Measures of adult shoulder function

Buchbinder, R.; Ramiro, S.; Huang, H.M.; Gagnier, J.J.; Jia, Y.X.; Whittle, S.L.

American College of Rheumatology
Empowering Rheumatology Professionals

Arthritis Care & Research
Vol. 72, No. S10, October 2020, pp 250–293
DOI 10.1002/acr.24230

Measures of Adult Shoulder Function

Rachelle Buchbinder,[1] Sofia Ramiro,[2] Hsiaomin Huang,[3] Joel J. Gagnier,[4] Yuanxi Jia,[5] and Samuel L. Whittle[6]

INTRODUCTION

In the first systematic review of randomized controlled trials evaluating efficacy of interventions for shoulder pain published in 1998, none of the 31 included trials utilized a measure of function (1). Since then, at least 50 instruments to measure adult shoulder function have been developed and used in trials for shoulder pain (2). A 2011 review of nine of these tools, chosen based on having been cited in at least 20 references and for which psychometric testing had been reported, was published in a special issue of this journal devoted to patient outcome measures relevant to rheumatology (3).

This updated review includes six of these nine tools and three new tools. Eight were chosen on the basis of being the most commonly used in randomized controlled trials in the last 5 years that were identified by searching Ovid MEDLINE and Ovid EMBASE from 2015 through December 9, 2019, using search strategies developed by Cochrane Musculoskeletal to identify common shoulder disorders, combined with Cochrane's highly sensitive search strategies for randomized controlled trials. We also included Patient-Reported Outcomes Measurement Information System (PROMIS) Physical Function (PF) and Upper Extremity Computerized Adaptive Tests (CATs). Studies evaluating the measurement properties of the nine tools in shoulder conditions were identified by combining the relevant tool and search terms for common shoulder disorders with a search filter developed by Terwee et al for identifying such studies (4).

There have been many psychometric studies of our chosen measures of adult shoulder function since the 2011 review. Many of the tools have been cross-culturally adapted into multiple languages, and these studies also provide important information regarding psychometric properties. Our review summarizes the available information about how these measures perform for different patient populations in different settings.

SHOULDER PAIN AND DISABILITY INDEX (SPADI)

Description

Purpose. The SPADI is a patient-reported outcome measure of shoulder pain and function. The original version was published in 1991 and developed as a joint-specific measure for any disorders of the shoulder joint for use in an outpatient setting (5). It was developed by a panel of rheumatologists and a physical therapist. It initially comprised 20 items that the panel considered to be measures of shoulder pain and function; seven items were subsequently removed because of inadequate test-retest reliability or poor correlation with shoulder range of motion (ROM), resulting in a 13-item scale.

Content or domains. The SPADI comprises two subscales: pain and disability.

Number of items. It consists of 13 items in total, including five items in the pain subscale and eight items in the disability subscale.

Response options/scale. All items were originally rated using a visual analog scale (VAS) (5). More recent versions have most commonly used an 11-point (0-10) numerical rating scale (NRS) (6). The anchors for each of the pain items are 0 = no pain and 10 = worst pain imaginable, and the anchors for the

Rachelle Buchbinder's work is supported by an Australian National Health and Medical Research Council Senior Principal Research Fellowship.

[1]Rachelle Buchbinder, MBBS, MSc, PhD, FRACP, FAHMS: Monash University and Cabrini Institute, Melbourne, Victoria, Australia; [2]Sofia Ramiro, MD, MSc, PhD: Leiden University Medical Center, Leiden, The Netherlands, and Zuyderland Medical Center, Heerlen, The Netherlands; [3]Hsiaomin Huang, MPH: Stanford University, Stanford, California; [4]Joel J. Gagnier, ND, MSc, PhD: University of Michigan, Ann Arbor; [5]Yuanxi Jia, PhD: Johns Hopkins University, Baltimore, Maryland; [6]Samuel L. Whittle, MBBS, MClinEpi, FRACP: Monash University and Cabrini Institute, Melbourne, Victoria, Australia, and The Queen Elizabeth Hospital, Woodville South, South Australia, Australia.

No potential conflicts of interest relevant to this article were reported.

Address correspondence to Rachelle Buchbinder, MBBS, MSc, PhD, FRACP, FAHMS, Department of Clinical Epidemiology, Monash University, 4 Drysdale St, Malvern, Victoria, Australia 3144. E-mail: rachelle.buchbinder@monash.edu.

Submitted for publication April 16, 2020; accepted in revised form April 21, 2020.

eight disability items are 0 = no difficulty and 10 = so difficult it requires help.

Recall period for items. One week

Cost to use. Free of charge.

How to obtain. The SPADI is printed in various references (7). A free PDF version is online at https://www.tac.vic.gov.au/files-to-move/media/upload/spi.pdf or it is available as an online calculator at https://www.orthotoolkit.com/spadi/.

Practical application

Method of administration. It is a self-assessment tool that can be completed using pen and paper or an electronic version. It may also be administered via the telephone (8).

Scoring. Each item is scored from 0 (best) to 10 (worst). A minimum of two-thirds of the questions must be answered in each subscale in order to calculate a score. The subscore is the sum of scored items divided by the maximum possible score multiplied by 100%. The total SPADI score is the unweighted mean of the pain and disability subscales.

Score interpretation. Possible scores range from 0 (best) to 100 (worst). There are no cutoff points to indicate severity. Normative data are available from a sample of 635 healthy volunteers in Australia (n = 323) and Canada (C = 312) (9). Subjects were included if they had no diagnosed shoulder pathology in the dominant arm and had no active shoulder pathology or shoulder surgery in the previous 3 years. Participants without a history of shoulder problems had a lower (better) mean score (3.3 on a 0-100 scale) than those who reported shoulder problems more than 3 years ago (mean score 6.1; $P < 0.0001$). Women had a higher (worse) mean score (4.2) than men did (2.7) after adjustment for nationality ($P = 0.026$), and scores increased with age.

Respondent time to complete. Median time to complete is 2 minutes (range 1-4 minutes) (10).

Administrative burden. Scoring is straightforward, and no special software or equipment is required. Some electronic tools include automated calculation of the final score. Administration and scoring takes 5 minutes (5).

Translations/adaptations. The SPADI is translated into multiple languages. Cross-cultural validation has been performed for the following languages: Spanish (11), Chinese (12), Arabic (13), Danish (14), Norwegian (15), Dutch (16), Indian (Tamil) (17), Hindi (18), Greek (19), Turkish (20), Brazilian Portuguese (21), Persian (22), Thai (23), Nepali (24), Italian (25), and German (26).

Psychometric information

Floor and ceiling effects. No floor or ceiling effects have been found in patients with rotator cuff disease (27) or after shoulder arthroplasty (28).

Reliability. There is evidence for high internal consistency, with high Cronbach's α scores for the total SPADI scale and the individual subscales in various settings, including shoulder disorders in outpatients (Cronbach's α: total SPADI = 0.96, disability = 0.95, and pain = 0.89) (29), population-based individuals with shoulder pain (Cronbach's α: total SPADI = 0.92, disability = 0.90, and pain = 0.85) (8), and Spanish patients with shoulder pain/dysfunction after surgery for breast cancer (Cronbach's $\alpha = 0.965$) (30).

Test-retest reliability was moderate in the original development study (intraclass correlation coefficient [ICC] = 0.66; 95% confidence interval [CI] 0.42-0.81), although only 37 subjects with shoulder pain were included (5). Higher ICCs have been reported in subsequent studies in various populations, including patients with general shoulder pain (ICC = 0.91) (31), adhesive capsulitis (ICC = 0.89 [95% CI 0.82-0.93], pain subscale ICC = 0.85 [95% CI 0.76-0.91], disability subscale ICC = 0.86 [95% CI 0.78-0.92]) (32), and Spanish patients with shoulder pain/dysfunction after surgery for breast cancer (ICC = 0.992) (30).

Validity. Analysis of structural validity suggests that the English SPADI consists of two factors (pain and disability), which load approximately onto the two subscales. Although factor analyses in the development paper (5) and in a large randomized controlled trial of subjects with full-thickness rotator cuff tears (33) produced two factors that did not map onto the subscales with complete fidelity, other factor analyses in larger populations of community-dwelling individuals with shoulder pain have demonstrated more distinct loading of pain and disability items onto separate factors, suggesting that the SPADI is indeed bidimensional (7,8).

A Rasch model analysis of 1030 patients referred for physical therapy for shoulder pain also demonstrated a bidimensional structure; however, there was evidence of differential item functioning for some items in the disability subscale (eg, washing hair and putting on a shirt were more difficult for women than men, and putting on trousers was more difficult for people aged 60 or older). This suggests greater structural validity for the pain subscale than the disability subscale and implies that the two subscales should be reported separately (34).

There is evidence for moderate to high correlation between SPADI scores and other generic and shoulder-specific pain and disability measures. This has been demonstrated for general shoulder disorders and specific conditions, including rotator cuff disease and adhesive capsulitis, in various settings, including primary care, hospital outpatients, and the general community. Pearson's or Spearman's correlations to other instruments or measures and the population they have been measured in are as follows:

- Shoulder ROM: $r = 0.55$ to 0.80 in patients with shoulder pain (5).
- Disabilities of Arm, Shoulder, and Hand (DASH): $r = 0.93$ following shoulder arthroplasty (28), and $r = 0.55$ in participants in two clinical trials of treatments for adhesive capsulitis (35).
- American Shoulder and Elbow Surgeons (ASES) Society Shoulder Score: $r = 0.81$ following shoulder arthroplasty (28), and $r = 0.77$ in patients referred to an upper extremity clinic for shoulder problems (36).
- Simple Shoulder Test (SST): $r = 0.74$ in patients referred to an upper extremity clinic for shoulder problems (36).
- Constant Murley Shoulder Scale (CS): $r = 0.82$ following shoulder arthroplasty (28), $r = 0.53$ in degenerative or inflammatory shoulder disease referred for surgery (37), and $r = 0.56$ in patients enrolled in physical therapy for shoulder dysfunction (38).
- Oxford Shoulder Score (OSS): $r = 0.85$ in patients attending a specialist shoulder clinic with subacromial impingement (39), $r = 0.74$ in degenerative or inflammatory shoulder disease referred for surgery (37); and Cronbach's $r = -0.74$ ($r = -0.71$ and -0.72 for the pain and disability subscales, respectively) in Turkish patients with shoulder problems (40).
- Shoulder Disability Questionnaire: $r = 0.57$ in patients with shoulder pain in primary care (10).
- Shoulder Rating Questionnaire (SRQ): $r = 0.83$ in patients with shoulder pain in primary care (10).
- Sickness Impact Profile: $r = 0.57$ in patients with shoulder pain in outpatient physical therapy clinic (41), and $r = 0.45$ in community volunteers with shoulder pain (42).
- Short Form 36 (SF-36) physical component scale (PCS): $r = -0.46$ in patients with shoulder pain in the general community (8).
- Health Assessment Questionnaire (HAQ): $r = 0.55$ in participants in two clinical trials of treatments for adhesive capsulitis (35).

Responsiveness. The total SPADI score exhibits good responsiveness as measured by the area under the curve (AUC) of the receiver operating characteristic (ROC) curve. The AUC was 0.87 (95% CI 0.79-0.95) in subjects in primary care receiving nonsurgical treatments for shoulder pain (10), ranged from 0.81 (95% CI 0.78-0.84) to 0.85 (95% CI 0.82-0.88) in subjects receiving physical therapy for shoulder pain (43), and ranged from 0.74 (95% CI 0.64-0.83) to 0.85 (95% CI 0.76-0.93) in clinical trials of interventions for adhesive capsulitis (35).

The SPADI has also been found to have greater, or at least comparable, responsiveness compared with other shoulder-specific measures. Reported effect sizes (ESs) and standardized response means (SRMs) of the SPADI total score and subscales in various shoulder conditions and settings are as follows:

- Total shoulder arthroplasty: ES = 2.10 and SRM = 1.72 (SPADI pain: ES = 2.12, SRM = 1.71; SPADI function: ES = 1.77, SRM = 1.51) (44). This was better than the DASH (ES = 1.19) and comparable with the ASES (ES = 2.13).
- Trials of oral steroids and hydrodilatation in adhesive capsulitis: ES = 1.20 to 1.64 and SRM = 1.27 to 1.68. Greater responsiveness was observed for the SPADI compared with the Croft Index (ES = 0.87-1.21) or DASH (ES = 0.55-0.83) (35).
- Physical therapy for shoulder pain: ES = 1.26 and SRM = 1.38 (41) at 6 weeks; ES = 1.26 to 1.71 and SRM = 1.35 to 1.75 at 6 months. This was similar to the DASH short version (QuickDASH) (ES = 1.04, SRM =1.26 at 6 weeks; ES =1, SRM = 1.56 at 6 months) (43).
- Rotator cuff surgery or total shoulder arthroplasty: SRM = 1.23 (31); this was better than the SRQ (SRM = 0.65) and comparable with the SST (SRM = 1.05).
- Occupational or physical therapy for various upper extremity disorders: ES = 1.21 and SRM = 1.08; this was similar to the DASH (ES =1.06, SRM = 1.08) and SF-36 PCS (ES = 1.20, SRM = 1.07) (45).
- Spanish patients with shoulder pain/dysfunction after surgery for breast cancer: ES = 0.59 and SRM = 0.82 (30).

Minimally important differences. The minimal detectable change (MDC) for the SPADI total score has been estimated to be 18.1 points (musculoskeletal upper extremity problems) (45) and 17 to 19.7 points (Norwegian SPADI in patients receiving treatment for adhesive capsulitis and rotator cuff disease, respectively) (15,32).

The minimal clinically important difference (MCID) has been estimated to be 13.2 points (musculoskeletal upper extremity problems) (45), 20 points (Norwegian SPADI in patients receiving nonsurgical treatment for rotator cuff disease) (27), and 8 to 10 points (shoulder pain presenting to primary care) (6,10). Variation in the estimated MCIDs is likely to reflect differences in methodology and populations; however, it is possible that in some situations the MCID may be smaller than the MDC. It has been suggested that a change of approximately 20 points in the SPADI is necessary to infer clinically important change (27).

Generalizability. The SPADI is a shoulder-specific measure that has been used in populations with various shoulder disorders, primarily nonspecific shoulder pain or rotator cuff disorders but also in adhesive capsulitis and after shoulder arthroplasty.

Use in clinical trials. The SPADI had been used in almost 50 randomized controlled trials of interventions for various shoulder disorders through the end of 2015 (2). Recent examples include trials of therapeutic deep heat for shoulder pain (46), exercise versus physical therapy for rotator cuff disease (47),

cervicothoracic manual therapy (48), glucocorticoid injections for shoulder pain (49), intra-articular hyaluronate and tramadol injections for adhesive capsulitis (50), and extracorporeal shockwave therapy for shoulder pain (51). In some cases in which the SPADI was used as an outcome measure for trials performed in countries where English is not the primary language, validated translations were used (51), but in many cases it was unclear whether this was the case (46,50).

Critical appraisal of overall value to the rheumatology community

Strengths. The SPADI is brief, is easy to administer and score, has good overall evidence for construct validity and responsiveness, and has been used and tested in numerous settings. It is readily obtained at no cost, and cross-cultural validation has been performed for several languages.

Caveats and cautions. Factor analysis and Rasch modelling suggest that the SPADI should be treated as separate subscales; however, there may not be clear demarcation between underlying concepts in the two scales, and the disability subscale may not provide true interval-level measurement (33,34). The upper anchor label for the disability subscale (so difficult it requires help) is potentially ambiguous because the perceived requirement for help may vary according to the level of help available to the individual respondent. The MCID may be smaller than the MDC, and therefore a change of at least the MDC (17-20 points) is required to be confident of a clinically important change.

Clinical usability. The SPADI is useful in a clinical context for assessing both joint-specific pain and function. It is brief, easy to administer and score, and responsive to change.

Research usability. The SPADI is brief, easy to administer, and responsive. It is valid for intervention and population-level studies.

AMERICAN SHOULDER AND ELBOW SURGEONS SOCIETY STANDARDIZED SHOULDER ASSESSMENT FORM

Description

Purpose. The ASES Society Shoulder Assessment form was developed to be a baseline measure of shoulder function applicable to all patients regardless of diagnosis (52). The requirements were that it should be easy to use, assess activities of daily living (ADLs), and include a patient self-evaluation section. A draft version was developed based on a review of all relevant forms that existed at the time, and it was modified based on two rounds of feedback from the ASES Society's membership. In 1998, the original ASES shoulder form was changed to the modified ASES by deleting two ADL items (sleep on painful side and throw ball overhead) and adding five new ADL items (open a jar of food, cut with a knife, use a phone, do up buttons, and carry shopping bag) to make a whole extremity questionnaire (31). This version does not appear to be used for shoulder conditions.

Content or domains. The patient self-assessment section comprises the following three domains: pain, ADLs, and instability. The physician assessment comprises the following four domains: ROM, signs, strength, and instability. Most commonly, however, only the two domains of patient self-assessment of pain and ADLs are used.

Number of items. The full ASES consists of 18 patient self-evaluation items (six for pain, ten for ADLs, and two for instability) and 29 physician-assessed items (five for ROM, 11 for signs, five for strength, and eight for instability). There is also one item that asks the patient to indicate where the pain is on a diagram and one open item that asks the physician about physical findings. Most commonly, only the 16 items evaluating the patient-reported assessment of pain and ADLs are used.

Response options/scale. *Pain.* For four items (Are you having pain in your shoulder? Do you have pain in your shoulder at night? Do you take pain medication [aspirin, ibuprofen, acetaminophen, etc.]? Do you take narcotic pain medication [codeine or stronger]?) the response option is yes/no. One item asks the respondent to specify how many pills they take each day (average), and one item is a 0 to 10 pain VAS (How bad is your pain today?) with anchors of 0 (no pain at all) to 10 (pain as bad as it can be). If there is pain in the shoulder, then the respondent marks the site on the front and back of a diagram.

ADLs. Response options for all ten items (put on a coat, sleep on your painful affected side, wash back/do up bra in back, manage toileting, comb hair, reach a high shelf, lift 10 lbs above shoulder, throw a ball overhand, do usual work, and do usual sport) are four-point ordinal Likert scales in which 0 = unable to do, 1 = very difficult, 2 = somewhat difficult, and 3 = not difficult. For the last two items (do usual work and do usual sport) the respondent is asked to list what these are.

Instability. For one item (does your shoulder feel unstable [as if it is going to dislocate]?) the response option is yes/no, and the other item is a 0 to 10 VAS (how unstable is your shoulder?) with anchors of 0 (very stable) to 10 (very unstable).

Physician-assessed ROM. Preferably using a goniometer, the physician measures both active and passive motion in forward elevation (maximum arm-trunk angle), external rotation (arm comfortably at side), external rotation (arm at 90° abduction), internal rotation (highest posterior anatomy reached with thumb),

and cross-body abduction (antecubital fossa to the opposite acromion).

Physician-assessed signs. Four items (supraspinatus/greater tuberosity tenderness, acromioclavicular joint tenderness, biceps tendon tenderness [or rupture], and other tenderness [list]) have 0 to 3 Likert-scale responses (0 = none, 1 = mild, 2 = moderate, and 3 = severe). The other seven items (impingement I, impingement II, impingement III, subacromial crepitus, scars, atrophy, and deformity) are all yes/no responses. If scars or atrophy are present, the respondent is asked to indicate the location, and if a deformity is present, the respondent is asked to describe it.

Physician-assessed strength. For one item (testing affected by pain), the response option is yes/no. The remaining four items (forward flexion, abduction, external rotation, and internal rotation) are each measured on 0 to 5 Likert scales (0 = no contraction, 1 = flicker, 2 = movement with gravity eliminated, 3 = movement against gravity, 4 = movement against some resistance, and 5 = normal power).

Physician-assessed instability. For four items (anterior translation, posterior translation, interior translation [sulcus sign], and anterior apprehension), the response options for instability are graded as 0 if absent, 1 if mild (0 to 1 cm translation), 2 if moderate (1- to 2-cm translation or translates to the glenoid rim), 3 if severe (greater than 2-cm translation or over the rim of the glenoid). For four items (reproduced symptoms, voluntary instability, positive relocation test, and generalized laxity), the response options are yes/no. The final item asks if there are any other physical signs and has an open field.

Recall period for items. For patient-reported outcomes, the recall period is not stated, except for some items in which the recall period is "today."

Cost to use. Free of charge.

How to obtain. It is printed in the original reference (52). The patient self-report section can be accessed and scored from various sites, including https://www.orthopaedicscore.com/scorepages/patient_completed_score.html, https://www.aaos.org/uploadedFiles/American%20Shoulder%20and%20Elbow%20Surgeons%20Standardized%20Shoulder%20Assessment%20Form.pdf, andhttps://www.orthotoolkit.com/ases/.

Practical application

Method of administration. The full ASES uses both patient self-assessment and physician assessment, but most commonly, only the self-assessment component assessing pain and ADLs is used. The remaining information applies to this abbreviated version of the ASES.

Scoring. The ASES score is comprised of only the patient self-evaluations of pain and ADLs, with equal weight given to degree of pain experienced by the patient (50 points) and the cumulative ADL score (50 points). The pain score is reversed (by subtracting the score from 10) and multiplied by 5 (so that a higher score is better). The cumulative ADL score is out of 30 and is multiplied by 5/3. The formula is as follows:

$$[(10 - \text{VAS pain score}) \times 5] + [5/3 \times \text{cumulative ADL score}]$$

This derives a score of a possible range of 0 to 100. Some studies also report separate ASES pain and function (ADL) subscale scores. The instability items, the remaining five pain items, and the physician assessment are not included in the ASES score. However, one study has devised a scoring method for physician assessment (28).

