measure shoulder instability (WOSI and OSIS). A slightly lower correlation (0.6 or greater) was expected between shoulder instability–specific PROMs and PROMs designed to measure shoulder function in patients with more general shoulder problems (DASH, OSS, SST). Because the DASH, OSS, and SST all assess common or general shoulder problems, we did not distinguish between these 3 PROMs. A difference of 0.1 between high (0.7 or greater) and lower (0.6 or greater) correlations was chosen based on recommendation and use in other studies.[10,11,15] The supplemental analysis regarded expected high correlations between the external anchor question and the change scores of the WOSI and OSIS (0.7 or greater). Two-tailed Pearson correlation coefficients were used, because the mean change scores were normally distributed. The responsiveness of the WOSI and OSIS was considered adequate if at least 75% of the results were in accordance with the hypotheses.[37]

Second, the area under the receiver operating characteristic (ROC) curve was calculated.[17] A ROC curve was plotted in which the population was divided into an "important improved" group (completely recovered, much improved, and slightly improved) versus a "not improved" group (unchanged), based on the external anchor question. Patients whose scores deteriorated were excluded. The area under the curve (AUC) was calculated as a measure of responsive-

TABLE 1 The Mean Change Scores for the WOSI and OSIS According to the Functional Anchor

		Mean Change Scores*	
Functional Anchor (n = 105)	**n**	**WOSI**	**OSIS**
Completely recovered	15	23.7 (13.4, 34.0)	12.7 (8.2, 17.3)
Much improved	45	22.3 (16.1, 28.4)	12.4 (10.0, 14.9)
Slightly improved	9	14.6 (3.5, 25.7)	6.0 (0.4, 11.6)
Unchanged	28	0.1 (−4.4, 4.6)	1.9 (−0.4, 0.1)
Slightly worse	6	0.1 (−25.5, 25.6)	−4.2 (−8.4, 0.1)
Much worse	2	−29.3 (−98.9, 58.5)	−7.5 (−51.2, 37.0)
Worse than ever	0	NA	NA

Abbreviations: NA, not applicable; OSIS, Oxford Shoulder Instability Score; WOSI, Western Ontario Shoulder Instability Index.
**Values in parentheses are 95% confidence interval.*

ness using the original change scores. An AUC of at least 0.70 was considered adequate responsiveness.

Third, the standardized response mean (SRM) was calculated by dividing the mean change score by the standard deviation of the change. The effect size was calculated by dividing the mean change score by the standard deviation of the group's baseline scores. As the WOSI and OSIS aim to measure the same construct, the PROM with the highest SRM and effect size can be considered the most responsive.

The MIC was estimated as the optimal cutoff point on the ROC curve (ie, the value for which the sum of the proportions of misclassifications {[1 – sensitivity] + [1 – specificity]} was smallest).[32] Because a minimal important improvement may not be the same as a minimal important deterioration, we did not include patients whose scores deteriorated.

For illustrative purposes, the anchor-based MIC distribution was plotted,[13] showing the distribution of change scores in the "important improved" group and the "not improved" group. In this figure (**FIGURES 1A** and **1B**), change scores were categorized per 5 OSIS points (1-5 points of improvement, 6-10 points of improvement, etc) and 10 WOSI points (1-10 points of improvement, etc). This was necessary because both PROMs have a large scale (0-48 points for the OSIS and 0-100 points for the WOSI), and most change scores (eg, 6-point improvement, 7-point improvement, etc) occurred only once or twice.

For illustrative purposes, we also calculated the mean change scores (95% CI) on the WOSI and OSIS for the patients within each change group.

It is suggested that a minimum of 50 patients is adequate for assessing measurement properties.[18,38]

RESULTS

Descriptive Statistics

A flow chart of the included patients is presented in **FIGURE 2**. A total of 138 patients completed baseline measurements, of whom 33 declined to complete the same PROMs twice. The demographic characteristics of the remaining 105 patients are presented in **TABLE 2**. There were no missing data on the PROMs that were completed by these 105 patients.

Eight patients completed both PROMs on paper, but most (97 patients) preferred to complete both PROMs on the internet.

The mean age was 32 years, 70% of the patients were male, and 36% of the patients were treated surgically between baseline and follow-up. Forty-nine percent of the patients suffered their first shoulder dislocation more than 2 years before they completed the first PROMs.

TABLE 3 presents the mean ± SD PROM scores at baseline and follow-up and the mean change scores with 95% CIs. All shoulder-related PROM scores improved over the 6-month interval.