Score interpretation. A higher ASES score indicates better pain and disability (0 = worst and 100 = best). A missing rule or distinct cutoffs to reflect severity have not been published. Normative data are available based on a sample of 343 patients (aged 6-87 years) from an outpatient orthopedic center who were being seen for conditions unrelated to the shoulder (patients with prior shoulder problems were also excluded) (53). Overall, the mean (SD) ASES score was 92.2 (14.5) points, and the score decreased with age. Those aged 60 years or older had decreased ability to lift above shoulder level and reach behind the back when compared with younger cohorts.

Normative data are also available in 635 asymptomatic, healthy volunteers in Australia (n = 323) and Canada (n = 312) (9). Participants were excluded if they had a history of active shoulder pathology or a history of recent surgery (within the last 3 years) or joint arthroplasty. People without a history of shoulder pathology reported a higher (better) mean ASES score compared with those with a history of a shoulder problems (96.7 [range 0-100] versus 93.0; Wilcoxon's rank sum test $P = 0.0003$). No differences in scores were observed in people with and without current wrist or elbow problem or handedness. Women had slightly lower scores (95 in women versus 97 in men; $P = 0.03$), and scores declined with increasing age.

Respondent time to complete. The ASES takes less than 5 minutes to complete (52,54). All items are easy to read and understand and are not suggestive or emotionally sensitive. Missing data are very rare.

Administrative burden. The patient section can be administered without the clinical section, and the score computation is easy and can be implemented in any database.

Translations/adaptations. The ASES is translated into multiple languages. Cross-cultural validation has been performed in German (55), Italian (56), Brazilian Portuguese (57,58), Spanish (59), Finnish (60), Turkish (61), and Tunisian Arabic (62) versions.

Psychometric information

Floor and ceiling effects. Many studies have reported acceptable floor and ceiling effects across different patient populations, including patients with planned surgery for rotator cuff disease or glenohumeral arthritis (0% floor and ceiling), with instability (0% floor; 1.3% ceiling) (63), following shoulder arthroplasty (1% floor; 8% ceiling) (28,55), with impingement, with adhesive capsulitis or glenohumeral arthritis (0% floor and ceiling) (56), with rotator cuff disease (2.3% floor and ceiling) (64), and with rotator cuff repair, arthroplasty or physical therapy for impingement or adhesive capsulitis (<4% floor and ceiling) (65). One study in people with impingement reported ASES function subscale floor and ceiling effects of 5% and 22.4% respectively (66).

Reliability. The internal consistency of the ASES has been found to range from Cronbach's α of 0.61 to 0.96 across various study populations with different shoulder conditions, including in a mixed population referred for physical therapy with some postsurgery patients (0.86) (67); in a mixed population with impingement, adhesive capsulitis, and glenohumeral arthritis (0.85) (56); in populations with instability (0.61), rotator cuff disease (0.64), and glenohumeral arthritis (0.62) (63); in a population post shoulder arthroplasty (0.96) (55); in a population with various shoulder conditions (nonsurgical: 0.89 to 0.92; after surgery: 0.91) (68); in a mixed outpatient population (0.88) (61); in a population with rotator cuff disease, including tears or adhesive capsulitis (0.813) (62); in a population with rotator cuff disease, glenohumeral, or acromioclavicular arthritis and instability (0.88, 95% CI 0.84-0.91) (60); in a mixed population with mainly rotator cuff tears (0.91) (59); and in a population with shoulder dysfunction (0.794) (58). One study that included people with rotator cuff disease, superior labral anterior posterior (SLAP) lesions, and instability following surgery reported acceptable internal consistency for the pain (0.711) and ADL (0.850) subscales (69).

Good to excellent test-retest reliability is reported across many studies involving various study populations with differing shoulder conditions. These include a mixed population referred for physical therapy including some postsurgery patients (ICC = 0.84, 95% CI 0.75-0.91) (67); a population with various shoulder conditions (nonsurgical: ICC = 0.84, 95% CI 0.66-0.92; after surgery: ICC = 0.91, 95% CI0.82-0.96) (68); a population with impingement, adhesive capsulitis, and glenohumeral arthritis (ICC = 0.91) (56); a population with rotator cuff disease, instability, and glenohumeral arthritis (ICC = 0.94, 95% CI 0.88-0.97) (63); a post–shoulder arthroplasty population (ICC = 0.93, 95% CI 0.90-0.95) (55); a population with rotator cuff disease, including tears or adhesive capsulitis (ICC = 0.96, 95% CI 0.918-0.981) (62); a mixed outpatient population (ICC = 0.94 [0.95 and 0.86 for pain and ADL subscales, respectively]) (61); a population with rotator cuff disease, glenohumeral or acromioclavicular arthritis, and instability (ICC = 0.83, 95% CI 0.70-0.90) (60); a population with rotator cuff repair, arthroplasty, or physical therapy impingement or adhesive capsulitis (ICC = 0.82) (65); a population with shoulder dysfunction (ICC = 0.75) (58); and a population with no shoulder pain (ICC = 0.96) (53).

One study reported excellent item reliability (whether patients with similar shoulder function answer the questions similarly; Pearson's correlation coefficient 0.98) and moderate person reliability (how precisely a test discriminates between patients of different abilities or reliably ranks respondents; Pearson's correlation coefficient 0.48) (64). One study of patients with impingement reported a good person reliability 0.86 (66), whereas one study in patients with rotator cuff disease reported person reliability to be fair (0.48) and inferior to that of the SST (0.71, moderate) and the PROMIS PF CAT (0.93, excellent) (64).

Validity. One study has examined the Spanish version of the patient-reported component of the ASES (pASES) using both confirmatory factor and Rasch analysis (59). Factor loadings for confirmatory factor analysis were more than 0.40, and the Rasch model confirmed the unidimensionality of the scale, although ADL item 10 (do usual sport) was suggested to be uninformative. Another study found that the ASES is likely not unidimensional, with 27.9% unexplained variance, which is consistent with its inclusion of both pain and function items (64). This compared with an unexplained variance of only 4.5% for the PROMIS PF CAT and 8.4% for the SST. Although the SST also includes pain and function items, only two of 12 questions relate to pain in contrast with the ASES, in which these contribute 50% of the final score.

Pearson's or Spearman's correlations of the pASES total score to other instruments are as follows:

- SPADI: 0.81 (28) and 0.92 (55) in patients post arthroplasty, –0.942 for post total arthroplasty and –0.932 for reverse hemiarthroplasty (70); 0.82 in a mixed outpatient population (61).
- Western Ontario Rotator Cuff Index (WORC): 0.81 in patients who were receiving physical therapy for impingement or after rotator cuff repair, acromioplasty, or decompression surgery (71).
- Western Ontario Stability Index (WOSI): 0.15 in a mixed population with rotator cuff disease, isolated SLAP lesions, and shoulder instability after surgery (69).
- Rowe, rating sheet for Bankart repair score: 0.15 in a mixed population with rotator cuff disease, isolated SLAP lesions,

and shoulder instability after surgery (69).

- DASH: 0.79 (28) and 0.84 (55) in patients post arthroplasty; 0.92 in patients with impingement, adhesive capsulitis, and glenohumeral arthritis (56); and 0.69 in patients with shoulder dysfunction (58).
- CS: 0.71 in patients post arthroplasty (28); 0.36 in a mixed population with rotator cuff disease, isolated SLAP lesions, and shoulder instability after surgery (69); 0.62 in a mixed population (the majority with rotator cuff tears) (59); and 0.871 in patients with rotator cuff tears (72).
- SST: 0.35 in a mixed population with rotator cuff disease, isolated SLAP lesions, and shoulder instability after surgery (69); 0.73 in a mixed outpatient population, including those with rotator cuff disease, glenohumeral or acromioclavicular arthritis, and instability (60); 0.536 for rotator cuff disease (64); and –0.89 post total arthroplasty and –0.87 for reverse hemiarthroplasty (70).
- University of California, Los Angeles (UCLA) Shoulder Rating Scale: 0.38 in a mixed population with rotator cuff disease, isolated SLAP lesions, and shoulder instability after surgery (69); very high relation with arthroscopic rotator cuff population overall ($r = 0.91$; $P < 0.001$), with a moderate correlation in the preoperative period ($r = 0.67$; $P < 0.001$), a high correlation at 6 months after surgery ($r = 0.87$; $P < 0.001$), and a very high relation at 12 and 24 months after surgery ($r = 0.90$ and 0.92, respectively; $P < 0.001$) (73).
- Rotator Cuff Quality of Life Index: 0.70 in patients who were receiving physical therapy for impingement or after rotator cuff repair, acromioplasty, or decompression surgery (71).
- Single Assessment (or *a*) Numerical Evaluation score: 0.75 in patients following rotator cuff repair, 0.88 following rotator cuff revision, and 0.78 in SLAP repair (0.8 overall) (74); 0.85 in rotator cuff repair, 0.72 for total shoulder replacement, and 0.82 for physical therapy (65).
- PROMIS PF CAT: 0.581 in rotator cuff disease (64); 0.43 in planned surgery for rotator cuff tear (75); and 0.694 in impingement, with 0.664 for the ASES function subscale and 0.426 for the pain subscale (66).
- PROMIS PF Upper Extremity CAT: 0.72 in shoulder pain (excluding patients with prior rotator cuff surgery, shoulder surgery in last 6 months, or partial rotator cuff tear) (76) and 0.59 in planned surgery for rotator cuff tear (75).
- PROMIS Pain Interference CAT: –0.43 in planned surgery for rotator cuff tear (75) and 0.729 in impingement, with 0.667 for the ASES function subscale and 0.594 for the pain subscale (66).
- SF-36 bodily pain: 0.60 in patients with impingement, adhesive capsulitis, and glenohumeral arthritis (56); 0.65 in patients post arthroplasty (55); 0.64 in a mixed population (61); 0.60 in shoulder dysfunction (58); 0.05 in a mixed population with rotator cuff disease, isolated SLAP lesions, and shoulder instability after surgery (69); 0.68 in a mixed outpatient population including those with rotator cuff disease, glenohumeral or acromioclavicular arthritis, and instability (60); and 0.74 in a mixed population, the majority of whom had rotator cuff tears (59).
- SF-36 PCS: 0.64 in patients post arthroplasty (28,55); 0.48 in patients with impingement, adhesive capsulitis, and glenohumeral arthritis (56); 0.40 in a mixed population referred for physical therapy, including some after surgery (67); 0.20 in a mixed population with rotator cuff disease, isolated SLAP lesions, and shoulder instability after surgery (69); 0.57 in a mixed outpatient population including those with rotator cuff disease, glenohumeral or acromioclavicular arthritis, and instability (60); and 0.65 in a mixed population, the majority of whom had rotator cuff tears (59).
- SF-36 physical functioning: 0.47 in patients with impingement, adhesive capsulitis, and glenohumeral arthritis (56); 0.57 in patients post arthroplasty (55); 0.41 in a mixed population referred for physical therapy, including some after surgery (67); 0.35 in a mixed population (61); 0.50 for shoulder dysfunction (58); 0.27 in a mixed population with rotator cuff disease, isolated SLAP lesions, and shoulder instability after surgery (69); 0.51 in a mixed outpatient population including those with rotator cuff disease, glenohumeral or acromioclavicular arthritis, and instability (60); and 0.59 in a mixed population, the majority of whom had rotator cuff tears (59).
- Clinician-assessed component of the ASES: 0.48 in patients post arthroplasty (28).

Spearman's correlation coefficients of the ASES scale to other instruments in a population with no current shoulder pathology were as follows: $r = 0.65$ for the modified UCLA Scale, $r = 0.48$ for the CS, $r = 0.71$ for the OSS, $r = –0.78$ for the SPADI, and $r = –0.61$ for the Stanmore Percentage of Normal Shoulder Assessment (SPONSA) (9).

Responsiveness. The ASES has been found to have good or at least comparable responsiveness compared with other shoulder-specific measures. Reported ESs and SRMs in various shoulder conditions and settings are as follows:

- Shoulder arthroplasty (rheumatoid or osteoarthritis): ES = 2.13, SRM = 1.81 (44).
- Mixed population, the majority with rotator cuff tears, followed to 6 months post treatment: standardized ES = 0.80, SRM = 0.75 (the ES was greater among those who received surgery versus other treatments) (59).
- Subacromial steroid for calcific tendinitis: ES = 1.65 to 1.84 (77).
- Mixed population referred for physical therapy, including some after surgery: ES = 1.39, SRM = 1.54 (67).
- Impingement or after rotator cuff repair, acromioplasty, or decompression surgery: SRM = 1.42 (71).
- Rotator cuff disease: SRM = 1.16; instability: SRM = 0.93; and glenohumeral arthritis: SRM = 1.11 (63).

- Rotator cuff disease, isolated SLAP lesions, and shoulder instability after surgery: ES = 0.617, SMR = 0.771 (69).

Minimally important differences. An MDC of 9.4 points (90% CI 15.5) has been reported in one study in patients referred for physical therapy, including some after surgery (67).

The MCID of the ASES has been reported to range from 6.3 to 26.9 points, depending on the study population and method of ascertainment, as follows:

- Mixed population referred for physical therapy including some after surgery: 6.4 points (67).
- Tendinitis or rotator cuff tear treated with nonoperative modalities: 12.01 points for the ASES function subscale and 16.92 points for the pain subscale (78).
- Post arthroplasty (total, reverse, or hemi): 20.9 points (79).
- Post arthroplasty (total or reverse): 6.3 (95% CI 2.3-15.0) to 13.5 (95% CI 4.8-22.3) points, depending on the anchor (80); 6.5 points for ASES function subscale and 8 points for ASES pain subscale (81).
- Rotator cuff repair: 11.1 points (82).
- Full-thickness rotator cuff tears: 21.9 points (anchor-based) and 26.9 points (distribution-based) (83).

Generalizability. The patient self-reported component of the ASES has been used widely to assess outcomes from different surgical and nonoperative treatments in people with varying shoulder conditions, including glenohumeral arthritis, rotator cuff disease, shoulder instability, and adhesive capsulitis.

Use in clinical trials. The ASES was used in 43 randomized clinical trials up to the end of 2015 (2). It remains extremely popular and has been used in at least 35 trials in the last 5 years, including the following:

- Double- versus single-row rotator cuff repair (84)
- Steroid injection versus arthroscopic capsular release for early stage adhesive capsulitis (85)
- Rotator cuff repair with and without distal clavicle resection (86-88)
- Relaxation exercises to reduce postoperative pain after rotator cuff repair (89)
- Subacromial autologous–conditioned plasma versus glucocorticoid for symptomatic partial rotator cuff tears (90)
- Subacromial autologous platelet-rich plasma versus glucocorticoid for symptomatic partial rotator cuff tears (91)
- A 135 degree versus a 155 degree reverse arthroplasty prosthesis for rotator cuff arthropathy (92)
- Liposomal bupivacaine versus continuous peripheral nerve block following arthroplasty (93)
- Glucocorticoid injection versus oral nonsteroidal anti-inflammatory drugs (NSAIDs) for adhesive capsulitis (94)
- A 12-month exercise program versus usual care after rotator cuff repair (95)
- Multimodal analgesia injection combined with glucocorticoid versus saline injection after arthroscopic rotator cuff repair (96)
- Interference screw versus suture anchor fixation for biceps tenodesis (97)
- Optimum versus maximum tension–bridging suture for rotator cuff repair (98)
- Platelet-rich plasma after rotator cuff repair (99)
- Arthroscopic versus open stabilization for anterior shoulder subluxation (100)
- High- versus low-dose intra-articular glucocorticoid for shoulder stiffness (101)
- Triple-loaded single-row versus suture-bridging double-row (20) rotator cuff repair augmented with platelet-rich plasma fibrin membrane (102)
- Arthroscopic rotator cuff repair with and without biceps tenodesis, using the percutaneous intra-articular transtendon technique (103)
- Pulley exercises versus rehabilitation without pulleys after rotator cuff repair (104)
- Open versus arthroscopic rotator cuff repair (105)
- Biceps tenotomy versus biceps tenodesis (106)
- Intramedullary nail versus locking plate for proximal humeral fracture (107)
- Arthroscopy-assisted versus standard intramedullary nail fixation for diaphyseal humerus fractures (108)
- Reverse total shoulder arthroplasty as primary procedure versus revision procedure for proximal humerus fractures (109)
- Proprioceptive exercises in addition to conventional physical therapy for impingement (110)
- Intraoperative platelet-rich plasma versus local anesthetic injection after arthroscopic rotator cuff repair (111)
- Arthroscopic suprapectoral versus open subpectoral biceps tenodesis (112)
- Addition of mesenchymal stem cells to suture repair of ruptured supraspinatus muscle tears (113)
- Cement-augmented locking plate versus proximal humerus nail for surgical neck proximal humerus fractures (114)
- Arthroscopic soft tissue tenodesis at the rotator interval versus bony interference fixation tenodesis at the distal bicipital groove for the long head of the biceps (115)
- A 0.5 ml versus 1 ml type 1 atelocollagen intratendinous injection or no injection for small intratendinous partial thickness rotator cuff tear (116)
- Suture-spanning augmentation of single-row repair for massive rotator cuff tears (117)
- Biceps tenotomy versus tenodesis for long head of biceps lesions (118)
- Arthroscopic Bankart repair with and without arthroscopic in-

fraspinatus remplissage for anterior shoulder instability (119)
- Bankart repair with or without arthroscopic electrothermal capsulorrhaphy (120)

Critical appraisal of overall value to the rheumatology community

Strengths. The ASES score has good reliability, construct validity, and responsiveness and has been widely used.

Caveats and cautions. It uses a mix of scales (binary, Likert, and VAS). In the case that an MDC95% is reported to be higher than the MCID, the MDC95% should be taken as the MCID.

Clinical usability. Most studies include only the components of the patient self-assessment (pain and ADLs), which are used to derive an ASES score. This component is easy to understand, use, and score.

Research usability. The ASES score has good applicability for research and good responsiveness. It is recommended for use by the ASES Society, and it is widely used, particularly in surgical trials.

CONSTANT (MURLEY) SHOULDER SCALE

Description

Purpose. The CS was first described in a university thesis in 1986 and was published in 1987 (121). The score was originally conceived as an overall measure of the functional state of the shoulder in states of disease, injury, normal health, or after treatment to be used in both clinical assessment and research. The development paper does not describe the rationale or method for item selection and weighting. Modifications and guidelines for use were published by the original author in 2008 (122).

Self-administered versions of the CS have been proposed (in both English and French) (123,124), in which the two objectively assessed subscales (ROM and strength) are estimated by the patient using explanatory photographs (for ROM) and either measured or estimated using common household items of varying weight (for strength).

Content or domains. The CS has the following four domains: pain, ADLs, shoulder ROM, and strength. Pain and ADLs are both patient-reported; ROM and strength are measured by an examiner.

Number of items. There are 10 items in total (one for pain, four for ADLs, four for shoulder ROM, and one for strength).

Response options/scale. The pain subscale is a rating of the most severe pain experienced during normal activities over the preceding 24 hours and was originally a four-item Likert scale (none = 15 points, mild = 10 points, moderate = 5 points, and severe = 0 points) but was later modified to a 15-cm VAS (scored in centimeters and rounded to the nearest integer) (122,125). The VAS anchors are 15 = no pain and 0 = intolerable pain; there are no internal markers.

In the ADL subscale, sleep is rated on a 0 to 2 scale (0 = disturbance every night, 1 = occasional disturbance, and 2 = undisturbed). Work and recreational activities are each rated on a 15-cm VAS in response to the questions "how much of your normal work does your shoulder allow?" and "how much of your normal recreational activity does your shoulder allow?" with anchors of none and all. Each VAS response is measured in 3-cm quintiles to give a score of 0 to 4. Functional mobility is rated according to the level that the patient can use their hand comfortably on a five-item scale scored from 2 to 10 in two-point increments (up to waist = 2, up to xiphoid = 4, up to neck = 6, up to top of head = 8, and above head = 10). Some versions include a further response category (below waist), which is allocated zero points (125).

Shoulder ROM is measured by the examiner as the maximum painless active movement in four planes (forward elevation, lateral elevation, external rotation, and internal rotation). Each movement is allocated up to ten points for a total ROM score of up to 40 points. Forward and lateral elevation are each measured with a goniometer in 30-degree increments (0-30, 31-60, 61-90, 91-120, 121-150, and more than 150 degrees) and are allocated two points for each increment from 0 to 10 points. External rotation is measured according to the functional capacity to reach various anatomical landmarks, with two points allocated for each of the following five successful movements: hand to back of head with elbow forward, hand to back of head with elbow back, hand to top of head with elbow forward, hand to top of head with elbow back, and full elevation. Internal rotation is measured by the ability to reach the following anatomical landmarks with the thumb in two-point increments from 0 to 10 points: lateral aspect of thigh, behind the buttock, sacroiliac joint, level of the waist, 12th thoracic vertebra, and interscapular level.

Shoulder strength was originally described as the maximum resisted force with the arm at 90° of forward flexion and abduction in the coronal plane using a spring balance held in the patient's hand (or using a wrist cuff in patients with poor grip strength due to disease) (121). Various modifications to the original method have been made subsequently. The revised CS (122) used a proprietary isometric dynamometer (Isobex) attached via a wrist cuff to measure the maximum force generated in a 5-second effort with the shoulder at 90° abduction in the scapular plane and a pronated wrist. The score is the force in kilograms with one point per 0.5 kg to a maximum of 25 points.

Recall period for items. The recall period is 24 hours for pain and one week for ADLs.

Cost to use. Free of charge.

How to obtain. A PDF version that includes a standardized protocol for use can be found at https://ugeskriftet.dk/dmj/standardised-test-protocol-constant-score-evaluation-functionality-patients-shoulder-disorders.

Practical application

Method of administration. The CS is a patient-written self-assessment (pen and paper) plus clinical examination.

Scoring. The CS is a 100-point scale comprising patient-reported data (35 points) and clinical assessment (65 points). Pain contributes up to a maximum of 15 points, ADLs contribute up to 20 points (0-2 for sleep, 0-4 for work, 0-4 for recreation, and 0-10 for functional mobility), ROM contributes up to 40 points, and strength contributes up to 25 points (122). The total score is the sum of these four subscales.

Score interpretation. A score of 0 = worst function, and a score of 100 = best function. Individual subscales are not reported separately. The instrument by Constant and Murley was originally described as a composite measure of shoulder function and deliberately included pain assessment in the overall functional result (121).

Comparison with the contralateral shoulder is possible; however, this assumes an absence of pathology in the other shoulder. Normative data are available from a sample of 1620 clinic patients (1046 men and 573 women, age range 11-87 years) who reported one painless shoulder and a further 115 healthy volunteers (56 men and 59 women, age range 20-69 years) with no history of shoulder disorders or other illness (126). By definition, subjects were only included if they achieved a maximum score on the subjective measures (pain and ADLs). The mean total score (± SD) was 89 points (± 7) in clinic patients and 87 points (± 5) in healthy volunteers. Mean scores were higher in men (91-92 points) than women (83-84 points) ($P < 0.01$), with much of this difference being accounted for by differences in strength. Scores declined progressively with age. There was no difference in scoring between experienced physician researchers and novice resident physicians.

Respondent time to complete. The complete CS (including both patient-reported data and physical examination) takes less than 20 minutes (37). The self-reported component is relatively brief.

Administrative burden. The CS requires time to perform; however, some of the examination maneuvers (particularly ROM) may be incorporated in the normal physical examination. Some special equipment is needed, including a goniometer for measuring ROM and a dynamometer or similar device for measuring strength. No special software is required. The calculation of the score is straightforward.

Translations/adaptations. A German language version was published by Constant (127). Cross-cultural adaptations have been performed for several languages, including Danish (125,128), Turkish (129), Brazilian Portuguese (130), Greek (131), and Chinese (132).

Psychometric information

Floor and ceiling effects. No floor or ceiling effects for the total score were detected in patients with inflammatory or degenerative shoulder conditions referred for orthopedic surgery (37), in Dutch patients with subacromial impingement or rotator cuff tears (133), or in patients undergoing total shoulder arthroplasty for glenohumeral osteoarthritis (134). There is evidence of a floor effect for the strength subscale in patients who are unable to position their arm in the measurement position; 51.9% of patients with inflammatory or degenerative rheumatic diseases who were referred to a hospital for shoulder surgery received 0 points for strength because of either pain or inability to reach the measurement position (37).

Reliability. Internal consistency (measured by Cronbach's α) has been estimated to 0.37 in Korean patients undergoing surgery for various shoulder disorders (69), 0.60 in Canadian patients with mild to severe rotator cuff disease (135), and 0.8 in Dutch patients with subacromial impingement or rotator cuff tears (133).

Test-retest reliability is reported to be good to excellent. In French patients with painful rotator cuff disease, ICC = 0.92 (95% CI 0.82-0.98) for the total CS score (including self-reported and objective measures, using a single assessor) (136). In patients with rheumatoid arthritis, test-retest reliability for the total CS score was slightly higher when the same examiner was used (intratester ICC = 0.95-0.96) than when two different experienced clinicians performed the assessment (intertester ICC = 0.84-0.87) (137).

The strength testing component of the CS is a potential source of measurement error because of the lack of standardization of the testing procedure and equipment. In a sample of healthy volunteers, however, both intratester and intertester reliability was high when assessed by senior physical therapy students with experience in the use of this tool (intratester ICC = 0.90-0.98; intertester ICC = 0.89-0.97) (138).

In some cases in which reliability or validity was tested in countries where English is not the primary language, it was not clear whether the English CS or a translated version was used (44,69,133,136,139).

Validity. The CS has been linked to seven second-level International Classification of Functioning, Disability and Health (ICF) categories, which is considerably fewer than other measures, including the DASH and ASES, but similar to the SPADI (140). The seven categories represented the following two ICF components: body functions (sleep functions, sensation of pain, mobility of joint functions, and muscle power functions) and activities and participation (hand and arm use, remunerative employment, and recreation and leisure). There is variation in the methods used for strength testing, and no gold standard method exists. Variation in the position of the arm or the cuff during strength testing can result in substantial variation in force measurement (141).

Pearson's or Spearman's coefficients for correlation of the CS with other general and shoulder-specific outcome measures (and the population they were measured in) are as follows:

- Standardized Index of Shoulder Function: $r = 0.91$ in various nonsurgical and postsurgical shoulder disorders (142).
- Shoulder Function Index: $r = 0.89$ at baseline in proximal humeral fracture (143).
- OSS: $r = 0.71$ in degenerative or inflammatory shoulder disease referred for surgery (37); $r = 0.65$ in patients undergoing surgery for rotator cuff disease (144); and $r = 0.53$ (at 6 weeks) and 0.79 (at 12 weeks) in patients with a proximal humerus fracture (145).
- ASES: $r = 0.36$ in patients undergoing surgery for various shoulder disorders (69); $r = 0.87$ in patients undergoing surgery for rotator cuff tears (72); and $r = 0.495$ in patients enrolled in physical therapy for shoulder dysfunction (38).
- Penn Shoulder Score: $r = 0.85$ in patients undertaking physical therapy for shoulder disorders (54).
- SST: $r = 0.52$ in patients undergoing surgery for various shoulder disorders (69); $r = 0.49$ (72) to 0.70 (146) after rotator cuff surgery; and $r = 0.65$ in patients enrolled in physical therapy for shoulder dysfunction (38).
- UCLA Scale: $r = 0.70$ (at 6 weeks) and 0.92 (at 12 weeks) in patients with a proximal humerus fracture (145); $r = 0.67$ in patients undergoing surgery for various shoulder disorders (69); $r = 0.66$ after rotator cuff surgery (146); and $r = 0.59$ in patients enrolled in physical therapy for shoulder dysfunction (38).
- SPADI: $r = 0.53$ in degenerative or inflammatory shoulder disease referred for surgery (37) and $r = 0.56$ in patients enrolled in physical therapy for shoulder dysfunction (38).
- DASH: $r = 0.50$ in degenerative or inflammatory shoulder disease referred for surgery (37); $r = 0.76$ in patients undergoing surgery for rotator cuff tears (72); $r = 0.62$ (at 6 weeks) and 0.86 (at 12 weeks) in patients with a proximal humerus fracture (145); $r = 0.82$ in German-speaking Swiss patients following shoulder arthroplasty (28).
- WOSI Index: $r = 0.18$ in patients undergoing surgery for various shoulder disorders (69) and $r = 0.59$ in patients with symptomatic shoulder instability (147).
- WORC: $r = 0.56$ (148) to 0.65 (149) in rotator cuff disease.
- Bostrom's shoulder movement impairment scale: r = 0.78 in degenerative or inflammatory shoulder disease referred for surgery (37).
- Shoulder Function Assessment (SFA) scale: r = 0.86 in degenerative or inflammatory shoulder disease referred for surgery (37).
- SF-36 PCS score: $r = 0.41$ in patients undergoing surgery for various shoulder disorders (69).

Agreement between the modified CS and patient-reported CS. In patients referred for shoulder surgery, mean scores were similar for the clinician-assessed (mean 48, SD 20) and patient-assessed CS (mean 47, SD 19.5). The agreement between clinician assessors and patients was high for ROM (weighted $\kappa = 0.8$-0.9) (124). In French patients with various shoulder disorders, the correlation between the total CS score assessed by an orthopedic surgeon and the autoconstant assessed by the patient was ICC = 0.87 (123). Of the four subscales, the ICC was lowest for the strength subscale (0.57) but was highest for the ROM subscale (0.85).

Responsiveness. Reported ESs and SRMs in various shoulder conditions and settings are as follows:

- Patients with rotator cuff disease seen in a tertiary orthopedic clinic after 6 months: SRM = 1.38 (149).
- One week after completion of a course of physical therapy plus acupuncture or mock transcutaneous electrical nerve stimulation for painful rotator cuff disease: ES = 0.4 (150).
- After subacromial decompression surgery: SRM = 1.12 and 2.09 at 3 and 6 months and ES = 1.23 and 1.92 at 3 and 6 months (151).
- Six months following surgery for various shoulder disorders: SRM = 0.58 and ES = 0.57 (69).
- Two to five years after total shoulder arthroplasty for osteoarthritis: SRM = 2.4 and ES = 2.9 (134).
- Six months after total shoulder arthroplasty: SRM = 1.99 and ES = 2.23 (44).
- At 3 months in patients with symptomatic shoulder instability: SRM = 0.59 (147).

Minimally important differences. The MDC in Dutch patients with subacromial impingement or rotator cuff tears was 23 points (133). Reported MCIDs range between 8 and 17 points. It was 8 points for patients at 1 year following reverse shoulder arthroplasty for cuff arthropathy (152), 9.4 points at 44 months after shoulder arthroplasty for glenohumeral arthritis or cuff arthropathy (153), 10.4 points at 3 months after arthroscopic repair of rotator cuff tear (154), 11 points after 3 months of physical therapy

following arthroscopic decompression surgery in Danish patients (139), and 17 points after 3 months of exercise therapy for subacromial pain (155).

Generalizability. The CS has been tested in healthy subjects and in people with various shoulder pathologies, including rotator cuff disease, proximal humeral fractures, glenohumeral arthritis, shoulder instability, and adhesive capsulitis.

Use in clinical trials. The CS is one of the most widely used instruments in clinical trials for shoulder disorders, with almost 130 clinical trials reported up to December 2015 (2). Examples of trials that have used the CS as an outcome measure in the last 5 years include various surgical interventions for conditions including rotator cuff tears (156), proximal humeral fracture (157), and osteoarthritis (158); physical therapies (including heat therapy for shoulder pain) (159); and other interventions, including suprascapular nerve blocks (160) and platelet-rich plasma injections (161) for rotator cuff tears and glucocorticoid injections for adhesive capsulitis (162).

Critical appraisal of overall value to the rheumatology community

Strengths. The CS appears to be widely accepted within the clinical and research communities and is frequently used. It covers several clinically relevant domains. It is readily accessible and requires relatively little equipment. Normative data are available. Responsiveness appears to be acceptable in a variety of settings.

Caveats and cautions. The CS requires assessment in person by a trained assessor; self-assessment versions have been developed but require further validation. Ambiguity in the description of the assessment tasks and inconsistent application of testing methods have the potential to result in excessive interrater variability. There are limited data on the MDC, but existing data create some concern that the MDC may be higher than the MCID. There is evidence of variation in the methods and equipment used to measure strength, which may affect overall scores, and floor effects for strength exist when shoulder mobility is severely restricted.

Clinical usability. The CS is commonly used in clinical practice, particularly in surgical specialties, and in Europe. It is easily incorporated into routine clinical assessment, although some special equipment is required for strength testing.

Research usability. Concerns regarding reliability (particularly intertester reliability) are an important caveat for use in a research setting.

SIMPLE SHOULDER TEST

Description

Purpose. The SST assesses functional limitations of the shoulder relative to the patient's ADLs before or after treatment and work (163). Questions were derived from Neer's evaluation (164), the ASES evaluation (52), and the most frequent complaints of patients observed in the shoulder practice at the University of Washington (163). Further details on how item content was selected have not been described. Item response theory was applied later (165).

Content or domains. The SST has the following three domains: pain, function/strength, and ROM.

Number of items. The SST contains 12 items (two for pain, seven for function/strength, and three for ROM).

Response options/scale. All items have yes/no responses.

Recall period for items. At the moment of assessment.

Cost to use. Free of charge for noncommercial use.

How to obtain. The SST can be obtained from the original publication (163) or is free online at https://orthop.washington.edu/patient-care/articles/shoulder/simple-shoulder-test.html.

Practical application

Method of administration. Self-assessment.

Scoring. For the original score, 0 = worst and 12 = best. It is transformed by summing the number of "yes" answers/number of completed items and multiplying this by 100 to get the percentage of "yes" responses.

Score interpretation. The score ranges from 0 (worst function) to 100 (best function). A missing rule, distinct cutoffs for severity, and normative data have not been published. There are no subscales.

Respondent time to complete. It takes 2 to 3 minutes to complete.

Administrative burden. The SST is free online. The score computation is very easy and possible by hand. No software is needed. The time to administer and determine the score is estimated to be 5 minutes.

Translations/adaptations. The SST has been translated, adapted, and validated in Dutch (166), Brazilian Portuguese (167), Spanish (168), Persian (169), Italian (170), and Lithuanian (171).

Psychometric information

Floor and ceiling effects. A range of floor and ceiling effects have been found across a range of shoulder conditions (36,64,166,169,172,173). Floor and ceiling effects were found in 1.2% and 5.1% of patients with rotator cuff diseases and in 2% and 9.3% of patients with shoulder instability in one study (172); floor and ceiling effects were found in 21% and 6.1% of patients in a study of people with rotator cuff disease (172); and ceiling effects were found in 15.3% of patients who had undergone shoulder arthroplasty, but no floor effects were found (173). No floor or ceiling effects were found in Iranian patients with a variety of shoulder disorders (169), whereas 1.8% floor and 13.6% ceiling effects were observed among Dutch patients with various shoulder disorders (166).

Reliability. Estimates of the internal consistency of the SST have been reported across various patient populations, including patients with various shoulder complaints: (Cronbach's α = 0.85) (29), Brazilian patients with various shoulder complaints: (Cronbach's α = 0.82, 95% CI 0.76-0.86) (167), Dutch patients with shoulder problems attending an orthopedic clinic (Cronbach's α = 0.78; removing items from the questionnaire did not result in higher Cronbach's α) (166), Italian patients after surgery for anterior instability (Cronbach's α = 0.87) (170), and Iranian patients with a variety of shoulder disorders (Cronbach's α = 0.73) (169).

One study reported that the SST was unidimensional, with 8.4% unexplained variance (64). However, although the STT was designed to measure a single construct, some factor analyses have questioned this and have more commonly identified two- or three-factor solutions. One factor analysis of the English version of the SST from a sample of patients with shoulder complaints revealed a two-factor solution that explained 52.6% of the variance (29). Two items relating to the ability to sleep comfortably and comfort at rest, which are both influenced by the amount of pain a person is experiencing, showed misfit. In another Rasch analysis, three items were considered misfit in data obtained from patients following shoulder arthroplasty or rotator cuff repair surgery (174). However, this study found that the SST was unidimensional and that this was not the reason for the misfit of items. Across the entire continuum of shoulder functioning, function was not measured with equal precision but with very large confidence intervals, ie, larger than those of the ASES and SPADI (165).

Moderate fit with one factor was identified in the Dutch version (166); however, three-factor solutions have been found in the Persian (169), Brazilian Portugese (167), and Spanish (168) versions. A three-factor solution, mainly related to activities performed with the arm at shoulder level (arm elevation), shoulder movement, and arm strength that explained 49.7% of variance was identified in the Persian STT (169). The Cronbach's α for these three factors was 0.7, 0.53, and 0.6, respectively. For the Brazilian Portuguese STT tested in patients with various shoulder problems, a three-factor solution explained over 40% of the variance, and Cronbach's α was 0.82 for the overall test, 0.82 for both arm elevation and shoulder movement subscales, 0.81 for comfort in rest subscale, and 0.59 for the overall global shoulder function value (167). For the Spanish version, also tested in patients with various shoulder problems, a three-factor structure explained 56.12% of the variance, with Cronbach's α of 0.73, 0.72, and 0.66 for the three factors, respectively, and 0.793 overall (168).

The test-retest reliability of the SST is high. One study reported ICCs of 1 (95% CI 1-1) and 0.97 (95% CI 0.91-0.99) measured in patients with instability and rotator cuff disease, respectively (172). In a study of patients attending a surgical clinic, the ICC = 0.99 (31); in studies of Dutch patients with shoulder complaints, the ICC = 0.86 (175) and 0.92 (166); in Italian patients after surgery for instability, the ICC = 0.97 (170); and in patients with various shoulder problems in Iran, the ICC = 0.94 (95% CI 0.86-0.97 (169) and in Spain, the ICC = 0.912 (168).

Validity. Pearson's or Spearman's coefficients for correlation of the STT with other general and shoulder-specific outcome measures (and the population they were measured in) are as follows:

- SPADI: ICC = 0.74 and 0.80 (29,36) in various shoulder problems; and ICC = 0.68, pain subscale ICC = 0.63, and function subscale ICC = 0.72 Italian patients after shoulder surgery for anterior instability (170).
- ASES: 0.73 (36) and 0.536 (64) in various shoulder problems; 0.76 in rotator cuff diseases (172); 0.89 in instability (172); and ICC = 0.8 in Italian patients after shoulder surgery for anterior instability (170).
- OSS: ICC = 0.61 (176) and ICC = 0.74 (166) for general shoulder problems in Dutch patients and ICC = 0.68 (169) in Iranian patients; ICC = 0.67 in Italian patients after shoulder surgery for anterior instability (170).
- DASH: 0.72 in rotator cuff surgery (42) and ICC = 0.73 for various shoulder problems in Spanish patients (168) and ICC = 0.74 in Dutch patients (166).
- CS: 0.70 in rotator cuff surgery (146).
- WORC: 0.68 in rotator cuff surgery (42).
- SF-36 bodily pain: 0.62 in various shoulder problems (36).
- UCLA Scale: ICC = 0.61 for the pain subscale and ICC = 0.51 for the function subscale in Italian patients after shoulder surgery for anterior instability (170).
- WOSI: ICC = 0.63 in Italian patients after shoulder surgery for anterior instability (170).
- PROMIS PF CAT: 0.635 in patients seen by a shoulder and elbow surgeon (64).

- SF-36 physical functioning: 0.58 (36) and 0.56 in various shoulder problems (166).
- SF-36 PCS: 0.40 in various shoulder problems and 0.60 in rotator cuff surgery 0.60 (36,42).
- SF-12 PCS: 0.44 in various shoulder problems (172) and 0.43 in Spanish patients (168).
- SF-36 mental component summary (MCS) score: 0.16 in rotator cuff surgery (42).

Responsiveness. ESs and SRMs of the SST are as follows:

- Shoulder arthroplasty: ES = 2.17-2.87, SRM = 1.43-2.05 (173,177).
- Rotator cuff repair: SRM = 1.09 (42).
- Rotator cuff surgery: ES = 1.08, SRM = 1.01 (172).
- Rotator cuff surgery and total shoulder arthroplasty: SRM = 0.87 (31).
- Instability surgery: ES = 0.61, SRM = 0.63 (172).
- After shoulder arthroplasty: ES = 2.29, SRM = 2.05 (173).

Minimally important differences. The MDC95% across various shoulder problems is 32.3% (0%-100% scale) (29), and after surgery for shoulder instability in Italian patients it is 1.12 (0-100% range) (170) and 2.8 points (0-12 scale) (175). In Spanish patients with various shoulder problems, MDC90 = 6.2% (0%-100% scale) (168); in Iranian patients with shoulder disorders, MDC = 3.7 (0-12 scale) (169); and in Dutch patients with various shoulder problems, MDC = 3.3 (0-12 scale (166).

The MCID is reported for the following shoulder conditions:

- Rotator cuff disease: 2.05 (using the 15-item function question as an anchor question) and 2.33 points (using the four-item improvement question as an anchor question) (0-12 range) (78).
- Shoulder arthroplasty 3 and 2.4 points (0-12 range) (79,177).
- Various shoulder problems: 2.2 points (0-12 range) or 17.1 to 25.0 (0-100 range) (175).

Generalizability. This measure is shoulder specific but has also been used to measure shoulder function in patients with other conditions, eg, following surgery for breast cancer (30). It has also been translated and culturally adapted in various countries.

Use in clinical trials. Up to December 2015, the SST had been used in 17 randomized controlled trials (2). Examples of trials that have used the SST to measure outcomes in the last 5 years include double-row versus single-row rotator cuff repair (84), arthroscopic versus open stabilization for anterior shoulder subluxations (100), multimodal analgesia injection combined with corticosteroids after arthroscopic rotator cuff repair (96), pulley exercises versus rehabilitation without pulleys after rotator cuff repair (104), precut kinesiology tape versus oral NSAIDs for impingement syndrome (178), arthroscopic release versus manipulation under anesthesia for frozen shoulder (179), interference screw versus suture anchor fixation for biceps tenodesis (97), arthroscopic Bankart repair with and without arthroscopic infraspinatus remplissage for anterior shoulder instability (119), tenotomy versus tenodesis for long head of biceps tendon lesions (180), and arthroscopic rotator cuff with or without acromioplasty (87).

Critical appraisal of overall value to the rheumatology community

Strengths. The SST is very short and easy to use. It has good construct validity with function subscales, good reliability, and acceptable floor and ceiling effects.

Caveats and cautions. Because of the binary response options, the use of the SST score as a metric measure is questionable. In the case that an MDC95% is reported to be higher than the MCID, the MDC95% should be taken as the MCID. However, more studies are needed to compare the MDC95% with the MCID within similar populations and contexts. Construct validity is better when correlating the SST with other shoulder-specific measures rather than generic measures. The SST can measure shoulder function but cannot infer overall health of the shoulder.

Clinical usability. The SST is easy to use; it has widespread use in the United States and is also translated and culturally adapted in many different countries.

Research usability. This measure is widely used because of its convenience and ease of use. However, we recommend it be used with caution because of its identified caveats. Further testing is needed to fully understand its psychometric properties for various patient groups.

OXFORD SHOULDER SCORE

Description

Purpose. The OSS was developed to be a self-assessment of pain and function of the shoulder, assessing outcomes from shoulder surgeries other than shoulder stabilization (181). The original version was published in 1996, and a revision published in 2009 clarified specifications for its use but only altered the method of scoring each item (182). A 22-item questionnaire was developed based on interviews with patients attending an outpatient shoulder clinic as well as established questionnaires (181). Another set of patients completed the draft questionnaire and were also invited to provide comments, including identifying

any further shoulder problems that had been omitted. Further testing over several steps resulted in the 12-item version. Factor analysis or item response theory was not used.

Content or domains. The OSS contains the following two domains: pain and daily functions. There are no subscales.

Number of items. The OSS has 12 items (two for pain, two for interference with pain, and eight for daily functions).