Responsiveness and MIC

TABLE 4 presents the expected and observed correlations (with 95% CIs) between change scores. Most of the observed correlations (78%) exceeded the expected correlations; only the correlations between the functional anchor and change scores were slightly lower than expected (0.64 and 0.63 for the WOSI and OSIS, respectively). According to the external anchor, 69 patients (66%) were improved, 28 patients (27%) were unchanged, and 8 patients (8%) were worse

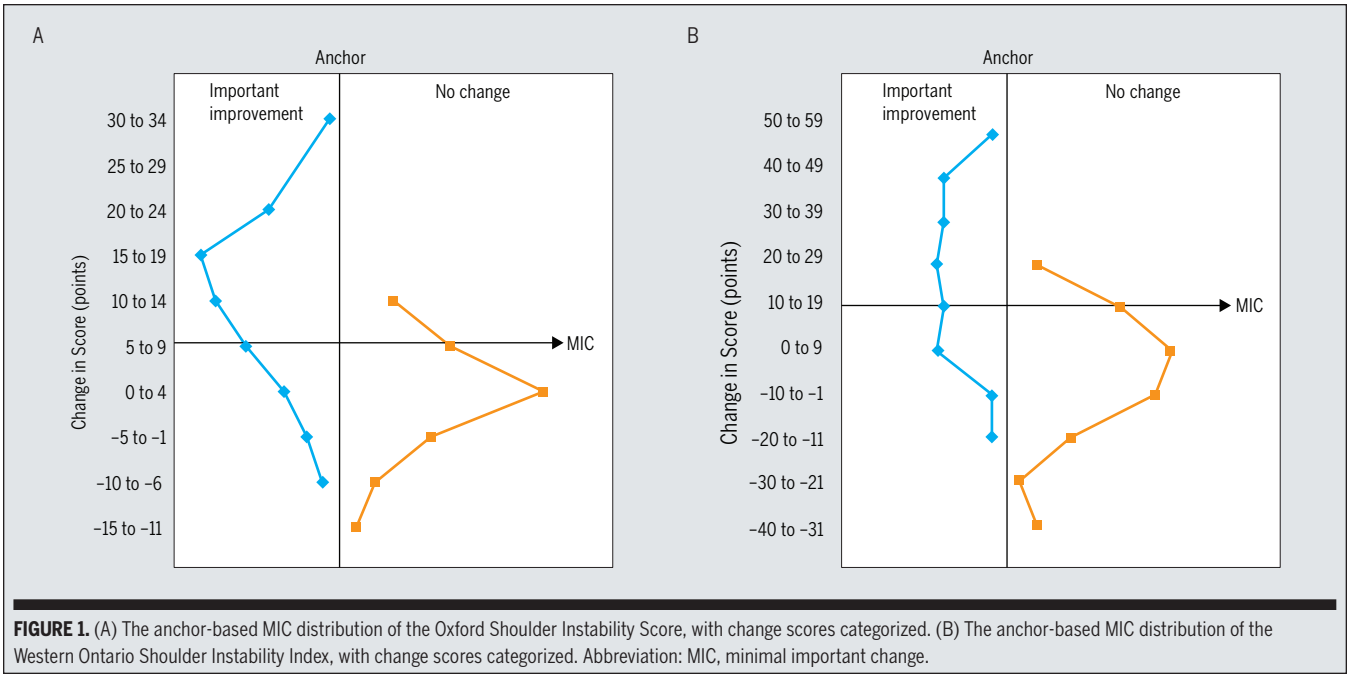

FIGURE 1. (A) The anchor-based MIC distribution of the Oxford Shoulder Instability Score, with change scores categorized. (B) The anchor-based MIC distribution of the Western Ontario Shoulder Instability Index, with change scores categorized. Abbreviation: MIC, minimal important change.