Response options/scale. In both the original and revised versions, there are five response options for each item. The original version: 1 = no pain/easy to do, 2 = mild pain/little difficulty, 3 = moderate pain/moderate difficulty, 4 = severe pain/extreme difficulty, and 5 = unbearable/impossible to do (range is 1 [best] to 5 [worst]) (181).

In the revised version, each of the 12 items are scored from 0 to 4 instead of 1 to 5, and the direction of the scores were also reversed (0 [worst] to 4 [best]) (182).

Recall period for items. The recall period is 4 weeks.

Cost to use. Free of charge for noncommercial use; fees for commercial users and academic studies that are funded by a commercial entity.

How to obtain. It is available at https://innovation.ox.ac.uk/outcome-measures/oxford-shoulder-score-oss/.

Practical application

Method of administration. Self-assessment.

Scoring. In the original OSS, the total score is the sum of the (completed) 12 items (scoring 1-5); 12 = best and 60 = worst (181). In the revised version, the total score is also the sum of the 12 items (scoring 0-4), but the scores are reversed, with score ranges from 0 = worst to 48 = best (182). There are no subscales.

At least ten of the 12 items have to be completed (182). If only one or two questions have been left unanswered, the mean value representing all of the other responses can be entered to fill the gaps. If more than two questions are unanswered, calculating an overall score is not recommended. If patients indicate two answers for one question, by convention it is recommended that the worse (more severe) response is adopted.

Score interpretation. In the original OSS, a score of 12 indicates no disability, whereas a score of 60 indicates maximal disability. In the revised OSS (and online form), a score of 0 indicates maximal disability, whereas a score of 48 indicates no disability. A score of 0 to 19 indicates severe arthritis, 20 to 29 indicates moderate to severe arthritis, 30 to 39 indicates mild to moderate arthritis, and 40 to 48 indicates satisfactory joint function.

Normative data for the OSS scale (range 0-48) in 635 asymptomatic, healthy volunteers in Australia (n = 323) and Canada (n = 312) are available (9). Participants were excluded if they had a history of active shoulder pathology or a history of recent surgery (within the last 3 years) or joint arthroplasty. People without a history of shoulder pathology reported higher (better) mean OSS scores compared with those with a history of a shoulder problem (46.8 versus 46.1, Wilcoxon's rank sum test; $P = 0.0013$). No differences in scores were observed in people with and without a current wrist or elbow problem or handedness. There was no difference between sexes adjusted for nationality (mean of 47 in both). The mean score declines with increased age.

Respondent time to complete. The OSS takes 2 to 4 minutes to complete (40,183-186).

Administrative burden. Score computation is easy and needs no explanation. No training is needed to administer and interpret the scores. It can be completed using pen and paper. The time to administer and score is approximately 5 minutes.

Translations/adaptations. The OSS has been translated and culturally adapted in Dutch (176), Italian (170,185), German (187), Brazilian Portuguese (188), Korean (183), French (189), Turkish (40), Chinese (190), Persian (169), Spanish (30), Romanian (184), Danish (191), and Polish (192).

Psychometric information

Floor and ceiling effects. No floor or ceiling effects have been reported in a mixed population of patients (most with rheumatoid arthritis or osteoarthritis) undergoing glenohumeral or acromioclavicular surgery (37) or in patients with mixed shoulder complaints in Iran (169), China (190), or Turkey (40). In one study in Korean patients with degenerative or inflammatory shoulder disorders, floor and ceiling effects of 0.1% and 8.5%, respectively, were reported (183).

Reliability. The OSS has acceptable internal consistency with numerous studies in varying populations, reporting Cronbach's α of 0.89 or more in both the original and translated versions as follows:

- Cronbach's $\alpha = 0.89$ in a preoperative mixed population that excluded patients with instability; Cronbach's $\alpha = 0.89$ in a subset of the same population 6 months following various types of surgery; all items correlated with the total score ($r >$ 0. 4) (181).
- Cronbach's $\alpha = 0.94$ in impingement syndrome with or without rotator cuff tear or calcific tendinitis (187).

- Cronbach's α = 0.843 in a mixed population of patients (most with rheumatoid arthritis or osteoarthritis) undergoing glenohumeral or acromioclavicular surgery (37).
- Cronbach's α = 0.947 in Spanish patients with shoulder pain/dysfunction after surgery for breast cancer (30).
- Cronbach's α = 0.91 for current shoulder symptoms in Persian patients; if an item was deleted, it ranged from 0.91 to 0.92 (169).
- Cronbach's α = 0.94 for impingement or calcific tendinitis in German patients, and it was not below 0.93 when single items were eliminated (187).
- Cronbach's α = 0.95 for degenerative or inflammatory shoulder conditions in Italian patients (185).
- Cronbach's α = 0.92 for shoulder problems in Turkish patients (40), Cronbach's α = 0.93 for in French patients (189), and Cronbach's α = 0.93 in Chinese patients (190).
- Cronbach's α = 0.96 post rotator cuff surgery in Polish patients (192).
- Cronbach's α = 0.91 for degenerative or inflammatory shoulder conditions in Korean patients (183) and Cronbach's α = 0.90 in Brazilian patients (186).
- Cronbach's α = 0.957 in Brazilian patients with rheumatoid arthritis (188).
- Cronbach's α = 0.954 in Romanian patients with rotator cuff disorders or proximal humerus fractures (184).

Factor analysis of the Persian OSS found a two-factor solution that explained 61.6% of the variance, which is consistent with the OSS being a two-dimensional outcome measure assessing pain and function (169). Principal component analysis of the Chinese OSS found a one-factor structure that accounted for 54.2% of the total variance (190).

Test-retest reliability is also reported to be acceptable, with values as follows:

- Preoperative mixed population excluded patients with instability: coefficient of reliability = 6.8 (Bland and Altman method) (181).
- Impingement with or without rotator cuff tear or calcific tendinitis: Pearson's correlation = 0.98 (187).
- Impingement: weighted κ = 0.13 for item 1; weighted κ values for other items = 0.44 to 0.79 (39).
- Patients with shoulder pain/dysfunction after surgery for breast cancer (Spanish OSS): ICC = 0.974 (30).
- Patients with current shoulder symptoms (Persian OSS): ICC = 0.90 (95% CI 0.77-0.95) (169).
- Patients with shoulder symptoms (Danish OSS): ICC = 0.98 (191).
- Patients with impingement or calcific tendinitis (German OSS): ICC = 0.98 (187).
- Patients with general shoulder problems (Danish and French versions of the OSS): ICC = 0.981 (176) and 0.91 (95% CI 0.88-0.94) (189), respectively.
- Post rotator cuff surgery (Polish OSS): ICC = 0.99 (192).
- Degenerative or inflammatory shoulder conditions (Korean OSS): ICC = 0.95 (183).
- Patients with nonspecific shoulder pain (Chinese OSS): ICC = 0.97 (190).
- Romanian patients with rotator cuff disorders and proximal humerus fractures: ICC = 0.953 and ICC = 0.976 (average ICCs of the office visit and 2 days from the office visit, respectively) (184).
- Turkish patients with mixed shoulder complaints: ICC = 0.99 (40).
- Brazilian patients with degenerative/inflammatory shoulder conditions or rheumatoid arthritis: ICC = 0.92 (186) and ICC = 0.917 (188), respectively.

Validity. Pearson's or Spearman's coefficients for the correlation of the OSS with other general and shoulder-specific outcome measures (and the population they were measured in) are as follows:

- CS: –0.74 and –0.75 in a mixed population of shoulder conditions before and after surgery, respectively (181); 0.66 in impingement with or without rotator cuff tear or calcific tendinitis (187); 0.71 in a mixed population of patients (most with rheumatoid arthritis or osteoarthritis) undergoing glenohumeral or acromioclavicular surgery (37); 0.84, 0.77, and 0.87 at baseline, 3 month follow-up, and 12 month follow-up, respectively, in proximal humeral fracture treated conservatively (193); 0.60 in impingement or calcific tendinitis in German patients (187); ICC = 0.64 in general shoulder problems in Dutch patients (176), 0.73 in French patients (189), and 0.66 in Chinese patients (190); ICC = 0.73 in degenerative or inflammatory shoulder conditions in Italian patients (185); ICC = 0.74 in Danish patients with shoulder complaints (191); ICC = 0.65 in rotator cuff disease both prior to surgery and 6 months post surgery (144); and 0.731 in Brazilian patients with rheumatoid arthritis (188). The correlation with the ADL, strength, and ROM subscales of the CS in Korean patients with degenerative or inflammatory shoulder conditions was ICC = 0.68, 0.48, and 0.42, respectively (183).
- SPADI: 0.74 in a mixed population of patients (most with rheumatoid arthritis or osteoarthritis) undergoing glenohumeral or acromioclavicular surgery (37); 0.85 in subacromial impingement (39); –0.674 in Spanish patients with shoulder pain/dysfunction after surgery for breast cancer (30): ICC = –0.74, 0.71, and 0.72 (total, pain subscale, and disability subscale, respectively) in mixed shoulder complaints in Turkish patients (40); r = 0.85 in patients attending a specialist shoulder clinic with subacromial impingement (39); and ICC = 0.79 in Italian patients after surgery for anterior instability (170).

- DASH: ICC = 0.79 in a mixed population of patients (most with rheumatoid arthritis or osteoarthritis) undergoing glenohumeral or acromioclavicular surgery (37); ICC = 0.77 in patients with degenerative or inflammatory shoulder conditions in Brazil (DASH work: ICC = 0.76; DASH sport/music: ICC = 0.62) (186); and ICC = 0.81, 0.69, and 0.59 for DASH disability/symptoms, DASH work, and DASH sport/music, respectively, in Korea (183).
- QuickDASH: –0.92 post rotator cuff surgery in Polish patients (192) and ICC = 0.633 and 0.672 (2 days later) in rotator cuff disorders and proximal humerus fractures in Romanian patients (184).
- SST: 0.68 in current shoulder symptoms in Persian patients (169); ICC = 0.61 (176) and 0.74 (166) in general shoulder problems in Dutch patients; and 0.67 in Italian patients after surgery for anterior instability (170).
- SFA scale: ICC = 0.72 in a mixed population of patients (most with rheumatoid arthritis or osteoarthritis) undergoing glenohumeral or acromioclavicular surgery (37).
- Subjective Shoulder Value (SSV): ICC = 0.68 in general shoulder problems in French patients (189).
- UCLA Scale: ICC = 0.67 in degenerative or inflammatory shoulder conditions in Italian patients (185).
- SF-36 bodily pain: ICC = –0.66 and –0.68 in a mixed population of shoulder conditions before and after surgery, respectively (181); ICC = 0.69 in subacromial impingement (39); ICC = 0.76 in impingement with or without rotator cuff tear or calcific tendinitis (187); 0.81 post rotator cuff surgery in Polish patients (192); 0.64 and 0.75 in rotator cuff disease prior to surgery and 6 months post surgery, respectively (144); ICC = 0.53 in nonspecific shoulder pain in Chinese patients (190); ICC = 0.66 in degenerative or inflammatory shoulder conditions in Italian patients (185); ICC = 0.74 in mixed shoulder complaints in Turkish patients (40); 0.624 in Brazilian patients with rheumatoid arthritis (188); and 0.7 (pain subscale) and 0.54 (function subscale) in Italian patients after surgery for anterior instability (170).
- SF-36 physical function: –0.61 and –0.62 in a mixed population of shoulder conditions before and after surgery, respectively (181); 0.62 in impingement with or without rotator cuff tear or calcific tendinitis (187); 0.57 in subacromial impingement (39); 0.82 post rotator cuff surgery in Polish patients: (192); ICC = 0.65 in nonspecific shoulder pain in Chinese patients: (190); –0.57 and 0.68 in rotator cuff disease prior to surgery and 6 months post surgery (144); ICC = 0.63 in mixed shoulder complaints in Turkish patients (40); ICC = 0.74 in degenerative or inflammatory shoulder conditions in Italian patients (185); and ICC = 0.589 in Brazilian patients with rheumatoid arthritis (188).
- SF-36 PCS: 0.63 in mixed shoulder complaints in Turkish patients (40) and ICC = 0.82 post rotator cuff surgery in Polish patients (192).
- HAQ: pain subscale 0.49 and 0.71 and function subscale 0.86 and 0.80 in mixed population of shoulder conditions before and after surgery, respectively (181) and 0.663 in Brazilian patients with rheumatoid arthritis (188).

One study assessed correlation of the OSS with other shoulder measures in people without current shoulder complaints and found the following: r = 0.71 for the ASES, r = 0.44 for the CS, r = 0.8 for the UCLA Scale, r = 0.78 for the SPADI, and r = 0.53 for the SPONSA (9).

Responsiveness. ESs and SRMs of the OSS have reported across various populations and settings as follows:

- Mixed population of patients with shoulder conditions planning to undergo surgery: ES = 1.2 (181).
- Mixed population of patients undergoing hemiarthroplasty: ES = 2.9 in osteoarthritis with intact rotator cuff, ES = 2.1 in osteoarthritis with a torn rotator cuff, and ES = 3.1 in rheumatoid arthritis (194).
- Mixed population of patients undergoing subacromial decompression with or without rotator cuff repair: ES = 1.88 for full rotator cuff repair, ES = 0.14 for partial repair, ES = 1.02 for no repair, and ES =1.55 for no tear (195).
- Subacromial impingement (regardless of treatment): ES = 0.24 and 0.96 at baseline to 6 weeks and to 18 weeks respectively (39).
- Mixed population of patients (most with rheumatoid arthritis or osteoarthritis) undergoing glenohumeral or acromioclavicular surgery: ES = 0.61 (37).
- Spanish patients with shoulder pain/dysfunction after surgery for breast cancer: ES = 0.50 and SRM = 0.70 (30).

Minimally important differences. An MDC of 3.15 points was reported in a study of Polish patients post rotator cuff surgery (192), whereas an MDC of 7.18 was reported in a study of Brazilian patients with degenerative or inflammatory shoulder conditions (186). Both a smallest detectable change (SDC) and an MCID of 6 points (OSS scale 0-48 points) were reported in a study including patients with shoulder problems attending an orthopedic outpatient clinic (baseline and 6 months follow-up, both surgery and nonoperative treatment, excluding fractures and frozen shoulder) (175). Another study that included patients following subacromial decompression also reported an MCID of 6 points (139). An MCID of 6.9 points for the original 12 to 60 point scale was reported in a study including a mixed population of patients (most with rheumatoid arthritis or osteoarthritis) undergoing glenohumeral or acromioclavicular surgery (196). An MCID of 18.8 on a 0 to 100 scale was reported in Persian patients with general shoulder problems (169).

Generalizability. Since the previous review there have been many studies evaluating the psychometric properties of the OSS. It has been widely translated and adapted into different languages and cultural settings, and it has been used widely to assess outcomes from different surgical and nonoperative treatments in people with varying shoulder conditions, including glenohumeral or acromioclavicular arthritis, rotator cuff disease, proximal humerus fractures, and adhesive capsulitis.

Use in clinical trials. Up to the end of 2015, the OSS had been used in seven randomized controlled trials (2). The OSS continues to be used as an outcome in trials published in the last 5 years, including trials for the following:

- Surgical versus nonsurgical treatment for proximal humerus fractures (197)
- Intraoperative platelet-rich plasma versus local anesthetic injection for arthroscopic repair (111)
- Ultrasound added to exercise and mobilization for frozen shoulder (198)
- Supervised exercise versus home exercise for frozen shoulder (199)
- Cement-augmented locking plate versus proximal humerus nail for surgical neck proximal humerus fractures (114)
- Arthroscopic subacromial decompression versus placebo for subacromial shoulder pain (200)
- Arthroscopic suprapectoral versus open subpectoral biceps tenodesis (112)
- Ultrasound-guided needling and lavage with or without glucocorticoid injection versus placebo for calcific tendinitis (201).
- Addition of platelet-rich plasma applied to tendon repair site after double-row arthroscopic supraspinatus repair (202)
- Extracorporeal shock wave therapy with or without kinesiotaping for calcific tendinitis (203)
- Arthroscopic implantation of a subacromial balloon spacer for treating massive rotator cuff tear (204)
- Stemmed versus stemless total shoulder arthroplasty for glenohumeral osteoarthritis (205)
- Reversed total shoulder arthroplasty versus nonoperative treatment for displaced proximal humerus fracture (206)

Critical appraisal of overall value to the rheumatology community

Strengths. This is a very short and responsive tool with good internal consistency. It is easy to complete and score. Guidelines for interpreting the scores are clear, and minimal floor or ceiling effects have been observed. This is a tool that was specially designed for surgical interventions. The construct validity with other measures is good, and there is no cost for noncommercial purposes. This tool has been translated and culturally adapted in many countries.

Caveats and cautions. Because the MDC is similar to the MCID, caution is necessary for measurements at an individual patient level. Similar to some other shoulder-specific tools, the OSS is a two-dimensional outcome measure but is used as a single scale without subscales.

Clinical usability. It is a short and easy-to-use tool for the assessment of shoulder surgery. Scores are easy to interpret.

Research usability. The OSS continues to be widely used in clinical trials investigating both nonoperative and surgical interventions for people with shoulder conditions.

DISABILITIES OF THE ARM, SHOULDER, AND HAND QUESTIONNAIRE AND ITS SHORT VERSION

Description

Purpose. The DASH was developed for self-assessment of symptoms and function of the entire upper extremities in patients with pain in the arm, shoulder, or hand (207). Draft items were identified by literature review (821 items were reduced to 67 plus 3 new) because of content overlap or off-target questions determined by a group consensus). Patient data were analyzed by different item to total correlation techniques, comparison to clinimetric ranking and clinical judgment, resulting in the final 30-item version (207,208). The newest manual contains extensive psychometric information (207). Psychometric analysis by item response theory (using Rasch analysis) was performed later for the DASH (209,210).

A shortened version called the QuickDASH (11 items) was published in 2005 (211), and the QuickDASH-9 (nine items) was published in 2009 (212). All relevant modern strategies were used in the development of the QuickDASH, comparing the following strategies: the concept retention method, the equidiscriminative item-total correlation, and the item response theory (Rasch modelling). The concept retention method was most similar to the DASH and was chosen to build the QuickDASH (211).

Content or domains. There are two domains (symptoms and function).

Number of items. The 30 items in the DASH are comprised of six items about symptoms (three for pain, one for tingling/numbness, one for weakness, and one for stiffness) and 24 about function (21 for physical function and three for social/role function). Two optional additional modules for work (four items) and sports/performing arts (four items) are more rarely used in patient settings but are used when assessing manual workers and athletes. The QuickDASH has 11 items (three for symptoms and

eight for function) (211). The QuickDASH-9 has nine items (one for pain and eight for function) (212) but is rarely used and not supported by the authors of the original questionnaire (207).

Response options/scale. All items are scored on a scale of five (Likert) levels in which 1 = no difficulty/symptoms, 2 = mild difficulty/symptoms, 3 = moderate difficulty/symptoms, 4 = severe difficulty/symptoms, and 5 = extreme difficulty (unable to do)/symptoms.

Recall period for items. The recall period is 1 week.

Cost to use. The DASH is free of charge for non-commercial use; licenses for commercial use are available at the Institute for Work and Health (IWH). Details on commercial use licences and costs can be found online (http://www.dash.iwh.on.ca/licences). The manual (3rd edition) (online and paper copies) can be found at http://www.dash.iwh.on.ca/dash-manual.

How to obtain. The property and copyright can be obtained at the IWH (online at http://www.dash.iwh.on.ca/). The links for the DASH and QuickDASH can be downloaded from the IWH website (http://www.dash.iwh.on.ca/about-dash and http://www.dash.iwh.on.ca/about-quickdash). Language versions are online at http://www.dash.iwh.on.ca/available-translations.

Practical application

Method of administration. The DASH is a self-assessment completed with paper and pencil or a computer.

Scoring. The arithmetic mean of at least 27 of the 30 items (missing rule) is transformed by (mean – 1) × 25 into the scale from 0 = no symptoms/full function to 100 = maximal symptoms/no function for the DASH total score (207). Five of six items are necessary for the determination of the symptoms score, and 22 of 24 items are needed for the function score. Similarly, ten of 11 items are necessary for the QuickDASH total score, three of three are needed for symptoms, and seven of eight are needed for function (211). Determination of the subscores for symptoms and function is possible (28,209,210,213,214), but this is not originally described (207).

Computer scoring is not necessary but is easier, eg, on Microsoft Excel or any calculation or statistics program. Scoring programs are online at http://www.orthopaedicscore.com/scorepages/disabilities_of_arm_shoulder_hand_score_dash.html and http://www.orthopaedicscore.com/scorepages/disabilities_of_arm_shoulder_hand_score_quickdash.html.

Score interpretation. Originally, 0 = best and 100 = worst. The reverse scale from 0 = worst to 100 = best (100 – original score) is also often used for comparison with other scores such as the SF-36. Several studies have showed varying distinct cutoff points to reflect severity. One example of this is cutoff scores of less than 15 = no problem, scores from 16 to 40 = problem, but working, and scores of more than 40 = unable to work (207).

Normative values of 1706 persons in the US general population stratified by sex, age, and comorbidity are available (US population mean = 10.1 [SD 14.7]) (28,207,215).

Respondent time to complete. The time to complete is 4 minutes for the DASH and 2 minutes for the QuickDASH (28,207,211,214). All items are easy to read and comprehend and are not emotionally sensitive, with the exception of item 21, which asks about sexual activity. This item is often left out by patients. For that reason, item 21 was not included in the QuickDASH (28,211).

Administrative burden. Item rating can be typed or scanned into an electronic database. Score computation is easy (see above) and missing data are rare (less than 10%) (209) other than item 21 on sexual activity, which has a high level of missingness (up to 30%) (211,216,217).The questionnaire contains instructions on how to complete it. The time to administer (including control of missing data) is 10 minutes for the DASH and 8 minutes for the QuickDASH. Little special training is necessary for these activities.

Translations/adaptations. The DASH is available for free for more than 50 languages and dialects (http://www.dash.iwh.on.ca/available-translations). Translations and cross-cultural adaptations have followed strict guidelines. (http://www.dash.iwh.on.ca/translation-guidelines). Versions in 20 other languages are in progress (as of January 6, 2020).

Psychometric information

Floor and ceiling effects. Very low floor or ceiling effects have been reported. In most studies, no floor or ceiling effects were reported, although two studies showed 0.5% and 1.8% ceiling effects for the DASH and 3% for the QuickDASH, and one study reported 2% floor effects (28,209,210,216,218-220).

Reliability. The DASH has high internal consistency, with Cronbach's α = 0.92 to 0.98 for the DASH (207,209,210,212,221) and Cronbach's α = 0.92 to 0.95 for the QuickDASH (211,222). High reported Cronbach's α may point toward some redundancy in the DASH, and Rasch analyses have shown some problems regarding the unidimensionality of the DASH score, which contains groups of items that measure different constructs, such as impairment activity limitations and participation restriction (210).

A Rasch analysis in a population of patients with upper extremity musculoskeletal disorders provides evidence that respondents were unable to discern the five response levels proposed (210). Additionally, an item misfit was found with the combination of conceptually disparate elements, such as pain, difficulty

sleeping, stiffness, tingling, and feeling less capable, into a symptom/disability total score. The two most misfitting items were 26 (tingling) and 21 (sexual activities) (210). During the development of the DASH, neither clinimetric nor psychometric strategies selected these two items during the process of item reduction; they were added only after clinicians' input (223).

Rasch analysis of the DASH has also been conducted in a population of patients with multiple sclerosis and showed that for a proportion of the items, the scoring function does not work as intended, because the patients' responses were not consistent with those predicted by the Rasch model (209). The corresponding results for the QuickDASH were better (211) but were also criticized (210).

The DASH demonstrates high test-retest reliability, with reported ICCs ranging from 0.89 to 0.98 for the DASH (45,136,207,216,218,220,224) and 0.90 to 0.94 for the QuickDASH (211,222,225).

Validity. Pearson's or Spearman's coefficients for correlation of the DASH with other general and shoulder-specific outcome measures (and the population they were measured in) are as follows:

- SPADI: $r = 0.93$ in patients who had undergone shoulder arthroplasty (28); $r = 0.83$ in patients with glenohumeral arthritis or soft tissue disorders around the shoulder, predominantly rotator cuff tendinitis (SPADI function, no overall SPADI assessed) (216); $r = 0.55$ at baseline and $r = 0.50$ for 3-week change scores in patients with adhesive capsulitis (before treatment and 3 weeks after arthrographic joint distension) (35); $r = 0.75$ in impingement (220); and $r = 0.82$ for the 3- to 4-month change score in patients with subacromial pain syndrome referred to physical therapy (226).
- CM: $r = 0.82$ in patients who had undergone shoulder arthroplasty (28).
- SF-36 PCS: $r = 0.67$ post shoulder arthroplasty (28); $r = 0.50$ for 6-month change scores post rotator cuff tear repair (7); $r = -0.59$ for impingement (220); and r = –0.70 in patients with rheumatoid arthritis and shoulder complaints (218).
- SF-12 PCS: r = –0.75 in proximal humeral fractures (224) and for musculoskeletal complaints of the neck, shoulder, and/or arm in primary care; $r = 0.61$ (pain in shoulder, arm, and hand only); $r = 0.63$ (pain in neck and shoulder); $r = 0.61$ (pain in neck, shoulder, arm, and hand); and $r = 0.62$ (pain in shoulder, arm, and hand) (227).
- SF-36 MCS: $r = 0.06$ post shoulder arthroplasty (28); $r = 0.29$ for 6-month change scores post rotator cuff tear repair (7); $r = -0.17$ in impingement (220); and $r = -0.27$ in patients with rheumatoid arthritis and shoulder complaints (218).
- SF-12 PCS: for musculoskeletal complaints of the neck, shoulder and/or arm in primary care; $r = 0.10$ (pain in shoulder, arm, and hand only), $r = 0.19$ (pain in neck and shoulder); $r = 0.16$ (pain in neck, shoulder, arm, and hand); and $r = 0.15$ (pain in shoulder, arm, and hand) (227).
- ASES: $r = 0.79$ post shoulder arthroplasty: (r = 0.59 for the clinician assessment component of the ASES) (28).
- HAQ: $r = 0.54$ at baseline and $r = 0.35$ for 3-week change scores in patients with adhesive capsulitis before treatment and 3 weeks after arthrographic joint distension (35), and $r = -0.88$ in patients with rheumatoid arthritis and shoulder complaints (218).
- SST: $r = 0.72$ for 6-month change scores in patients undergoing operative repair of a rotator cuff tear (7).
- WORC: $r = 0.71$ for 6-month change scores in patients undergoing operative repair of a rotator cuff tear (7).
- EuroQol-5D (EQ-5D): $r = -0.75$ in patients with a proximal humeral fracture (224).
- 28-Joint Disease Activity Score: $r = 0.42$ in patients with rheumatoid arthritis and shoulder complaints (218).
- Global Disability Rating: $r = 0.71$ and $r = 0.67$ 3-month change after physical therapy in patients with shoulder problems (45).

Correlation of the QuickDASH with other general and shoulder-specific outcome measures, and the population measured in, are as follows: for the SPADI, $r = 0.84$ post shoulder arthroplasty (228) and for the SF-36 PCS, $r = 0.68$ post shoulder arthroplasty (228).

Responsiveness. Reported ESs and SRMs of the DASH total score (according to patient population) are as follows:

- Total shoulder arthroplasty: ES = 1.19; SRM = 1.22 points (44).
- Arthroscopic acromioplasty: ES = 0.9, SRM = 0.5 points (221).
- Physical therapy for shoulder problems: ES = 1.06, SRM = 1.08 points (45).
- Surgical intervention for glenohumeral arthritis or soft tissue disorders around the shoulder, predominantly rotator cuff tendinitis: ES = 0.64; SRM = 0.81 points (216).
- Adhesive capsulitis and arthrographic joint distension: ES = 0.58 and SRM = 0.96 points in patients receiving treatment of known efficacy, ES = 0.55 and SRM = 0.91 points in patients reporting marked or moderate improvement, and ES = 0.83 and SRM = 1.45 points in patients reporting marked improvement (35).

ESs and SRMs of the QuickDASH total score according to patient population have been reported as follows: ES = 1.26 for total shoulder arthroplasty (228), and SRM = 0.79 in conservative treatment of shoulder or hand complaints (211).

Minimally important differences. The reported MDC for the DASH ranges between 12 and 19 points. It is 12.2 points in patients with shoulder problems treated with physical therapy

(45), 12.75 points in glenohumeral arthritis or rotator cuff tendinitis treated surgically (216), 13.1 points in impingement with no treatment (220), 11.8 points in impingement with physical therapy, 19 points in humeral shaft fractures treated both operative and nonoperatively (229), and 16.3 points in surgical and non-surgical interventions for shoulder complaints (175).

The reported MCIDs range between 4.4 and 12.4, which is notably less than reported MDCs. The MCID is 6.7 points in humeral shaft fractures treated both operative and nonoperatively (229), 4.4 points in physical therapy for impingement (226), and 10.2 points in shoulder problems treated with physical therapy (45); the minimum important change (MIC) is 12.4 points in surgical and nonsurgical interventions for shoulder complaints (175).

MDCs and MCIDs reported for the QuickDASH include the following: MDC = 11.2 percentage points and MCID 8.0 points in physical therapy for shoulder pain (225); SDC = 17.1 points and MCID = 13.4 points in surgical and nonsurgical interventions for shoulder complaints (175).

Generalizability. The DASH has been used to assess outcomes from various nonoperative and surgical interventions in people with varying shoulder conditions, including glenohumeral arthritis, rotator cuff disease, proximal humerus fractures, adhesive capsulitis, rheumatoid arthritis, and multiple sclerosis. The availability of the DASH in so many languages adds to its generalizable use.

Use in clinical trials. The DASH or QuickDASH has been used in 35 randomized controlled trials of interventions for various shoulder disorders through the end of 2015 (2). Recent examples include trials of open reduction internal fixation for humeral shaft fracture (230), exercise to prevent shoulder problems in women undergoing breast cancer treatment (231), and platelet-rich plasma therapy for degenerative tendinopathies (232).

Critical appraisal of overall value to the rheumatology community

Strengths. The DASH has been extensively tested and is widely used as a self-assessment instrument for the shoulders and other disorders of the upper extremities. It is particularly useful in cases in which measurements of symptoms and function of the entire upper extremity are wanted. Because shoulder function determines the position of the elbow and the hand, the DASH is also useful in all elbow and hand conditions. The QuickDASH total score yields very similar values to those of the DASH, and the total scores correlate highly with each other.

Caveats and cautions. The DASH covers the whole upper extremity, not only the shoulder. Its specificity and responsiveness are generally lower than those of shoulder-specific tools, although these are higher than those of generic quality of life tools. This limits its usefulness in assessing outcomes from shoulder-specific conditions. The DASH includes items relating to the lower part of the limb in which (dis)ability is not expected to change, and this may account for its poorer performance. There is evidence that the DASH score is also influenced by the disability of the lower extremity. In addition, Rasch analysis has revealed scoring issues and item misfit. Compared with other instruments, the strict 90% missing rule produces a relatively high percentage of missing data.

Clinical usability. The DASH is a good tool for the comprehensive assessment of upper extremity conditions, ie, if shoulder problems cannot be differentiated from hand problems. It is easy to apply, analyze, and interpret. The comparison of empirical and normative data allows for valid description of the patient's upper extremity status. The QuickDASH provides the necessary short assessment for clinical visits.

Research usability. The DASH is good for research purposes in various upper extremity conditions. It is well tested, and there is a large body of data for the comparison of different settings and different upper extremity instruments, especially for the analysis of construct validity, compared with other instruments. However, concerns about its lack of specificity to shoulder function and its relatively limited responsiveness indicate that it is not the instrument of choice for shoulder function.

UNIVERSITY OF CALIFORNIA, LOS ANGELES SHOULDER RATING SCALE

Description

Purpose. This multidimensional scale was originally developed to obtain information about patients with shoulder arthritis undergoing shoulder arthroplasty (233). It has subsequently been used to assess the outcomes of patients with other shoulder conditions undergoing surgery or nonoperative treatment. The original three-item version of the UCLA Shoulder Rating Scale was published in 1981 and completed by the clinician (233). No details on the method of development were published. A modified UCLA Scale that includes both self-assessment and clinician-assessed items to assess outcome from rotator cuff surgery was published in 1986 (234); this is the version that is most widely used. A self-reported version of the modified UCLA Scale that modified the two clinician-assessed items for self-report was published in 2008 (235).

Content or domains. The original UCLA Scale includes the following three domains: pain, function, and muscle power and motion (233). The most commonly used version (234), as well as the patient-completed version (235), includes the following five domains: pain, function, active forward elevation, strength, and

patient satisfaction (after surgery). Other versions include ones with two additional domains (abduction ROM and abduction strength) (236), the inclusion of the patient satisfaction item to the original UCLA Scale (146), the exclusion of patient satisfaction (because it asks about satisfaction of treatment so is not suitable as a measure of status) (237), and an Italian version that includes only pain and function (25).

Number of items. The number of items varies between three items (original UCLA Scale) (233) and seven items (modified version with the addition of two new items for abduction) (236). The most commonly used version (the modified UCLA Scale) and the patient-reported UCLA Scale both have five items (234,235). One version has four items (excluding the patient satisfaction item) (237), and the Italian version has only two items (pain and function) (25).

Response options/scale. In the original UCLA Scale, all items have weighted scores out of 10, based on categorical responses (233). For pain: constant and unbearable, with strong medication frequently = 1 point; constant but bearable with strong medication occasionally = 2 points; none or little at rest, occurs with light activities, with salicylates = 4 points; with heavy or particular activities only, with salicylates occasionally frequently = 5 points; occasional and slight = 8 points; and no pain = 10 points. For function, unable to use arm = 1 point, very light activities only = 2 points, light housework or most daily living activities = 4 points, most housework, washing hair, putting on brassiere, shopping, and driving = 5 points, slight restriction only or able to work above shoulder level = 8 points, and normal activities = 10 points. For muscle power and motion, ankylosis with deformity = 1 point, ankylosis with good functional position = 2 points; muscle power poor to fair and elevation less than 60° with internal rotation less than 45° = 4 points, muscle power fair to good and elevation 90° with internal rotation 90° = 5 points, muscle power good or normal and elevation 140° with external rotation 20° = 8 points, and normal muscle power and motion near normal = 10 points.

In the modified version, there are six similar response options for pain and function, although the numerical weightings differ (234). Pain options are as follows: present always and unbearable, with strong medication frequently (1 point); present always but bearable, with strong medication occasionally (2 points); none or little at rest, present during light activities, with salicylates used frequently (4 points); present during heavy or particular activities only, with salicylates used occasionally (6 points); occasional and slight (8 points); and none (10 points). Function options are as follows: unable to use limb (1 point); only light activities possible (2 points); able to do light housework or most ADLs (4 points); most housework, shopping, and driving possible and able to do hair and to dress and undress, including fastening a bra (6 points); slight restriction only and able to work above shoulder level (8 points); and normal activities (10 points). There are also six response options for both active forward flexion and strength of forward flexion (manual muscle testing), and there are two response options for satisfaction of patient. The six response options for active forward flexion are 150° (5 points), 120° to 150° (4 points), 90° to 120° (3 points), 45° to 90° (2 points), 30° to 45° (1 point), and less than 30° (0 points). The responses for strength of forward flexion (manual muscle testing) are Grade 5 (normal; 5 points), Grade 4 (good; 4 points), Grade 3 (fair; 3 points), Grade 2 (poor; 2 points), Grade 1 (muscle concentration; 1 point), and Grade 0 (nothing; 0 points). The response options for the satisfaction of patient are satisfied and better (5 points) or not satisfied and worse (0 points).

In the patient-reported UCLA Scale, the response options for pain, function, and satisfaction are the same (235). For the active forward flexion, the patient is asked how high they can lift their arm forward, with the degrees shown pictorially. For the strength of forward flexion, they are asked how strong their arm is by comparison with the power/strength in their other arm. There is a picture asking them to ask someone to resist them as they lift their arm up. The response options for this item are normal strength (5 points), good strength/a bit weaker (4 points), fair strength/moderately weaker (3 points), poor strength/much weaker (2 points), muscle contraction only (1 point), and nothing (0 points). In the version that added items measuring abduction range of motion and abduction strength, the response items and scores (up to 5 points) are the same as those of the active forward flexion and strength of forward flexion items (236).

Recall period for items. The recall period is not specified in the original version, but the current online modified UCLA Scale version specifies the past 4 weeks (see below).

Cost to use. Free of charge.

How to obtain. The modified UCLA Scale can be downloaded for free from https://www.orthopaedicscores.com and https://www.orthotoolkit.com/UCLA-shoulder/.

An abridged Italian version (only pain and function items) can be found in refs. 25 and 170, and a Brazilian Portuguese version is published in ref. 238.

Practical application

Method of administration. In the original UCLA Scale, all three items are completed by the clinician (233). In the modified version, three items are completed by the patient (pain, function, and satisfaction), and two are completed by the clinician or observer (active forward flexion and strength of forward flexion) (234). In the modified patient-reported version, all items are completed by the patient (235).

Scoring. In the original 1981 version of the UCLA Scale, each of the three items is scored out of a maximum of 10 points (233). It was noted in this version that occasionally the analysis of pain, function, and ROM does not fit exactly the numerical criteria indicated, and the interval numbers provide flexibility for in-between indications.

In the modified UCLA Scale and the patient-reported version, each of the five items contribute unequally to an overall score out of 35, with higher scores indicating better outcomes, as follows: pain (10 points; weighting 28.6%), function (10 points; weighting 28.6%), active elevation (5 points; weighting 14.3%), strength of forward flexion (5 points; weighting 14.3%), and satisfaction (5 points; weighting 14.3%) (234,235). The reason for the weighting was not described in the modified version. The modified scale can also be converted into a 100-point scale. There are no subscales.

Computer scoring is not necessary, but online scoring programs are available at http://www.orthopaedicscores.com and https://www.orthotoolkit.com/UCLA-shoulder/.

Score interpretation. In the original UCLA Scale, a score greater than 8 for pain, function, and ROM was considered to be excellent, greater than 6 was considered good, 2 to 4 was fair, and less than 3 was poor (233). In the modified and patient-reported versions of the UCLA, a score of 34 or 35 is considered excellent, 28 to 33 is considered good, 21 to 27 is considered fair, and 20 or less is considered poor (234,235). Good or excellent scores (above 27) indicate a satisfactory result, whereas fair or poor scores (27 and below) indicate unsatisfactory results. One study used cutoffs of 34 or 35, 29 to 33, and less than 29 to indicate excellent, good, and poor results, respectively (239). In the seven-item version of the UCLA Scale (range of possible scores 2-45), scores of 41 to 45 and 36 to 40 indicate excellent and good scores, respectively (236).

Normative data for the modified four-item UCLA Scale (leaving out the patient satisfaction item) (range 2-30) have been reported in 635 asymptomatic, healthy volunteers in Australia (n = 323) and Canada (n = 312) (9). People without a history of shoulder pathology reported a higher (better) mean UCLA Scale score compared with those with a history of a shoulder problem (28.9 [range 2-30] versus 28.3; Wilcoxon's rank sum test $P = 0.0005$). No differences in scores were observed in people with and without a current wrist or elbow problem or handedness. There was no important difference between sexes adjusted for nationality (mean of 28 in women versus 29 in men; $P = 0.4768$). Mean score declined slightly with increased age after ages 50 to 59. An unexplained statistically significant difference in scores overall (and by age group) was observed between countries, with a trend for lower scores among Canadians.

Normative data for the modified four-item UCLA Scale in 120 shoulder-throwing and nonthrowing college athletes and recreational athletes (aged less than 40 years or 40 years and older without a history of shoulder or neck pain) have also been reported (237). Across all groups of athletes, the normalized score was 98% (95% CI 75-100) and ranged from 97% (85.7%-100%) for recreational athletes 40 years or older to 99% (89.3%-100%) for nonthrowing college athletes. Men were found to score higher on the active forward flexion subscale score (0.15 points higher of 5 points; $P = 0.004$). Subgroup type contributed significant but small effects for the pain subscale.

Respondent time to complete. Time to complete the UCLA Scale has not been reported. However, it is easy to understand and relatively short (five items for the modified UCLA Scale) and is likely completed within a few minutes.

Administrative burden. All versions other than the self-reported version of the UCLA Scale require an assessor and face-to-face interaction to assess the forward flexion and strength of forward flexion items. Item ratings can be typed or scanned into an electronic database. Score computation is easy (see above). Little special training is necessary for these activities.

Translations/adaptations. There is a Brazilian Portuguese version of the modified UCLA Shoulder Rating Scale (238). It modified five response options for pain and four for function. There is also an Italian version of the pain and function items of the modified UCLA (25). Both translated versions included cross-cultural adaptation.

Psychometric information

Floor and ceiling effects. The presence/absence of floor and ceiling effects have not been reported.

Reliability. The modified UCLA Scale has single items measuring each domain, so internal consistency cannot be measured. However, one study found that the Italian version of the UCLA Scale pain and function items had a high internal consistency, with a Cronbach's α of more than 0.89, and both items correlated with the total score ($r > 0.54$) (25).

Good test-retest reliability of the three self-reported items in the modified UCLA Scale has been reported in a mixed post–shoulder surgery population (N = 31) with mostly rotator cuff tears (n = 20) and instability (n = 9); results are as follows: for pain, ICC = 0.78 (95% CI 0.58-0.89); for function, ICC = 0.89 (95% CI 0.78-0.94); and for satisfaction ICC = 0.79 (95% CI 0.59-0.89) (68). However, the same study reported poor test-retest reliability in a mixed group who had not undergone surgery, with the following results: for pain, ICC = 0.59 (95% CI 0.25-0.80) and for function, ICC = 0.51 (95% CI 0.14-0.75); satisfaction was unable to be estimated. The authors suggested that the modified UCLA Scale may not be appropriate for use in a nonsurgical population. Another study, however, found good

test-retest reliability in 20 patients who had surgical or conservative management at 6, 12, and 13 weeks after proximal humerus fracture (ICC = 0.93 [95% CI 0.76 to 0.97]) (145). The Pearson correlation coefficient was $r > 0.91$ for test-retest of the Italian version of the pain and function items in 40 patients with damage to the spinal accessory nerve during neck resection for head and neck cancer causing shoulder syndrome (25). There was also acceptable test-retest reliability of the two-item Italian version of the modified UCLA Scale in a population of patients who had been treated with surgery for anterior shoulder instability ($r = 0.93$ for the UCLA pain subscale; $r = 0.95$ for UCLA function subscale) (170).

Validity. One study in patients undergoing rehabilitation post proximal humerus fracture assessed the content validity of the modified UCLA score linked to ICF codes (145). Apart from subjective shoulder value ('not definable'), the items linked with the following two ICF components: body functions (pain, joint mobility, and muscle strength/endurance) and activities and participation (using transport/driving, self-care, dressing, shopping, and housework). All items linked with body functions were separate items, whereas all items in activities and participation were combined within a single item.