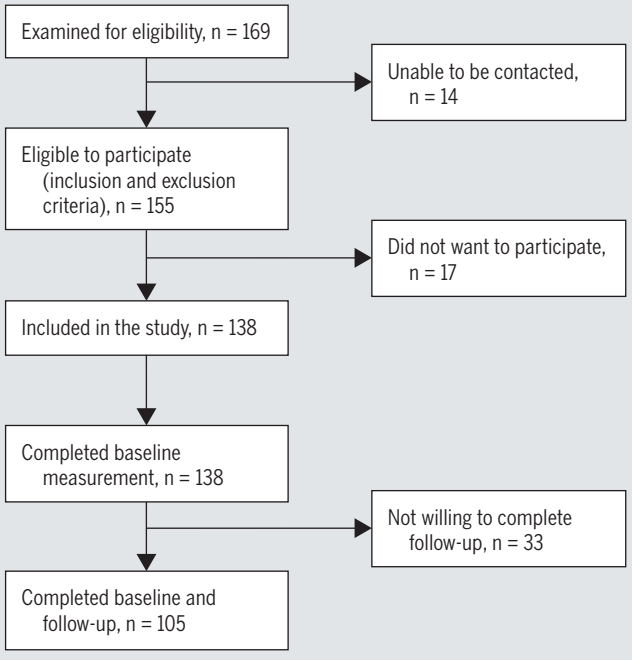

FIGURE 2. Flow chart of the included patients.

TABLE 2 Demographic Characteristics of the Study Population*

Characteristic	Value
Patients, n	105
Mean ± SD age, y	32 ± 12
Sex (male)	74 (70)
Treated surgically between baseline and follow-up	38 (36)
Interval from first instability to baseline measurement	
0-4 wk	8 (8)
1-6 mo	20 (19)
6 mo to 2 y	26 (25)
>2 y	51 (49)

**Values are n (%) unless otherwise indicated.*

at follow-up. The ROC curves were based on the 69 patients who reported that they improved and the 28 patients who reported that they were unchanged. These are presented in **FIGURES 3A** and **3B**. The AUC (calculated with the uncategorized scores) was 0.83 (95% CI: 0.75, 0.91) for the OSIS and 0.82 (95% CI: 0.74, 0.90) for the WOSI.

The SRM was 0.61 (13.6/22.3) for the WOSI and 0.86 (7.8/9.1) for the OSIS. The effect size was 0.25 for the WOSI (13.6/54.0) and 0.29 (7.8/27.3) for the OSIS.

The optimal cutoff value of the ROC curve, shown in **FIGURES 3A** and **3B** (ie, the point on the curve closest to the upper left corner), was used as an estimation of the MIC and corresponds to 6 points for the OSIS and 14 points for the WOSI. The sensitivity and specificity for these cutoffs are summarized in **TABLE 5**.

FIGURES 1A and **1B** present the anchor-based MIC distribution after change scores were categorized. **TABLE 1** presents the mean (95% CI) change scores for the WOSI and OSIS for each subgroup of patients according to the anchor question.

DISCUSSION

We found that the WOSI and OSIS have an adequate responsiveness: 7 out of 9 hypotheses could be confirmed (78%). Moreover, the AUCs were 0.82 (95% CI: 0.74, 0.90) and 0.83 (95% CI: 0.75, 0.91) for the WOSI and OSIS, respectively. This indicates that both PROMs are able to measure changes in shoulder function in patients with shoulder instability. The SRM was higher for the OSIS compared to the WOSI (0.86 versus 0.61), whereas the effect sizes were similar. This indicates that the variation in change scores was higher for the WOSI.

Using the ROC method, the MIC was estimated to be about 6 points for the OSIS and about 14 points for the WOSI.

To identify the MIC, we chose to include patients who reported "slightly improved" shoulder function in the "important improved" group. If these patients had been included in the unchanged group, the estimated MIC values would be slightly bigger, but, because only 9 patients declared themselves to be "slightly improved," the difference would have been small.[13] We reported the smaller MIC values obtained when including the slightly improved patients in the important improved group so as to capture all important changes.

A strong point of the study is the homogeneity of this patient population, including only patients with shoulder instability. However, there are several limitations to this study.

First, the 9 hypotheses that were tested to assess responsiveness are mostly based on the assumption that the correlation between PROMs specifically for shoulder instability will be stronger compared to the correlation between PROMs that are not specifically for shoulder instability. A limitation of this approach is

TABLE 3 PROM Scores at Baseline and Follow-up and the Mean Change Score

PROM*	Baseline†	Follow-up†	Mean Change‡
WOSI (0-100)	54.0 ± 22.3	69.4 ± 21.8	13.6 (9.5, 17.6)
OSIS (0-48)	27.3 ± 9.1	35.4 ± 8.9	7.8 (6.0, 9.6)
DASH (100-0)	22.2 ± 16.7	13.7 ± 12.7	-7.7 (-10.8, -4.7)
OSS (0-48)	36.3 ± 7.8	40.8 ± 5.6	4.1 (2.7, 5.4)
SST (0-12)	8.8 ± 3.1	10.0 ± 2.4	1.2 (0.7, 1.7)

Abbreviations: DASH, Disabilities of the Arm, Shoulder and Hand assessment; OSIS, Oxford Shoulder Instability Score; OSS, Oxford Shoulder Score; PROM, patient-reported outcome measurement; SST, Simple Shoulder Test; WOSI, Western Ontario Shoulder Instability Index.
**The ranges following each PROM reflect most impaired to least impaired function.*
†Values are mean ± SD.
‡Values in parentheses are 95% confidence interval.

TABLE 4 Correlations for Evaluating Responsiveness

Correlations Between Change Scores	Hypothesized	Observed*
WOSI and OSIS	≥0.7	0.77 (0.64, 0.89)†
WOSI and DASH	≥0.6	0.75 (0.62, 0.88)†
WOSI and OSS	≥0.6	0.68 (0.53, 0.82)†
WOSI and SST	≥0.6	0.69 (0.55, 0.83)†
WOSI and external anchor	≥0.7	0.64 (0.52, 0.77)
OSIS and DASH	≥0.6	0.73 (0.60, 0.87)†
OSIS and OSS	≥0.6	0.61 (0.45, 0.76)†
OSIS and SST	≥0.6	0.63 (0.48, 0.78)†
OSIS and external anchor	≥0.7	0.63 (0.52, 0.78)

Abbreviations: DASH, Disabilities of the Arm, Shoulder and Hand assessment; OSIS, Oxford Shoulder Instability Score; OSS, Oxford Shoulder Score; SST, Simple Shoulder Test; WOSI, Western Ontario Shoulder Instability Index.
**Values in parentheses are 95% confidence interval.*
†Correlations that met the prespecified hypotheses.

TABLE 5 Sensitivity and Specificity for the MIC Cutoffs of the OSIS and WOSI, Based on the Uncategorized Change Scores

	MIC Cutoff	Sensitivity	Specificity
OSIS	6 points	0.77	0.79
WOSI	14 points	0.70	0.89

Abbreviations: MIC, minimal important change; OSIS, Oxford Shoulder Instability Score; WOSI, Western Ontario Shoulder Instability Index.

that the responsiveness of the comparison instruments is not well studied.

Second, the distribution of change scores of the patients who reported themselves to be improved and that of the patients who reported that they had not changed showed considerable overlap. If the MICs of 6 points for the OSIS or 14 points for the WOSI are used to determine which patients are importantly changed, misclassification can occur. Moreover, the sample size within the group classified as unchanged was small (n = 28).

Third, the correlations between the external anchor question and the change scores of the WOSI and OSIS were slightly lower (0.64 and 0.63) than expected (0.70 or greater). This might be explained by the fact that the WOSI and OSIS actually measure a broader concept than function, because they also include questions about emotional aspects (eg, the OSIS item, "During the last 4 weeks, how often has your shoulder been 'on your mind'—how often have you thought about it?" and the WOSI item, "How concerned are you about your shoulder becoming worse?"). Another explanation could be recall bias: when patients have difficulty recalling their function at baseline, it can also be difficult to indicate how much their function has improved 6 months later, which may explain these lower correlations. In general, a retrospective anchor has the disadvantage that it can potentially be influenced by other (not shoulder related) problems. Such a problem either at baseline or during the interval could influence the way patients value their change in shoulder function. However, the observed correlations were still higher than recommended values of 0.30[32] or 0.50[15] for estimating the MIC, so we feel that our analyses were appropriate.

Responsiveness of the WOSI was previously assessed by Kirkley et al,[26] by Salomonsson et al,[35] and by Cacchio et al,[5] who determined the effect size and SRM. The SRM and the effect size that we found for the WOSI were lower compared to those previously described (0.61 versus 0.93 to 1.94 and 0.25 versus 1.47 to 1.67).[5,26,35] This may be explained by the fact that effect size and SRM are measures of the magnitude of change scores, not the validity of change scores. Therefore, the differences in SRM and effect size between studies probably reflects the differences in treatment effects rather than differences in the responsiveness of the PROMs. Therefore, we prefer responsiveness assessment by hypothesis testing regarding correlations between