Pearson's or Spearman's correlations of the modified UCLA Shoulder Rating Scale to other instruments (and the population they were measured in) are as follows:

- ASES: very high correlation overall ($r = 0.91$) in an arthroscopic rotator cuff population, with moderate correlation in the preoperative period ($r = 0.67$), high correlation at 6 months after surgery ($r = 0.87$), and very high correlation at 12 and 24 months after surgery ($r = 0.90$ and 0.92, respectively) (73); and 0.38 in a mixed population with rotator cuff disease, isolated SLAP lesions, and shoulder instability after surgery (69).
- WORC: $r = 0.80$ in a mixed population of Brazilian patients with rotator cuff disorders (240).
- WOSI: $r = 0.649$ at baseline (prior to intervention) and $r = 0.694$ for change score (baseline to 3 months) in patients having treatment for instability (147).
- OSS: $r = 0.77$ (at 6 weeks), 0.83 (at 12 weeks), and 0.73 (for change scores from 12 to 13 weeks) in proximal humerus fractures (145); $r = 0.66$ in impingement or calcific tendinitis in German patients (187); and $r = 0.67$ in Italian patients after shoulder surgery for anterior instability (185)
- DASH: $r = 0.65$ (at 6 weeks), 0.84 (at 12 weeks), and –0.65 (for change scores from 12 to 13 weeks) in proximal humerus fractures (145).
- SSV: $r = 0.45$ (at 6 weeks), 0.81 (at 12 weeks), and 0.46 (for change scores from 12 to 13 weeks) in proximal humerus fractures (145).
- CS: $r = 0.70$ (at 6 weeks), 0.92 (at 12 weeks), and 0.83 (for change scores from 12 to 13 weeks) in proximal humerus fractures (145).

One study assessed the correlation of the modified UCLA Scale with other shoulder measures in people without shoulder complaints, with the following results: $r = 0.65$ for the ASES, $r = 0.42$ for the CM Shoulder Score, $r = 0.80$ for the OSS, $r = –0.68$ for the SPADI, and $r = –0.63$ for the SPONSA (9). Correlation of the four-item UCLA Scale has been reported in patients with previous rotator cuff repair as $r = 0.76$ for the SST and $r = 0.66$ for the CS (146). The UCLA Scale scores were higher than the CS scores in almost all participants, whereas correlation with forward motion and the abduction ratio was only 0.37 (poor) and 0.48 (fair), respectively.

Correlations of the modified UCLA function subscale score in a mixed shoulder population (46% of whom were post-surgery) were: $r = 0.64$ for the SPADI disability subscale and $r = 0.60$ for the SST (29). Correlations with the UCLA pain subscale score were $r = 0.63$ for the SPADI pain subscale, whereas discriminant validity of the modified UCLA pain and disability subscales was $r = –0.64$, indicating greater convergence than divergence between these subscales. In a study including patients after shoulder surgery for anterior instability, correlations of the Italian version of the pain item of the modified UCLA Scale were as follows: $r = –0.63$ for the total SPADI, $r = –0.61$ for the SPADI pain subscale, $r = –0.64$ SPADI function subscale, and $r = 0.61$ for the SST. In the same study, correlations with the function item were $r = –0.57$ for the total SPADI, $r = –0.52$ for the SPADI pain subscale, $r = –0.60$ for the SPADI function subscale, $r = 0.51$ for the SST (170).

Agreement between the modified UCLA Scale and the patient-reported UCLA Scale. There was acceptable agreement in a postoperative population of patients who had undergone either arthroscopic subacromial decompression or rotator cuff repair ($N = 100$), with values as follows: ICC = 0.910 (95% CI 0.87-0.94) for the whole cohort, ICC = 0.951 (95% CI 0.92-0.97) for the cohort with subacromial decompression (n = 46), and ICC = 0.734 (95% CI 0.61-0.83) for the cohort with rotator cuff repair (n = 54) (235). Patient-derived scores were slightly lower (worse outcomes) in the order of one point compared with two clinician assessors, which is mostly explained by differences in the strength of forward flexion item. This could have arisen because of patient apprehension in pushing against resistance when performing the self-assessment compared with being more confident when this was performed in the presence of the clinician and/or greater clinician encouragement to push against resistance despite any pain.

Responsiveness. Reported ESs and SRMs of the modified UCLA Scale according to patient population are as follows:

- Subacromial decompression: ES = 1.17, 2.0, and 2.73 points

at 6 weeks, 6 months, and 12 months, respectively and SRM = 0.83, 1.41, and 1.69 points at 6 weeks, 6 months, and 12 months, respectively (151).
- Surgery for instability: SRM = 0.385 points (147).
- Rotator cuff disorders three months after treatment (surgery or physical therapy) (Brazilian Portuguese version): ES = 1.17 and SRM = 1.66 points (241).

Minimally important differences. In patients being treated for proximal humerus fractures, the MCID for improvement was found to be 2.4 points from an anchor-based method (mean score of five patients who reported a small overall improvement was used to anchor the MCID for improvement) and 2.0 points from a statistical distribution-based method (145). No patients deteriorated, so no MCID for deterioration was reported.

The MDC of the Italian version of the UCLA Scale was found to be 0.90 (1.77) for pain and 0.15 (0.30) for function (170).

Generalizability. The UCLA Scale and later versions have been predominantly used to assess outcomes from various surgical procedures (shoulder arthroplasty, arthroscopic subacromial decompression, open and arthroscopic rotator cuff repair, and surgery for anterior instability) but has also been used to assess proximal humerus fractures and shoulder syndrome due to damage to the spinal accessory nerve during neck dissections of head and neck cancers.

Use in clinical trials. Through December 2015, the modified UCLA Scale or other versions has been used in 35 randomized controlled trials (2). Examples from the last 5 years include therapeutic ultrasound (242), mirror therapy (243), and platelet-rich plasma injection (244) for adhesive capsulitis; deep heat versus ultrasound for shoulder pain (46); immediate versus delayed passive mobilization (245,246), doxycycline (247); and platelet-rich plasma (248) following rotator cuff repair; double- versus single-row rotator cuff repair (249,250); suture-spanning augmentation of single-row rotator cuff repair (117); rotator cuff repair with or without acromioplasty (87); acupuncture (251), kinesiotaping, and subacromial corticosteroid injection (252), ultrasound-guided subacromial NSAID, or glucocorticoid injection (253) for impingement syndrome; and open versus arthroscopic repair of traumatic anterior shoulder instability (254). Only one older trial to date, which assessed the efficacy of ropivacaine infusion versus placebo following rotator cuff surgery, has used the patient self-reported version of the UCLA Scale (255).

In many trials conducted in countries where English is not the first language, it is unclear whether or not the UCLA Scale was cross-culturally translated into the local language, eg, in trials conducted in Korea (46), Turkey (243,252), China (244,246,253), Belgium (245), Mexico (247), Italy (249), Taiwan (117), Brazil (248), Spain (251), and Japan (250,254). One ongoing trial comparing figure-of-eight bandages versus arm slings for treating middle-third clavicle fractures in adults is using the Brazilian Portuguese version of the modified UCLA Scale (256).

Critical appraisal of overall value to the rheumatology community

Strengths. The psychometric properties of the UCLA Scale and its modified versions have been tested across many noninflammatory shoulder disorders in surgical and nonsurgical populations. Normative data are available, and the UCLA Scale score does not appear to be influenced by sex or handedness, but scores do decline slightly with age. A self-administered version, which performs acceptably compared with an observer's assessment of forward flexion and strength of forward flexion, is available and may be useful in situations in which face-to-face assessment is not possible or desirable. The modified UCLA Scale appears to have good test-retest reliability and has moderate to good construct validity in comparison with similar instruments. It is responsive to change across a range of conditions, but the minimally important difference has only been reported for people following proximal humeral fractures.

Caveats and cautions. The modified UCLA Scale is a multidimensional tool with only single items measuring pain, function, forward flexion, and strength of forward flexion. An additional item, patient satisfaction, can only be applied following treatment. A major issue with many of the item responses for pain is that they combine frequency and severity of pain with the type and amount of medication required to relieve the pain within single response options, which may make it difficult to choose an appropriate response. This problem also affects some of the response options for function that combine multiple functional activities within the single option. The validity of the weightings of responses within the pain and function items has not been evaluated. The satisfaction item is also problematic; only two response options are possible (satisfied and better or not satisfied and worse), so it is unclear how to respond if the respondent considers themselves unchanged. In addition, this item implies an intervention has taken place and could not be administered before an intervention. Finally, a patient might be unchanged or worse but be satisfied, or they might be better but still not be satisfied.

Combining multiple domains into a single total score means that the score cannot inform specifically about the pain or function or other constructs a patient experiences. Multiple-item scales also yield much more reliable measurements than do single-item scales. It is doubtful, therefore, that the UCLA Scale is precise enough to effectively follow the progress of individual patients in the clinic setting (29). Floor and ceiling effects have not been reported.

Clinical usability. Versions with a combination of self and clinical assessment and solely self-assessment are available. It is widely used to assess outcomes from shoulder surgery and nonoperative treatments across a range of conditions. Further

assessment of its measurement properties is still needed, and caution is necessary when interpreting the total score, which combines single items across several domains.

Research usability. Use with caution because of its many identified caveats. Further testing is needed to fully understand its psychometric properties for various patient groups. Cross-culturally appropriate translations into languages other than Brazilian Portuguese and Italian are needed and should be used for studies for trials and other research conducted in languages other than English. The patient self-assessment version of the UCLA Scale might be preferred when face-to-face assessment is not possible or desirable.

WESTERN ONTARIO ROTATOR CUFF INDEX

Description

Purpose. The original version of the WORC Index was published in 2003 and developed to specifically measure disorders of the rotator cuff (257). It has also been used and evaluated for patients with rotator cuff repair (258), shoulder instability (259), and winged scapula (260).

Content or domains. The WORC Index contains the following five domains: physical symptoms, sports/recreation, work, lifestyles, and emotions (257).

Number of items. There are 21 items in total (six for physical symptoms, four for sports/recreation, four for work, four for lifestyles, and three for emotions) (257).

Response options/scale. All items are rated using a VAS (257). On the paper version, the lines are 100 mm long.

Recall period for items. The recall period is 2 weeks.

Cost to use. Free of charge.

How to obtain. It is available in various references, eg, ref. 257.

Practical application

Method of administration. The WORC Index is a self-assessment using pen and paper (257), but it has also been deliverd electronically in many studies.

Scoring. Each item is scored from 0 (best) to 100 (worst). The total score ranges from 0 to 2100 (257) but is often normalized as a percentage (out of 100%), with a higher percentage indicating better function.

Score interpretation. A higher score indicates poorer function, but on the percentage scale, a higher percentage indicates better function. There are no cutoff points to indicate severity. One study reported that before treatment, the average WORC score among patients with rotator cuff repair, disorders of the rotator cuff, and shoulder instability was 37.8, 48, and 55.7, respectively (259), and it increased to 78.8, 737.7, and 89.8, respectively, 6 months after treatment. Another study recruiting a mixture of patients with impingement syndrome or rotator cuff pathology reported that the WORC score increased from 39.85 before treatment to 70.15 at 6 months after treatment (149).

Respondent time to complete. The mean time to complete the WORC Index is 3 minutes.

Administrative burden. Scoring is straightforward, and no special software or equipment is required (259). The VAS is less favored by some patients and may take more time to record scores.

Translations/adaptations. The WORC Index has been translated into multiple languages. Cross-cultural validation has been performed for the following languages: Chinese (261), Swedish (262), Canadian French (263), Polish (264), Turkish (265), Brazilian Portuguese (240), Persian (240,266,), and Norwegian (15,27).

Psychometric information

Floor and ceiling effects. The floor and ceiling effects of the WORC Index overall score were 1.1% and 5.6%, respectively, in a study of patients with rotator cuff disease and shoulder instability (259). They were low across all domains, except for floor effects in the work domain (12.2%) and ceiling effects in the emotions domain (14.4%). A substantial ceiling effect (39.1%) of the WORC Index overall score was observed 6 months after treatment when patients had improved, and this was evident across all domains (range 34.5%-56.3%) except for the shoulder hindrance score (0%).

Reliability. There is evidence for high test-retest validity at 2-week intervals, with ICCs ranging from 0.54 (work) to 0.91 (physical symptoms) for the domains and 0.96 for the total score (257).

Validity. The convergent validity of the WORC Index is relatively good. The Pearson correlation coefficient between the WORC and the absolute CS, the relative CS (age-adjusted), and the ASES was r = 0.65, 0.66, and 0.73, respectively, in patients with impingement or rotator cuff pathology, a third of whom fulfilled

criteria for surgery (149). In another study, the correlation between the WORC Index and the physical summary score of the SF-36, the SST, and the DASH was $r = 0.58$, 0.91, and 0.88, respectively, in patients following rotator cuff repair (267). However, lower correlations were reported between the WORC Index and the ASES, the UCLA Scale, and the DASH in another study of patients with rotator cuff disease receiving various treatments ($r = 0.68$, 0.48, and 0.63, respectively) (257).

The longitudinal convergent validity has been reported to be relatively high. In one study that measured change in patients with impingement or rotator cuff pathology, a third of whom fulfilled criteria for surgery, at 6 months after treatment, the correlation between the change in the WORC Index and the change in the absolute and relative (age-adjusted) CS and the ASES was $r = 0.77$, 0.70, and 0.85, respectively (149). Another study in patients with rotator cuff disease receiving various treatments, reported correlations of the change in scores between the WORC Index and the ASES, UCLA Scale, and DASH of $r = 0.76$, 0.72, and 0.66, respectively (257).

The discriminant validity of the WORC Index is satisfactory, as shown by correlations with objective measurements. For example, the correlation between the WORC Index and external and internal rotation strength and external and internal rotation ROM was $r = 0.38$, 0.45, 0.27, and 0.31, respectively, in patients following rotator cuff repair (267). The WORC Index also has been reported to have low correlations with the subscales of the SF-36 ($r < 0.5$), except for physical function ($r = 0.56$), in patients with rotator cuff disease (257). Another study reported correlations for the WORC and the strength of the affected arm as well as the unaffected arm as $r = 0.42$ and 0.1, respectively, in patients with impingement or rotator cuff pathology, a third of whom fulfilled the criteria for surgery (149).

The longitudinal discriminant validity of the WORC Index is also satisfactory, as shown by the low correlations between the change of the WORC Index and the change of all subscales of the SF-36 ($r < 0.5$), except for role physical ($r = 0.52$), when evaluated 3 months after various treatments for rotator cuff disease (257). The correlations between the change in the WORC Index and change of strength of the affected and unaffected arms was $r = 0.37$ and 0.20, respectively, when evaluated 6 months after various treatments in patients with rotator cuff disease in another study (149).

The WORC Index has good known-group validity. For example, the WORC Index can differentiate between patients whose occupations are affected by shoulder symptoms and those whose occupations are not (149). The WORC Index can also detect the difference between patients whose ROM in external rotation is greater than 45° and the ones whose ROM is less than 45° and between patients receiving or not receiving workers' compensation (267). The longitudinal known-group validity of the WORC Index was questioned in one study on the basis that it failed to detect change in patients with less pathology compared with those with more pathology, as determined by the requirement for more extensive surgery at 3 and 6 months after treatment (149).

Responsiveness. As a specialized instrument, the WORC Index has good responsiveness. The correlation between the WORC change score and the shoulder hindrance score is $r = 0.51$, 0.55, and 0.43 among patients with rotator cuff repair, disorders of the rotator cuff without rupture, and shoulder instability, respectively (259). In another study, an analysis of variance showed that the WORC score can differentiate between change from baseline at 3 months and 6 months after rotator cuff surgery treatment, with an SRM of 1.33 (149).

Minimally important differences. The MDC of the WORC Index among patients with rotator cuff repair, disorders of the rotator cuff without rupture, and shoulder instability in one study was 16.7%, 20.3%, and 25.4%, respectively (259). In the same study, the MIC of the WORC estimated by the ROC cutoff was 34%, 22.9%, and 31.8%, respectively, whereas the MIC estimated by the 95% limit cutoff was 35.3%, 41.9%, and 46%, respectively.

Generalizability. The WORC Index is a rotator cuff–specific measurement instrument that has been evaluated among participants with various rotator cuff disorders as well as shoulder instability. It has also been shown to have moderate validity among patients with winged scapula (260).

Critical appraisal of overall value to the rheumatology community

Strengths. The WORC Index is relatively short, is easy to administer and score, and has good overall evidence for construct validity and responsiveness. It is readily obtained at no cost, and cross-cultural validation has been performed for multiple languages.

Caveats and cautions. More studies are needed to evaluate measurement error, interpretability (268), and reliability in patients with rotator cuff disorders. The responsiveness of the WORC Index has been questioned for some other languages (27,269).

Clinical usability. The WORC Index is useful in a clinical context for assessing rotator cuff disorders. It is relatively brief, easy to administer and score, and responsive to change.

Research usability. The WORC Index is brief, easy to administer, and responsive, and it is valid for assessing various treatments in many settings. It is available for use in multiple languages.

PROMIS PHYSICAL FUNCTION UPPER EXTREMITY CAT

Description

Purpose. The PROMIS PF Upper Extremity CAT measures upper extremity function (270). It was originally developed in 2014 as part of the PROMIS initiative, which is a National Institutes of Health Common Fund project involving the dynamic assessment of patient-reported outcomes. The PROMIS includes item banks that measure key health symptoms/concepts for both the general population and several chronic conditions. PROMIS item banks assess physical (physical function, fatigue, sleep disturbance, sleep related impairment, pain behavior, and pain interference), emotional (depression, anxiety, and anger), cognitive (applied cognition abilities and applied cognition general concerns), and social (ability to participate, satisfaction with social roles and activities, emotional support, instrumental support, informational support, and social isolation) health. These item banks were developed following rigorous protocols that involved extensive formative research and statistical analysis (271-275). This included state-of-the-art psychometric analysis, including classical test theory approaches and item response theory. Items can be administered as a full set, a CAT, or a calibrated short form (preselected set of items). An item bank that can be administered as a CAT provides an advantage to any preselected set of items (ie, short form) in maximizing assessment sensitivity and simultaneously minimizing the number of items needed.

Content or domains. It measures upper extremity physical function.

Number of items. The CAT uses a maximum of 12 of the 16 items. The CAT presents a maximum of 12 questions in a dynamic order, with the exact question determined by the response to a prior question. The questionnaire is considered completed when the expected change in the score with additional questions drops below a specific threshold.

Response options/scale. Response options range from 1 to 5 (1 = unable to do, 2 = with much difficulty, 3 = with some difficulty, 4 = with a little difficulty, and 5 = without any difficulty).

Recall period for items. All items are phrased in the present tense.

Cost of use. Free of charge.

How to obtain. Free versions are available online at http://www.healthmeasures.net/explore-measurement-systems/promis.

Practical application

Method of administration. It is a self-assessment measure, using pen and paper for the static version or completed electronically for the CAT version.

Scoring. The questionnaire is considered completed when the expected change in the score with additional questions drops below a specific threshold. PROMIS measures are scored using a *t* metric with a mean of 50 and an SD of 10; higher scores indicate more of the construct (eg, higher scores indicate greater function).

Score interpretation. This approach allows for the estimation of an individual's functioning relevant to the reference group (which, for PROMIS, is the general population). For example, scores of 60 or greater on a PROMIS measure indicate that the individual has more of a specific trait (eg, physical function of the upper extremity) than 68.27% of people in the general population. Scores above 70 indicate that self-reported physical function exceeds 95.45% of people in the general population. These scores can be used as referents for making appropriate clinical comparisons.

Respondent time to complete. The PROMIS PF Upper Extremity CAT takes approximately 70 seconds to complete (range 25-307 seconds) according to one study (276), 61.3 seconds on average (SD 28.8) in another study (277), and 96.9 seconds (SD 25.1) in a third study (278). Because completion of the CAT depends on the exact responses to each item, this dictates the total number of items that are completed and the time involved.

Administrative burden. The PROMIS instruments can be scored on the Assessment Center web-based platform or by other electronic data collection methods (eg, RedCap) that have access to scoring algorithms for arriving at T scores. The Assessment Center enables the investigator to create a data collection website including any PROMIS item banks and short forms (279). It supports data collection designs that include multiple time points and multiple treatment arms and enables investigators to monitor the enrollment of participants and completeness of data collection during the course of their research.

Translations/adaptations. It has been translated into Dutch-Flemish (280).

Psychometric information

Floor and ceiling effects. No floor or ceiling effects have been found in studies including patients with hand or upper extremity conditions (276), patients with shoulder arthritis (278), patients with shoulder pain (277), patients undergoing primary

total shoulder arthroplasty (281), patients with varying types of rotator cuff disease (282), and patients with upper extremity fractures (283).

A study including patients who underwent operative treatments for shoulder instability reported ceiling effects at 6 months (68.1%) and 2 years (67.0%) (284). In a subgroup aged 21 years or younger, ceiling effects were 71.1% at 6 months and 81% at 2 years. Ceiling effects were found for 3% of patients (T score of 56 or higher) undergoing arthroscopic rotator cuff repair (285) and 7.2% (T score of 56 or higher) of patients with hand or shoulder conditions (286). The latter study also reported 1.2% floor effects. Another study showed ceiling effects in 11.4% of patients with shoulder instability, with varying ceiling effects for different age groups (287). A further study showed ceiling effects when inspecting item loadings on lower functional ability levels (288). A study in patients with upper extremity trauma found no floor effects but found ceiling effects in 5.2% (at a mean of 13 weeks post surgery) and 18.2% (at a mean of 37.9 weeks post surgery) of patients (289).

In a systematic review of floor and ceiling effects across 12 studies including 18113 patients, Gulledge et al reported floor effects ranging from 0% to 1.6% and ceiling effects ranging from 0% to 28% (290).

Reliability. There is some evidence of person and item reliability in patients with shoulder instability, with $r = 0.82$ to 0.96 for item reliability and $r = 0.84$ to 0.85 for person reliability (287,288). In patients with upper extremity trauma, the average marginal reliability was $r = 0.90$ (289).

Validity. A Rasch analysis of 1197 adult patients with varying hand and upper extremity complaints (nonshoulder) showed the item bank fit a unidimensional model (Eigen value of 1.96), with 4.2% unexplained variance (288). There was also no differential item functioning (local independence of items $r = –0.37$ to 0.34), and item fit was also adequate in the Rasch model. In another study of 734 patients with isolated shoulder, elbow, or wrist fractures, a factor analysis revealed that the PROMIS Physical Function Upper Extremity loads onto one factor only, reflecting capability, which was separate from another factor reflecting quality of life, although it did not explore differential item functioning (283).

There is evidence for low to high correlations between the PROMIS PF Upper Extremity and various generic quality of life measures, measures of psychological function, upper extremity measures, general measures of physical function and pain, shoulder-specific measures, and rotator cuff measures. These associations have been reported in a variety of settings (eg, academic medical centers and outpatient clinics) and in patients with a variety of shoulder conditions, such as humeral fractures, rotator cuff muscle and tendon tears, adhesive capsulitis, bursitis, tendinitis, impingement, and instability, and, in some cases, in mixed samples of hand, elbow, or shoulder conditions. Pearson's or Spearman's correlations or other relevant statistics of the PROMIS PF Upper Extremity CAT to other instruments (and the population they are measured in) are as follows:

- Short Musculoskeletal Functional Assessment: $r = –0.76$ to –0.69 in upper extremity trauma (289).
- QuickDASH: $r = –0.82$ to –0.75 in upper extremity trauma (289), $r = –0.81$ in hand and upper extremity conditions (276), and $r = –0.47$ to –0.83 (1 week and 9 months, respectively) in upper extremity fractures (283).
- PROMIS PF Short Form 8a: $r = 0.73$ to 0.77 in upper extremity trauma (289).
- PROMIS PF CAT: $r = 0.63$ in shoulder instability (287), $r = 0.70$ in rotator cuff pathology (282), $r = 0.69$ (baseline) and $r = 0.53$ (change) in upper extremity conditions (286), and $r = 0.48$ in hand and upper extremity conditions (276).
- PROMIS Pain Interference CAT: $r = –0.60$ in hand and upper extremity conditions (276).
- PROMIS Mobility CAT: $r = 0.41$ in hand and upper extremity conditions (276), and $r = –0.75$ in pediatric and adolescent patients in an ambulatory sports medicine clinic (291).
- PROMIS Depression CAT: $r^2 = –0.196$ (preoperative) and $r^2 = –0.431$ (postoperative) in patients undergoing arthroscopic rotator cuff repair (292), and $r = –0.21$ in pediatric and adolescent patients in an ambulatory sports medicine clinic (291).
- ASES: $r = 0.71$ in shoulder instability (287), $r = 0.77$ in rotator cuff pathology (282), $r = 0.55$ in total shoulder arthroplasty (281), $r = 0.72$ in shoulder pain (across conditions) (277), $r = 0.86$ in adhesive capsulitis (277), $r = 0.49$ in failed arthroplasty (277), $r = 0.87$ in fractures (277), $r = 0.55$ to 0.73 in instability (277,284), $r = 0.88$ impingement syndrome (277), $r = 0.65$ in rotator cuff disease (277), $r = 0.74$ in other shoulder pain (277), $r = 0.57$ in shoulder arthritis (278), and $r = 0.59$ in rotator cuff repair (285).
- WOSI: $r = 0.63$ in shoulder instability (287).
- Marx Shoulder Activity Scale: $r = 0.06$ in shoulder instability (287), $r = 0.23$ in rotator cuff pathology (282), and $r = 0.06$ in total shoulder arthroplasty (281).
- SF-36 Health Survey PF Subscale: shoulder instability: $r = 0.78$ (287).
- SF-36: $r = 0.66$ in rotator cuff pathology (282), and $r = 0.53$ in total shoulder arthroplasty (281).
- SF-36 Health Survey General Health: $r = 0.30$ in rotator cuff pathology (282).
- EQ-5D: $r = 0.66$ in shoulder instability (287), $r = 0.73$ in rotator cuff pathology (282), and $r = 0.48$ in total shoulder arthroplasty (281).
- WORC Index: $r = 0.73$ in rotator cuff pathology (282).
- Western Ontario Osteoarthritis Shoulder Index: $r = 0.34$ in total shoulder arthroplasty (281).

- SST: r = 0.82 in shoulder pain (across conditions) (277), r = 0.91 in adhesive capsulitis (277), r = 0.19 in failed arthroplasty (277), r = 0.93 in fractures (277), r = 0.81 in instability (277), r = 0.93 in impingement syndrome (277), r = 0.78 rotator cuff disease (277), r = 0.75 in other shoulder pain (277), r = 0.64 in shoulder arthritis (278), and r = 0.62 rotator cuff repair (285).
- Two-item Patient Health Questionnaire: r = –0.30 in hand and upper extremity conditions (276).
- Two-item Pain Self-Efficacy Questionnaire: r = 0.47 in hand and upper extremity conditions (276).
- Pain NRS: r = –0.59 in hand and upper extremity conditions (276).

Responsiveness. Several studies report good responsiveness with the SRM, responsiveness to change, and ESs in a variety of shoulder conditions. In patients with upper extremity trauma, one study reported good responsiveness to the treatment of the injury (289). This study also reports that ESs and SRMs were large for the measure for the full sample, for a subsample split by occurrence/nonoccurrence of a secondary treatment, and for a subsample split by fracture severity, although actual values are not reported in the published paper or in its supplemental material. Another study of patients with upper extremity conditions reported responsiveness to changes across time, and the magnitude of change was reported as mean = 6.1 (SD 5.8), which was comparable with an absolute mean difference of 0.80 (286). It also reported no differences in magnitude of change in those with hand conditions versus those with shoulder or elbow conditions. In patients undergoing arthroscopic rotator cuff repair, scores on the instrument were responsive to surgery, with scores improving by over 10 points at 6 months follow-up (292). In patients with shoulder instability, the instrument detected improvement at 6 months postoperatively, with medium to large effects sizes (Cohen's d = 1.09) with a corresponding SRM of 0.92 (284).

Minimally important differences. The minimally important difference was reported as 2.1 in one study of patients with nonshoulder upper extremity conditions (293).

Generalizability. This PROMIS instrument has been used in samples with various shoulder disorders, including general shoulder pain, upper extremity trauma (including fractures), rotator cuff disorders, instability, and shoulder arthritis before and after shoulder arthroplasty.

Use in clinical trials. All of the studies that have explored the measurement properties of this instrument were cohort studies. We could find no randomized clinical trials using the PROMIS PF Upper Extremity CAT.

Critical appraisal of overall value to the rheumatology community

Strengths. The PROMIS PF Upper Extremity CAT is short, uses item response theory, has very good evidence of validity and good evidence for responsiveness, has been used in varying shoulder conditions, and is free to use.

Caveats and cautions. The instrument appears to have serious ceiling effects, and there is only minimal evidence for reliability. Furthermore, many of the studies included above include mixed samples of patients with varying upper extremity disorders, and almost all studies do not delineate shoulder disorders from elbow or hand disorders in any subgroup analyses. In addition, more work should be done to establish the SDC and the MCID/minimally important difference in known groups.

Clinical usability. It may be useful in a selection of shoulder-related disorders. It is brief and easy to use and score, and appears to be valid and responsive.

Research usability. It is brief, easy to administer, and responsive. It is valid for intervention and population-level studies.

CONCLUSIONS

We reviewed nine of the at least 50 instruments that have been developed to measure adult shoulder function. Since the last review of some of these shoulder measures was published in a special issue of this journal in 2011 (3), there has been an enormous body of research evaluating their psychometric properties in patients with varying shoulder disorders and in different settings. Many tools have also been cross-culturally translated into multiple languages. Most have also been used in randomized controlled trials. This extensive review provides researchers with the necessary information for the eight instruments most commonly used in trials in the last 5 years as well as information for the PROMIS PF Upper Extremity CAT. See Tables 1 and 2 for a summary of the measures.

AUTHOR CONTRIBUTIONS

All authors drafted the article, revised it critically for important intellectual content, and approved the final version to be published.

REFERENCES

1. Green S, Buchbinder R, Glazier R, Forbes A. Systematic review of randomised controlled trials of interventions for painful shoulder: selection criteria, outcome assessment, and efficacy. BMJ 1998;316:354–60.
2. Page MJ, Huang H, Verhagen AP, Gagnier JJ, Buchbinder R. Outcome reporting in randomized trials for shoulder disorders: literature review to inform the development of a core outcome set. Arthritis Care Res (Hoboken) 2018;70:252–9.

3. Angst F, Schwyzer H-K, 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:S174–88.
4. Terwee CB, Jansma EP, Riphagen IR, de Vet HC. Development of a methodological PubMed search filter for finding studies on measurement properties of measurement instruments. Qual Life Res 2009;18:1115–23.
5. Roach KE, Budiman-Mak E, Songsiridej N, Lertratanakul Y. Development of a shoulder pain and disability index. Arthritis Care Res 1991;4:143–9.
6. Williams JW, Holleman DR, Simel DL. Measuring shoulder function with the Shoulder Pain and Disability Index. J Rheumatol 1995;22:727–32.
7. MacDermid JC, Solomon P, Prkachin K. The Shoulder Pain and Disability Index demonstrates factor, construct and longitudinal validity. BMC Musculoskelet Disord 2006;7:12.
8. Hill CL, Lester S, Taylor AW, Shanahan ME, Gill TK. Factor structure and validity of the shoulder pain and disability index in a population-based study of people with shoulder symptoms. BMC Musculoskelet Disord 2011;12:8.
9. McLean JM, Awwad D, Lisle R, Besanko J, Shivakkumar D, Leith J. An international, multicenter cohort study comparing 6 shoulder clinical scores in an asymptomatic population. J Shoulder Elbow Surg 2018;27:306–14.
10. Paul A, Lewis M, Shadforth MF, Croft PR, van der Windt DA, Hay EM. A comparison of four shoulder-specific questionnaires in primary care. Ann Rheum Dis 2004;63:1293–9.
11. Membrilla-Mesa MD, Cuesta-Vargas AI, Pozuelo-Calvo R, Tejero-Fernández V, Martín-Martín L, Arroyo-Morales M. Shoulder pain and disability index: cross cultural validation and evaluation of psychometric properties of the Spanish version. Health Qual Life Outcomes 2015;13:200.
12. Wang W, Jia Z-Y, Liu J, Xie Q-Y, Cui J, Zheng W, et al. Cross-cultural adaptation and validation of the Chinese version of the shoulder pain and disability index in patients with symptomatic shoulder pain: a prospective case series. Medicine (Baltimore) 2018;97:e11227.
13. Alsanawi HA, Alghadir A, Anwer S, Roach KE, Alawaji A. Cross-cultural adaptation and psychometric properties of an Arabic version of the Shoulder Pain and Disability Index. Int J Rehabil Res 2015;38:270–5.
14. Christiansen DH, Andersen JH, Haahr JP. Cross-cultural adaption and measurement properties of the Danish version of the Shoulder Pain and Disability Index. Clin Rehabil 2013;27:355–60.
15. Ekeberg OM, Bautz-Holter E, Tveita EK, Keller A, Juel NG, Brox JI. Agreement, reliability and validity in 3 shoulder questionnaires in patients with rotator cuff disease. BMC Musculoskelet Disord 2008;9:68.
16. Thoomes-de Graaf M, Scholten-Peeters GG, Duijn E, Karel Y, Koes BW, Verhagen AP. The Dutch Shoulder Pain and Disability Index (SPADI): a reliability and validation study. Qual Life Res 2015;24:1515–9.
17. Jeldi AJ, Aseer AL, Dhandapani AG, Roach KE. Cross-cultural adaption, reliability and validity of an Indian (Tamil) version for the Shoulder Pain and Disability Index. Hong Kong Physiotherapy Journal 2012;30:99–104.
18. Sharma N, Sharma S, Kataria C. Translation and adaptation of Shoulder Pain and Disability Index (SPADI) into Hindi. Part 1. Indian Journal of Physiotherapy & Occupational Therapy 2011;5:142–5.
19. Vrouva S, Batistaki C, Koutsioumpa E, Kostopoulos D, Stamoulis E, Kostopanagiotou G. The Greek version of Shoulder Pain and Disability Index (SPADI): translation, cultural adaptation, and validation in patients with rotator cuff tear. J Orthop Traumatol 2016;17:315–26.
20. Bumin G, Tüzün EH, Tonga E. The Shoulder Pain and Disability Index (SPADI): Cross-cultural adaptation, reliability, and validity of the Turkish version. BMR J Aust Geol Geophys 2008;21:57–62.
21. Martins J, Napoles BV, Hoffman CB, Oliveira AS. The Brazilian version of Shoulder Pain and Disability Index: translation, cultural adaptation and reliability. Rev Bras Fisioter 2010;14:527–36.
22. Ebrahimzadeh MH, Birjandinejad A, Golhasani F, Moradi A, Vahedi E, Kachooei AR. Cross-cultural adaptation, validation, and reliability testing of the Shoulder Pain and Disability Index in the Persian population with shoulder problems. Int J Rehabil Res 2015;38:84–7.
23. Phongamwong C, Choosakde A. Reliability and validity of the Thai version of the Shoulder Pain and Disability Index (Thai SPADI). Health Qual Life Outcomes 2015;13:136.
24. Kc S, Sharma S, Ginn K, Almadi T, Reed D. Nepali translation, cross-cultural adaptation and measurement properties of the Shoulder Pain and Disability Index (SPADI). J Orthop Surg Res 2019;14:284.
25. Marchese C, Cristalli G, Pichi B, Manciocco V, Mercante G, Pellini R, et al. Italian cross-cultural adaptation and validation of three different scales for the evaluation of shoulder pain and dysfunction after neck dissection: University of California - Los Angeles (UCLA) Shoulder Scale, Shoulder Pain and Disability Index (SPADI) and Simple Shoulder Test (SST). Acta Otorhinolaryngol Ital 2012;32:12–7.
26. Angst F, Goldhahn J, Pap G, Mannion AF, Roach KE, Siebertz D, et al. Cross-cultural adaptation, reliability and validity of the German Shoulder Pain and Disability Index (SPADI). Rheumatology (Oxford) 2007;46:87–92.
27. Ekeberg OM, Bautz-Holter E, Keller A, Tveita EK, Juel NG, Brox JI. A questionnaire found disease-specific WORC index is not more responsive than SPADI and OSS in rotator cuff disease. J Clin Epidemiol 2010;63:575–84.
28. Angst F, Pap G, Mannion AF, Herren DB, Aeschlimann A, Schwyzer H-K, et al. Comprehensive assessment of clinical outcome and quality of life after total shoulder arthroplasty: usefulness and validity of subjective outcome measures. Arthritis Rheum 2004;51:819–28.
29. Roddey TS, Olson SL, Cook KF, Gartsman GM, Hanten W. Comparison of the University of California-Los Angeles Shoulder Scale and the Simple Shoulder Test with the Shoulder Pain and Disability Index: single-administration reliability and validity. Phys Ther 2000;80:759–68.
30. Torres-Lacomba M, Sanchez-Sanchez B, Prieto-Gomez V, Pacheco-da-Costa S, Yuste-Sanchez MJ, Navarro-Brazalez B, et al. Spanish cultural adaptation and validation of the Shoulder Pain and Disability Index, and the Oxford Shoulder Score after breast cancer surgery. Health Qual Life Outcomes 2015;13:63.
31. Beaton D, Richards RR. Assessing the reliability and responsiveness of 5 shoulder questionnaires. J Shoulder Elbow Surg 1998;7:565–72.
32. Tveitå EK, Ekeberg OM, Juel NG, Bautz-Holter E. Responsiveness of the shoulder pain and disability index in patients with adhesive capsulitis. BMC Musculoskelet Disord 2008;9:161.
33. Dawson J, Harris KK, Doll H, Fitzpatrick R, Carr A. A comparison of the Oxford shoulder score and shoulder pain and disability index: factor structure in the context of a large randomized controlled trial. Patient Relat Outcome Meas 2016;7:195–203.

34. Jerosch-Herold C, Chester R, Shepstone L, Vincent JI, MacDermid JC. An evaluation of the structural validity of the Shoulder Pain and Disability Index (SPADI) using the Rasch model. Qual Life Res 2018;27:389–400.
35. Staples MP, Forbes A, Green S, Buchbinder R. Shoulder-specific disability measures showed acceptable construct validity and responsiveness. J Clin Epidemiol 2010;63:163–70.
36. Beaton DE, Richards RR. Measuring function of the shoulder: a cross-sectional comparison of five questionnaires. J Bone Joint Surg Am 1996;78:882–90.
37. Christie A, Hagen KB, Mowinckel P, Dagfinrud H. Methodological properties of six shoulder disability measures in patients with rheumatic diseases referred for shoulder surgery. J Shoulder Elbow Surg 2009;18:89–95.
38. Placzek JD, Lukens SC, Badalanmenti S, Roubal PJ, Freeman DC, Walleman KM, et al. Shoulder outcome measures: a comparison of 6 functional tests. Am J Sports Med 2004;32:1270–7.
39. Cloke DJ, Lynn SE, Watson H, Steen IN, Purdy S, Williams JR. A comparison of functional, patient-based scores in subacromial impingement. J Shoulder Elbow Surg 2005;14:380–4.
40. Tuğay U, Tuğay N, Gelecek N, Özkan M. Oxford Shoulder Score: cross-cultural adaptation and validation of the Turkish version. Arch Orthop Trauma Surg 2011;131:687–94.
41. Heald SL, Riddle DL, Lamb RL. The shoulder pain and disability index: the construct validity and responsiveness of a region-specific disability measure. Phys Ther 1997;77:1079–89.
42. MacDermid JC, Drosdowech D, Faber K. Responsiveness of self-report scales in patients recovering from rotator cuff surgery. J Shoulder Elbow Surg 2006;15:407–14.
43. Chester R, Jerosch-Herold C, Lewis J, Shepstone L. The SPADI and QuickDASH are similarly responsive in patients undergoing physical therapy for shoulder pain. J Orthop Sports Phys Ther 2017;47:538–47.
44. Angst F, Goldhahn J, Drerup S, Aeschlimann A, Schwyzer H-K, Simmen BR. Responsiveness of six outcome assessment instruments in total shoulder arthroplasty. Arthritis Rheum 2008;59:391–8.
45. Schmitt JS, di Fabio RP. Reliable change and minimum important difference (MID) proportions facilitated group responsiveness comparisons using individual threshold criteria. J Clin Epidemiol 2004;57:1008–18.
46. Kim G-W, Won YH, Park S-H, Seo J-H, Kim D-H, Lee HN, et al. Effects of a newly developed therapeutic deep heating device using high frequency in patients with shoulder pain and disability: a pilotsStudy. Pain Res Manag 2019;2019:8215371.
47. Littlewood C, Malliaras P, Mawson S, May S, Walters SJ. Self-managed loaded exercise versus usual physiotherapy treatment for rotator cuff tendinopathy: a pilot randomised controlled trial. Physiotherapy 2014;100:54–60.
48. Mintken PE, McDevitt AW, Cleland JA, Boyles RE, Beardslee AR, Burns SA, et al. Cervicothoracic manual therapy plus exercise therapy versus exercise therapy alone in the management of individuals with shoulder pain: a multicenter randomized controlled trial. J Orthop Sports Phys Ther 2016;46:617–28.
49. Marks D, Bisset L, Comans T, Thomas M, Ng SK, O'Leary S, et al. Increasing capacity for the treatment of common musculoskeletal problems: a non-inferiority RCT and economic analysis of corticosteroid injection for shoulder pain comparing a physiotherapist and orthopaedic surgeon. PLoS One 2016;11:e0162679.
50. Kim K-H, Suh J-W, Oh KY. The effect of intra-articular hyaluronate and tramadol injection on patients with adhesive capsulitis of the shoulder. J Back Musculoskelet Rehabil 2017;30:913–20.
51. Kvalvaag E, Roe C, Engebretsen KB, Soberg HL, Juel NG, Bautz-Holter E, et al. one year results of a randomized controlled trial on radial extracorporeal shock wave treatment, with predictors of pain, disability and return to work in patients with subacromial pain syndrome. Eur J Phys Rehabil Med 2018;54:341–50.
52. Richards RR, An KN, Bigliani LU, Friedman RJ, Gartsman GM, Gristina AG, et al. A standardized method for the assessment of shoulder function. J Shoulder Elbow Surg 1994;3:347–52.
53. Sallay PI, Reed L. The measurement of normative American Shoulder and Elbow Surgeons scores. J Shoulder Elbow Surg 2003;12:622–7.
54. Leggin BG, Michener LA, Shaffer MA, Brenneman S, Iannotti J, Williams G Jr. The Penn Shoulder Score: reliability and validity. J Orthop Sports Phys Ther 2006;36:138–51.
55. Goldhahn J, Angst F, Drerup S, Pap G, Simmen BR, Mannion AF. Lessons learned during the cross-cultural adaptation of the American Shoulder and Elbow Surgeons shoulder form into German. J Shoulder Elbow Surg 2008;17:248–54.
56. Padua R, Padua L, Ceccarelli E, Bondi R, Alviti F, Castagna A. Italian version of ASES questionnaire for shoulder assessment: cross-cultural adaptation and validation. Musculoskelet Surg 2010;94 Suppl 1:S85–90.
57. Knaut LA, Moser AD, Melo SD, Richards RR. Translation and cultural adaptation to the Portuguese language of the American Shoulder and Elbow Surgeons Standardized Shoulder assessment form (ASES) for evaluation of shoulder function. Rev Bras Reumatol 2010;50:176–89.
58. Moser AD, Knaut LA, Zotz TG, Scharan KO. Validity and reliability of the Portuguese version of the American Shoulder and Elbow Surgeons Standardized Shoulder Assessment Form. Rev Bras Reumatol 2012;52:348–56.
59. Vrotsou K, Cuéllar R, Silió F, Ángel Rodriguez MA, Garay D, Busto G, et al. Patient self-report section of the ASES questionnaire: a Spanish validation study using classical test theory and the Rasch model. Health Qual Life Outcomes 2016;14:147.
60. Piitulainen K, Paloneva J, Ylinen J, Kautiainen H, Häkkinen A. Reliability and validity of the Finnish version of the American Shoulder and Elbow Surgeons Standardized Shoulder Assessment Form, patient self-report section. BMC Musculoskelet Disord 2014;15:272.
61. Çelik D, Atalar AC, Demirhan M, Dirican A. Translation, cultural adaptation, validity and reliability of the Turkish ASES questionnaire. Knee Surg Sports Traumatol Arthrosc 2013;21:2184–9.
62. Yahia A, Guermazi M, Khmekhem M, Ghroubi S, Ayedi K, Elleuch MH. Translation into Arabic and validation of the ASES index in assessment of shoulder disabilities. Ann Phys Rehabil Med 2011;54:59–72.
63. Kocher MS, Horan MP, Briggs KK, Richardson TR, O'Holleran J, Hawkins RJ. Reliability, validity, and responsiveness of the American Shoulder and Elbow Surgeons subjective shoulder scale in patients with shoulder instability, rotator cuff disease, and glenohumeral arthritis. J Bone Joint Surg Am 2005;87:2006–11.
64. Beckmann JT, Hung M, Bounsanga J, Wylie JD, Granger EK, Tashjian RZ. Psychometric evaluation of the PROMIS Physical Function computerized adaptive test in comparison to the American Shoulder and Elbow Surgeons score and Simple Shoulder Test in patients with rotator cuff disease. J Shoulder Elbow Surg 2015;24:1961–7.
65. Thigpen CA, Shanley E, Momaya AM, Kissenberth MJ, Tolan SJ, Tokish JM, et al. Validity and responsiveness of the Single Alpha-numeric Evaluation for shoulder patients. Am J Sports Med 2018;46:3480–5.
66. Strong B, Maloney M, Baumhauer J, Schaffer J, Houck JR, Hung M, et al. Psychometric evaluation of the Patient-Reported Outcomes Measurement Information System (PROMIS) Physical Function and Pain Interference Computer Adaptive Test for