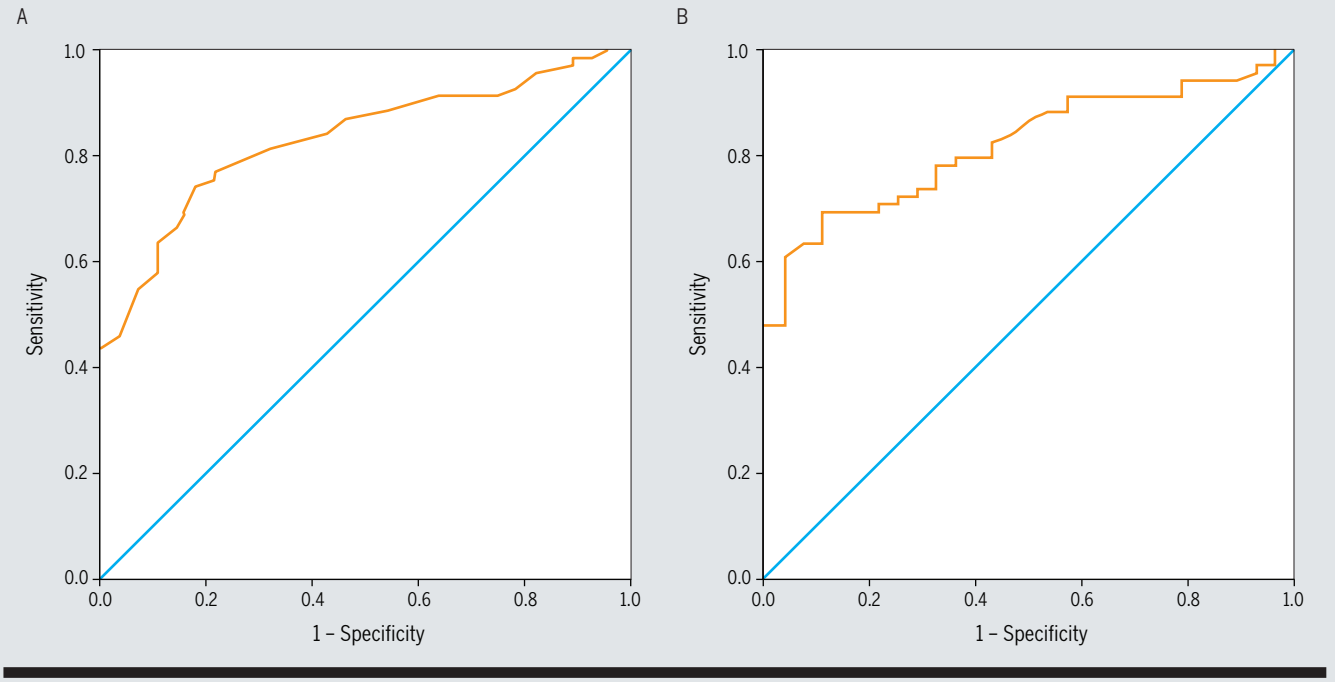

FIGURE 3. The receiver operating characteristic curves of the (A) Oxford Shoulder Instability Score and (B) Western Ontario Shoulder Instability Index.

change scores on different PROMs or by calculating the area under the ROC curve. Effect size and SRM can only be used for a head-to-head comparison of the relative responsiveness of PROMs that measure the same construct, such as the WOSI and OSIS, within the same population.

According to our expectations, the highest correlation was found between change scores of the instability-specific WOSI and OSIS. However, a high correlation was also observed between change scores of the DASH and the WOSI (0.75; 95% CI: 0.62, 0.88) and OSIS (0.73; 95% CI: 0.60, 0.87). The WOSI and OSIS also had high correlations with the DASH when their validity was originally tested, indicating that they may (partly) measure the same construct.[42,43] The DASH could thus potentially be useful in studies evaluating the shoulder function of patients with shoulder instability. The fact that correlations between change scores of instability-specific PROMs are higher compared to correlations of the instability-specific PROMs with change scores of the other shoulder PROMs is no proof that these are substantially better to use.

Change scores (including the MIC) should be interpreted in relation to the smallest change that can be detected beyond the measurement error in individual patients (the SDC).[14,36,45] Ideally, the measurement error should be smaller than the MIC of the same instrument to be 95% sure that a change in an individual patient as large as the MIC is not due to measurement error.[39] The SDC was previously calculated as 23.0 points for the WOSI and 9.0 points for the OSIS.[42,43] This means that individual patients should improve at least 23.0 and 9.0 points, respectively, to ensure that their change is not a result of the measurement error. Regarding the WOSI, the MIC was estimated to be 14 points, which cannot be distinguished with 95% confidence from measurement error (SDC, 23.0 points) in individual patients. Regarding the OSIS, the MIC was estimated to be 6 points, which can also not be distinguished with 95% confidence from measurement error (SDC, 9.0 points). Clinicians should, therefore, be aware that only a change in score that exceeds both the MIC and SDC can be considered a real and relevant change with 95% confidence. The interpretation of WOSI and OSIS change scores in individual patients should therefore be considered with caution.

CONCLUSION

Both the WOSI and OSIS are able to measure change in shoulder function in patients with shoulder instability. The MIC was estimated to be 6 points for the OSIS and 14 points for the WOSI. The MIC values are within the measurement error of the WOSI and OSIS, and therefore change scores in individual patients should be considered with caution. ◉

KEY POINTS

FINDINGS: Both the Western Ontario Shoulder Instability Index (WOSI) and Oxford Shoulder Instability Score (OSIS) have an adequate responsiveness. The estimated minimal important change (MIC) is 6 points for the OSIS (on a scale from 0 to 48) and 14 points for the WOSI (on a scale from 0 to 100).

IMPLICATIONS: Both the WOSI and OSIS are able to measure change in shoulder function in patients with shoulder instability.
CAUTION: The smallest detectable change (SDC) values for the WOSI and the OSIS have been reported in previous research to be 23 and 9 points, respectively. These values are larger than the MIC values reported here (14 points and 6 points, respectively); thus, clinicians should be aware that, to be confident that a change in score can be considered a real and relevant change, the change should exceed both the MIC and SDC.

REFERENCES

1. Angst F, Goldhahn J, Drerup S, Aeschlimann A, Schwyzer HK, Simmen BR. Responsiveness of six outcome assessment instruments in total shoulder arthroplasty. *Arthritis Rheum*. 2008;59:391-398. https://doi.org/10.1002/art.23318
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. https://doi.org/10.1016/S0894-1130(01)80043-0
3. Berendes T, Pilot P, Willems J, Verburg H, te Slaa R. Validation of the Dutch version of the Oxford Shoulder Score. *J Shoulder Elbow Surg*. 2010;19:829-836. https://doi.org/10.1016/j.jse.2010.01.017
4. 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.
5. Cacchio A, Paoloni M, Griffin SH, et al. Cross-cultural adaptation and measurement properties of an Italian version of the Western Ontario Shoulder Instability Index (WOSI). *J Orthop Sports Phys Ther*. 2012;42:559-567. https://doi.org/10.2519/jospt.2012.3827
6. Christiansen DH, Frost P, Falla D, Haahr JP, Frich LH, Svendsen SW. Responsiveness and minimal clinically important change: a comparison between 2 shoulder outcome measures. *J Orthop Sports Phys Ther*. 2015;45:620-625. https://doi.org/10.2519/jospt.2015.5760
7. Dawson J, Fitzpatrick R, Carr A. The assessment of shoulder instability. The development and validation of a questionnaire. *J Bone Joint Surg Br*. 1999;81:420-426.
8. Dawson J, Fitzpatrick R, Carr A. Questionnaire on the perceptions of patients about shoulder surgery. *J Bone Joint Surg Br*. 1996;78:593-600.
9. Dawson J, Rogers K, Fitzpatrick R, Carr A. The Oxford shoulder score revisited. *Arch Orthop Trauma Surg*. 2009;129:119-123. https://doi.org/10.1007/s00402-007-0549-7
10. de Boer MR, Terwee CB, de Vet HC, Moll AC, Völker-Dieben HJ, van Rens GH. Evaluation of cross-sectional and longitudinal construct validity of two vision-related quality of life questionnaires: the LVQOL and VCM1. *Qual Life Res*. 2006;15:233-248. https://doi.org/10.1007/s11136-005-1524-9
11. de Groot IB, Reijman M, Terwee CB, et al. Validation of the Dutch version of the Hip disability and Osteoarthritis Outcome Score. *Osteoarthritis Cartilage*. 2007;15:104-109. https://doi.org/10.1016/j.joca.2006.06.014
12. Desai AS, Dramis A, Hearnden AJ. Critical appraisal of subjective outcome measures used in the assessment of shoulder disability. *Ann R Coll Surg Engl*. 2010;92:9-13. https://doi.org/10.1308/003588410X12518836440522
13. de Vet HC, Ostelo RW, Terwee CB, et al. Minimally important change determined by a visual method integrating an anchor-based and a distribution-based approach. *Qual Life Res*. 2007;16:131-142. https://doi.org/10.1007/s11136-006-9109-9
14. de Vet HC, Terwee CB, Knol DL, Bouter LM. When to use agreement versus reliability measures. *J Clin Epidemiol*. 2006;59:1033-1039. https://doi.org/10.1016/j.jclinepi.2005.10.015
15. de Vet HC, Terwee CB, Mokkink LB, Knol DL. *Measurement in Medicine: A Practical Guide*. Cambridge, UK: Cambridge University Press; 2011.
16. 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. https://doi.org/10.1186/1477-7525-4-54
17. Deyo RA, Centor RM. Assessing the responsiveness of functional scales to clinical change: an analogy to diagnostic test performance. *J Chronic Dis*. 1986;39:897-906.
18. Donner A, Eliasziw M. Sample size requirements for reliability studies. *Stat Med*. 1987;6:441-448.
19. Hatta T, Shinozaki N, Omi R, et al. Reliability and validity of the Western Ontario Shoulder Instability Index (WOSI) in the Japanese population. *J Orthop Sci*. 2011;16:732-736. https://doi.org/10.1007/s00776-011-0141-4
20. Hofstaetter JG, Hanslik-Schnabel B, Hofstaetter SG, Wurnig C, Huber W. Cross-cultural adaptation and validation of the German version of the Western Ontario shoulder instability index. *Arch Orthop Trauma Surg*. 2010;130:787-796. https://doi.org/10.1007/s00402-009-1033-3
21. Huang H, Grant JA, Miller BS, Mirza FM, Gagnier JJ. A systematic review of the psychometric properties of patient-reported outcome instruments for use in patients with rotator cuff disease. *Am J Sports Med*. 2015;43:2572-2582. https://doi.org/10.1177/0363546514565096
22. Hudak PL, Amadio PC, Bombardier C, The Upper Extremity Collaborative Group. Development of an upper extremity outcome measure: the DASH (disabilities of the arm, shoulder and hand) [corrected]. *Am J Ind Med*. 1996;29:602-608. https://doi.org/10.1002/(SICI)1097-0274(199606)29:6<602::AID-AJIM4>3.0.CO;2-L
23. Janse AJ, Gemke RJ, Uiterwaal CS, van der Tweel I, Kimpen JL, Sinnema G. Quality of life: patients and doctors don't always agree: a meta-analysis. *J Clin Epidemiol*. 2004;57:653-661. https://doi.org/10.1016/j.jclinepi.2003.11.013
24. Judge A, Arden NK, Kiran A, et al. Interpretation of patient-reported outcomes for hip and knee replacement surgery: identification of thresholds associated with satisfaction with surgery. *J Bone Joint Surg Br*. 2012;94:412-418. https://doi.org/10.1302/0301-620X.94B3.27425
25. Kirkley A, Griffin S, Dainty K. Scoring systems for the functional assessment of the shoulder. *Arthroscopy*. 2003;19:1109-1120. https://doi.org/10.1016/j.arthro.2003.10.030
26. Kirkley A, Griffin S, McLintock H, Ng L. The development and evaluation of a disease-specific quality of life measurement tool for shoulder instability. The Western Ontario Shoulder Instability Index (WOSI). *Am J Sports Med*. 1998;26:764-772.
27. Lippitt SB, Harryman DT, 2nd, Matsen FA, 3rd. A practical tool for evaluating function: the Simple Shoulder Test. In: Matsen FA, 3rd, Fu FH, Hawkins RJ, eds. *The Shoulder: A Balance of Mobility and Stability*. Rosemont, IL: American Academy of Orthopaedic Surgeons; 1993:501-518.
28. Mohtadi NG, Chan DS, Hollinshead RM, et al. A randomized clinical trial comparing open and arthroscopic stabilization for recurrent traumatic anterior shoulder instability: two-year follow-up with disease-specific quality-of-life outcomes. *J Bone Joint Surg Am*. 2014;96:353-360. https://doi.org/10.2106/JBJS.L.01656
29. Mokkink LB, Terwee CB, Knol DL, et al. The COSMIN checklist for evaluating the methodological quality of studies on measurement properties: a clarification of its content. *BMC Med Res Methodol*. 2010;10:22. https://doi.org/10.1186/1471-2288-10-22
30. Mokkink LB, Terwee CB, Patrick DL, et al. The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. *J Clin Epidemiol*. 2010;63:737-745. https://doi.org/10.1016/j.jclinepi.2010.02.006
31. Provencher MT, McCormick F, Dewing C, McIntire S, Solomon D. A prospective analysis of 179 type 2 superior labrum anterior and posterior repairs: outcomes and factors associated with success and failure. *Am J Sports Med*. 2013;41:880-886. https://doi.org/10.1177/0363546513477363
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. https://doi.org/10.1016/j.jclinepi.2007.03.012

33. Rouleau DM, Faber K, MacDermid JC. Systematic review of patient-administered shoulder functional scores on instability. *J Shoulder Elbow Surg*. 2010;19:1121-1128. https://doi.org/10.1016/j.jse.2010.07.003

34. Roy JS, MacDermid JC, Faber KJ, Drosdowech DS, Athwal GS. The Simple Shoulder Test is responsive in assessing change following shoulder arthroplasty. *J Orthop Sports Phys Ther*. 2010;40:413-421. https://doi.org/10.2519/jospt.2010.3209

35. Salomonsson B, Ahlström S, Dalén N, Lillkrona U. The Western Ontario Shoulder Instability Index (WOSI): validity, reliability, and responsiveness retested with a Swedish translation. *Acta Orthop*. 2009;80:233-238. https://doi.org/10.3109/17453670902930057

36. 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. https://doi.org/10.2522/ptj.20120002

37. Terwee CB, Bot SD, de Boer MR, et al. Quality criteria were proposed for measurement properties of health status questionnaires. *J Clin Epidemiol*. 2007;60:34-42. https://doi.org/10.1016/j.jclinepi.2006.03.012

38. Terwee CB, Mokkink LB, Knol DL, Ostelo RW, Bouter LM, de Vet HC. Rating the methodological quality in systematic reviews of studies on measurement properties: a scoring system for the COSMIN checklist. *Qual Life Res*. 2012;21:651-657. https://doi.org/10.1007/s11136-011-9960-1

39. Terwee CB, Roorda LD, Knol DL, De Boer MR, De Vet HC. Linking measurement error to minimal important change of patient-reported outcomes. *J Clin Epidemiol*. 2009;62:1062-1067. https://doi.org/10.1016/j.jclinepi.2008.10.011

40. Thoomes-de Graaf M, Scholten-Peeters GG, Schellingerhout JM, et al. Evaluation of measurement properties of self-administered PROMs aimed at patients with non-specific shoulder pain and "activity limitations": a systematic review. *Qual Life Res*. 2016;25:2141-2160. https://doi.org/10.1007/s11136-016-1277-7

41. van der Linde JA, van Kampen DA, Terwee CB, Dijksman LM, KleinJan G, Willems WJ. Long-term results after arthroscopic shoulder stabilization using suture anchors: an 8- to 10-year follow-up. *Am J Sports Med*. 2011;39:2396-2403. https://doi.org/10.1177/0363546511415657

42. van der Linde JA, van Kampen DA, van Beers LW, van Deurzen DF, Terwee CB, Willems WJ. The Oxford Shoulder Instability Score; validation in Dutch and first-time assessment of its smallest detectable change. *J Orthop Surg Res*. 2015;10:146. https://doi.org/10.1186/s13018-015-0286-5

43. van der Linde JA, Willems WJ, van Kampen DA, van Beers LW, van Deurzen DF, Terwee CB. Measurement properties of the Western Ontario Shoulder Instability Index in Dutch patients with shoulder instability. *BMC Musculoskelet Disord*. 2014;15:211. https://doi.org/10.1186/1471-2474-15-211

44. van Kampen DA, van Beers LW, Scholtes VA, Terwee CB, Willems WJ. Validation of the Dutch version of the Simple Shoulder Test. *J Shoulder Elbow Surg*. 2012;21:808-814. https://doi.org/10.1016/j.jse.2011.09.026

45. Veehof MM, Sleegers EJ, van Veldhoven NH, Schuurman AH, van Meeteren NL. Psychometric qualities of the Dutch language version of the Disabilities of the Arm, Shoulder, and Hand questionnaire (DASH-DLV). *J Hand Ther*. 2002;15:347-354. https://doi.org/10.1016/S0894-1130(02)80006-0 EFERENCES