subacromial impingement syndrome. J Shoulder Elbow Surg 2019;28:324–9.
67. Michener LA, McClure PW, Sennett BJ. American Shoulder and Elbow Surgeons standardized shoulder assessment form, patient self-report section: reliability, validity, and responsiveness. J Shoulder Elbow Surg 2002;11:587–94.
68. Cook KF, Roddey TS, Olson SL, Gartsman GM, Valenzuela FF, Hanten WP. Reliability by surgical status of self-reported outcomes in patients who have shoulder pathologies. J Orthop Sports Phys Ther 2002;32:336–46.
69. Oh JH, Jo KH, Kim WS, Gong YS, Han SG, Kim YH. Comparative evaluation of the measurement properties of various shoulder outcome instruments. Am J Sports Med 2009;37:1161–8.
70. Michael RJ, Williams BA, Laguerre MD, Struk AM, Schoch BS, Wright TW, et al. Correlation of multiple patient-reported outcome measures across follow-up in patients undergoing primary shoulder arthroplasty. J Shoulder Elbow Surg 2019;28:1869–76.
71. Razmjou H, Bean A, van Osnabrugge V, MacDermid JC, Holtby R. Cross-sectional and longitudinal construct validity of two rotator cuff disease-specific outcome measures. BMC Musculoskelet Disord 2006;7:26.
72. Skutek M, Fremerey RW, Zeichen J, Bosch U. Outcome analysis following open rotator cuff repair. Early effectiveness validated using four different shoulder assessment scales. Arch Orthop Trauma Surg 2000;120:432–6.
73. Assuncao JH, Malavolta EA, Gracitelli ME, Hiraga DY, da Silva FR, Ferreira Neto AA. Clinical outcomes of arthroscopic rotator cuff repair: correlation between the University of California, Los Angeles (UCLA) and American Shoulder and Elbow Surgeons (ASES) scores. J Shoulder Elbow Surg 2017;26:1137–42.
74. Cunningham G, Ladermann A, Denard PJ, Kherad O, Burkhart SS. Correlation between American Shoulder and Elbow Surgeons and Single Assessment Numerical Evaluation Score after rotator cuff or SLAP repair. Arthroscopy 2015;31:1688–92.
75. Patterson BM, Orvets ND, Aleem AW, Keener JD, Calfee RP, Nixon DC, et al. Correlation of Patient-Reported Outcomes Measurement Information System (PROMIS) scores with legacy patient-reported outcome scores in patients undergoing rotator cuff repair. J Shoulder Elbow Surg 2018;27:S17–23.
76. Minoughan CE, Schumaier AP, Fritch JL, Grawe BM. Correlation of Patient-Reported Outcome Measurement Information System Physical Function Upper Extremity computer adaptive testing, with the American Shoulder and Elbow Surgeons Shoulder Assessment Form and Simple Shoulder Test in patients with shoulder pain. Arthroscopy 2018;34:1430–6.
77. Yoo JC, Koh KH, Park WH, Park JC, Kim SM, Yoon YC. The outcome of ultrasound-guided needle decompression and steroid injection in calcific tendinitis. J Shoulder Elbow Surg 2010;19:596–600.
78. Tashjian RZ, Deloach J, Green A, Porucznik CA, Powell AP. Minimal clinically important differences in ASES and simple shoulder test scores after nonoperative treatment of rotator cuff disease. J Bone Joint Surg Am 2010;92:296–303.
79. Tashjian RZ, Hung M, Keener JD, Bowen RC, McAllister J, Chen W, et al. Determining the minimal clinically important difference for the American Shoulder and Elbow Surgeons score, Simple Shoulder Test, and visual analog scale (VAS) measuring pain after shoulder arthroplasty. J Shoulder Elbow Surg 2017;26:144–8.
80. Werner BC, Chang B, Nguyen JT, Dines DM, Gulotta LV. What change in American Shoulder and Elbow Surgeons Score represents a clinically important change after shoulder arthroplasty? [Measurement study]. Clin Orthop Relat Res 2016;474:2672–81.
81. Wong SE, Zhang AL, Berliner JL, Ma CB, Feeley BT. Preoperative patient-reported scores can predict postoperative outcomes after shoulder arthroplasty. J Shoulder Elbow Surg 2016;25:913–9.
82. Cvetanovich GL, Gowd AK, Liu JN, Nwachukwu BU, Cabarcas BC, Cole BJ, et al. Establishing clinically significant outcome after arthroscopic rotator cuff repair. J Shoulder Elbow Surg 2019;28:939–48.
83. Gagnier JJ, Robbins C, Bedi A, Carpenter JE, Miller BS. Establishing minimally important differences for the American Shoulder and Elbow Surgeons score and the Western Ontario Rotator Cuff Index in patients with full-thickness rotator cuff tears. J Shoulder Elbow Surg 2018;27:e160–6.
84. Nicholas SJ, Lee SJ, Mullaney MJ, Tyler TF, Fukunaga T, Johnson CD, et al. Functional outcomes after double-row versus single-row rotator cuff repair: a prospective randomized trial. Orthop Journal Sports Med 2016;4:2325967116667398.
85. Zhu QX, Jiang Y, Wang H, He Y. Comparison of intra-articular injection and arthroscopic capsular release in stages I and II of primary frozen shoulder, a randomized clinical trial. International Journal of Gerontology 2019;13:81–6.
86. Yoo JC, Park YB, Kum DH, Park KM, Lee ES. Arthroscopic distal clavicle resection of symptomatic acromioclavicular joint arthritis combined with rotator cuff tear: a prospective randomized trial. Arthroscopy 2015;31:e7.
87. Waterman BR, Newgren J, Gowd A, Cabarcas B, Bach BR, Cole BJ, et al. Randomized prospective trial of arthroscopic rotator cuff with or without acromioplasty: no difference in patient-reported outcomes at long-term follow-up. Arthroscopy 2018;34:e3–4.
88. Park YB, Koh KH, Shon MS, Park YE, Yoo JC. Arthroscopic distal clavicle resection in symptomatic acromioclavicular joint arthritis combined with rotator cuff tear: a prospective randomized trial. Am J Sports Med 2015;43:985–90.
89. Weekes DG, Wicks ED, Campbell RE, Hadley C, Carter A, Chaudhry Z, et al. Do relaxation exercises decrease postoperative pain after rotator cuff repair? A randomized controlled trial. Orthop J Sports Med 2019;7 Suppl 5:2325967119S00380.
90. Von Wehren L, Blanke F, Todorov A, Heisterbach P, Sailer J, Majewski M. The effect of subacromial injections of autologous conditioned plasma versus cortisone for the treatment of symptomatic partial rotator cuff tears. Knee Surg Sports Traumatol Arthrosc 2016;24:3787–92.
91. Shams A, El-Sayed M, Gamal O, Ewes W. Subacromial injection of autologous platelet-rich plasma versus corticosteroid for the treatment of symptomatic partial rotator cuff tears. Eur J Orthop Surg Traumatol 2016;26:837–42.
92. Shishani Y, Lederman ES, Denard PJ, Gobezie R. Can a functional difference be detected in reverse arthroplasty with 135° vs. 155° prosthesis for the treatment of rotator cuff arthropathy: a prospective randomized study. J Shoulder Elbow Surg. 2019;28:813–8.
93. Sabesan VJ, Shahriar R, Petersen-Fitts G, Whaley JD, Bou-Akl T, Sweet M, et al. A prospective randomized trial to identify the optimal postoperative pain management in shoulder arthroplasty: liposomal bupivacaine vs. continuous peripheral nerve block. Journal of shoulder and elbow surgery. 2017;26:1810–7.
94. Ranalletta M, Rossi LA, Bongiovanni SL, Tanoira I, Elizondo CM, Maignon GD. Corticosteroid injections accelerate pain relief and recovery of function compared with oral NSAIDs in patients with adhesive capsulitis: a randomized controlled trial. Am J Sports Med 2016;44:474–81.
95. Piitulainen K, Hakkinen A, Salo P, Kautiainen H, Ylinen J. Does adding a 12-month exercise programme to usual care after a rotator cuff repair effect disability and quality of life at 12 months? A randomized controlled trial. Clin Rehabil 2015;29:447–56.
96. Perdreau A, Joudet T. Efficacy of multimodal analgesia injection combined with corticosteroids after arthroscopic rotator cuff repair. Orthop Traumatol Surg Res 2015;101 Suppl:S337–45.

97. Park JS, Kim SH, Jung HJ, Lee YH, Oh JH. A prospective randomized study comparing the interference screw and suture anchor techniques for biceps tenodesis. Am J Sports Med 2017;45:440–8.
98. Oh JH, Park JS, Rhee SM, Park JH. Maximum bridging suture tension provides better clinical outcomes in transosseous-equivalent rotator cuff repair: a clinical, prospective randomized comparative study. Am J Sports Med 2019;8:2129–36.
99. Pandey V, Bandi A, Madi S, Agarwal L, Acharya KK, Maddukuri S, et al. Does application of moderately concentrated platelet-rich plasma improve clinical and structural outcome after arthroscopic repair of medium-sized to large rotator cuff tear? A randomized controlled trial. J Shoulder Elbow Surg 2016;25:1312–22.
100. Owens BD, Cameron KL, Peck KY, DeBerardino TM, Nelson BJ, Taylor DC, et al. Arthroscopic versus open stabilization for anterior shoulder subluxations. Orthop J Sports Med 2015;3:1–4.
101. Kim YS, Lee HJ, Lee DH, Choi KY. Comparison of high- and low-dose intra-articular triamcinolone acetonide injection for treatment of primary shoulder stiffness: a prospective randomized trial. J Shoulder Elbow Surg 2017;26:209-15.
102. Barber FA. Triple-loaded single-row versus suture-bridge double-row rotator cuff tendon repair with platelet-rich plasma fibrin membrane: a randomized controlled trial. Arthroscopy 2016;32:753–61.
103. Baumgarten KM, Chang PS, Foley EK. Patient-determined outcomes after arthroscopic rotator cuff repair with and without biceps tenodesis utilizing the PITT technique. J Shoulder Elbow Surg 2019;28:1049–55.
104. Baumgarten KM, Osborn R, Schweinle WE, Zens MJ, Helsper EA. Are pulley exercises initiated 6 weeks after rotator cuff repair a safe and effective rehabilitative treatment? a randomized controlled trial. Am J Sports Med 2016;44:1844–51.
105. Bayle X, Pham TT, Faruch M, Gobet A, Mansat P, Bonnevialle N. No difference in outcome for open versus arthroscopic rotator cuff repair: a prospective comparative trial. Arch Orthop Trauma Surg 2017;137:1707–12.
106. Belay ES, Wittstein JR, Garrigues GE, Lassiter TE, Scribani M, Goldner RD, et al. Biceps tenotomy has earlier pain relief compared to biceps tenodesis: a randomized prospective study. Knee Surg Sports Traumatol Arthrosc 2019;27:4032–7.
107. Boyer P, Dukan R. Displaced 3- and 4-Part Proximal Humeral Fractures: a Prospective Randomized Study Comparing Intramedullary Nail and Locking Plate. American Academy of Orthopaedic Surgeons 2019. URL: https://aaos2019.conferencespot.org/#/event/event-data/details.
108. Dedeoglu SS, Imren Y, Cabuk H, Tekin AC, Kir MC, Gurbuz H. Arthroscopy-assisted versus standard intramedullary nail fixation in diaphyseal fractures of the humerus. J Orthop Surg (Hong Kong) 2017;25:2309499017727949.
109. Dezfuli B, King JJ, Farmer KW, Struk AM, Wright TW. Outcomes of reverse total shoulder arthroplasty as primary versus revision procedure for proximal humerus fractures. J Shoulder Elbow Surg 2016;25:1133–7.
110. Dilek B, Gulbahar S, Gundogdu M, Ergin B, Manisali M, Ozkan M, et al. Efficacy of proprioceptive exercises in patients with subacromial impingement syndrome: a single-blinded randomized controlled study. Am J Phys Med Rehabil 2016;95:169–82.
111. Flury M, Rickenbacher D, Schwyzer HK, Jung C, Schneider MM, Stahnke K, et al. Does pure platelet-rich plasma affect postoperative clinical outcomes after arthroscopic rotator cuff repair? A randomized controlled trial. Am J Sports Med 2016;44:2136–46.
112. Forsythe B, Zuke W, Go B, Yanke A, Verma N, Romeo A. Randomized prospective analysis of arthroscopic suprapectoral and open sub-pectoral biceps tenodesis: 1 year follow-up. Orthop J Sports Med 2017;5 Suppl 6:2325967117S00211.
113. Hart R, Smid P. Tears of the supraspinatus muscle sutured with added mesenchymal stem cells. Osteoporosis International 2018;29:S153.
114. Helfen T, Siebenburger G, Mayer M, Bocker W, Ockert B, Haasters F. Operative treatment of 2-part surgical neck fractures of the proximal humerus (AO 11–A3) in the elderly: cement augmented locking plate Philos™ vs. proximal humerus nail MultiLoc®. BMC Musculoskeletal Disord 2016;17:448.
115. Hwang JT, Yang CJ, Noh K-C, Yoo Y-S, Hyun YS, Lee YB, et al. Which is better for arthroscopic tenodesis of the long head of the biceps: soft tissue or bony interference fixation? [retrospective comparative study]. Arthroscopy 2016;32:560–7.
116. Kim JH, Kim DJ, Lee HJ, Kim BK, Kim YS. Atelocollagen injection improves tendon integrity in partial-thickness rotator cuff tears: A prospective comparative study. Orthop J Sports Med 2020;8:232596712090401.
117. Ma HH, Chen KH, Chiang ER, Chou TF, Ma HL. Does arthroscopic suture-spanning augmentation of single-row repair reduce the retear rate of massive rotator cuff tear? [RCT]. Am J Sports Med 2019;47:1420–6.
118. Macdonald P, McRae S, Stranges G, Old J, Dubberley J, Mascarenhas R, et al. Biceps tenodesis versus tenotomy in treatment of lesions of long head of biceps brachii in patients undergoing arthroscopic shoulder surgery. Arthroscopy 2017;33:e12–3.
119. MacDonald PB, Old J, Mascarenhas R, McRae S, Marsh J, Dubberley J, et al. Arthroscopic bankart repair with and without arthroscopic infraspinatus remplissage in anterior shoulder instability with hill-sachs defect: randomized controlled trial. Orthop. J Sports Med 2018;6 Suppl 4:2325967118S00095.
120. McRae S, Leiter J, Subramanian K, Litchfield R, MacDonald P. Randomized controlled trial of arthroscopic electrothermal capsulorrhaphy with Bankart repair and isolated arthroscopic Bankart repair. Knee Surg Sports Traumatol Arthrosc 2016;24:414–21.
121. Constant CR, Murley AH. A clinical method of functional assessment of the shoulder. Clin Orthop Relat Res 1987:160–4.
122. Constant CR, Gerber C, Emery RJH, Søjbjerg JO, Gohlke F, Boileau P. A review of the Constant score: modifications and guidelines for its use. J Shoulder Elbow Surg 2008;17:355–61.
123. Chelli M, Levy Y, Lavoué V, Clowez G, Gonzalez J-F, Boileau P. The, "Auto-Constant": Can we estimate the Constant-Murley score with a self-administered questionnaire? A pilot study. Orthop Traumatol Surg Res 2019;105:251–6.
124. Levy O, Haddo O, Massoud S, Mullett H, Atoun E. A patient-derived Constant-Murley score is comparable to a clinician-derived score. Clin Orthop Relat Res 2014;472:294–303.
125. Ban I, Troelsen A, Christiansen DH, Svendsen SW, Kristensen MT. Standardised test protocol (Constant Score) for evaluation of functionality in patients with shoulder disorders. Dan Med J 2013;60:A4608.
126. Yian EH, Ramappa AJ, Arneberg O, Gerber C. The Constant score in normal shoulders. J Shoulder Elbow Surg 2005;14:128–33.
127. Constant CR. Assessment of shoulder function [article in German]. Orthopade 1991;20:289–94.
128. Moeller AD, Thorsen RR, Torabi TP, Bjoerkman A-S, Christensen EH, Maribo T, et al. The Danish version of the modified Constant-Murley shoulder score: reliability, agreement, and construct validity. J Orthop Sports Phys Ther 2014;44:336–40.
129. Çelik D. Turkish version of the modified Constant-Murley score and standardized test protocol: reliability and validity. Acta Orthop Traumatol Turc 2016;50:69–75.
130. Barreto RPG, Barbosa MLL, Balbinotti MAA, Mothes FC, da Rosa LHT, Silva MF. The Brazilian version of the Constant-Murley Score

(CMS-BR): convergent and construct validity, internal consistency, and unidimensionality. Rev Bras Ortop 2016;51:515–20.

131. Ntourantonis D, Panagopoulos A, Iliopoulos I, Tatani I, Tsoumpos P, Kouzelis A, et al. Translation, cultural adaptation, validity and reliability of the Greek version of the modified Constant Score. JSES Open Access 2017;1:45–50.
132. Yao M, Yang L, Cao Z-Y, Cheng S-D, Tian S-L, Sun Y-L, et al. Chinese version of the Constant-Murley questionnaire for shoulder pain and disability: a reliability and validation study. Health Qual Life Outcomes 2017;15:178.
133. Henseler JF, Kolk A, van der Zwaal P, Nagels J, Vliet Vlieland TP, Nelissen RG. The minimal detectable change of the Constant score in impingement, full-thickness tears, and massive rotator cuff tears. J Shoulder Elbow Surg 2015;24:376–81.
134. Sciascia AD, Morris BJ, Jacobs CA, Edwards TB. Responsiveness and internal validity of common patient-reported outcome measures following total shoulder arthroplasty. Orthopedics 2017;40:e513–9.
135. Razmjou H, Bean A, Macdermid JC, van Osnabrugge V, Travers N, Holtby R. Convergent validity of the constant-murley outcome measure in patients with rotator cuff disease. Physiother Can 2008;60:72–9.
136. James-Belin E, Roy AL, Lasbleiz S, Ostertag A, Yelnik A, Orcel P, et al. Comparative study of psychometric properties of three assessment tools for degenerative rotator cuff disease. Clin Rehabil 2019;33:277–84.
137. Van Den Ende CH, Rozing PM, Dijkmans BA, Verhoef JA, Voogt-van der Harst EM, Hazes JM. Assessment of shoulder function in rheumatoid arthritis. J Rheumatol 1996;23:2043–8.
138. Johansson KM, Adolfsson LE. Intraobserver and interobserver reliability for the strength test in the Constant-Murley shoulder assessment. J Shoulder Elbow Surg 2005;14:273–8.
139. 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;45:620–5.
140. Roe Y, Soberg HL, Bautz-Holter E, Ostensjo S. A systematic review of measures of shoulder pain and functioning using the International classification of functioning, disability and health (ICF). BMC Musculoskelet Disord 2013;14:73.
141. Ziegler P, Kühle L, Stöckle U, Wintermeyer E, Stollhof LE, Ihle C, et al. Evaluation of the Constant score: which is the method to assess the objective strength? [measurement study]. BMC Musculoskelet Disord 2019;20:403.
142. Dupeyron A, Gelis A, Sablayrolles P, Bousquet P-J, Julia M, Herisson C, et al. Heterogeneous assessment of shoulder disorders: validation of the Standardized Index of Shoulder Function. J Rehabil Med 2010;42:967–72.
143. Van de Water AT, Davidson M, Shields N, Evans MC, Taylor NF. The Shoulder Function Index (SFInX): evaluation of its measurement properties in people recovering from a proximal humeral fracture. BMC Musculoskelet Disord 2016;17:295.
144. Dawson J, Hill G, Fitzpatrick R, Carr A. The benefits of using patient-based methods of assessment: medium-term results of an observational study of shoulder surgery. J Bone Joint Surg Br 2001;83:877–82.
145. Van de Water AT, Shields N, Davidson M, Evans M, Taylor NF. Reliability and validity of shoulder function outcome measures in people with a proximal humeral fracture. Disabil Rehabil 2014;36:1072–9.
146. Romeo AA, Mazzocca A, Hang DW, Shott S, Bach BR. Shoulder scoring scales for the evaluation of rotator cuff repair. Clin Orthop Relat Res 2004:107–14.
147. 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;26:764–72.
148. De Witte PB, Henseler JF, Nagels J, Vliet Vlieland TP, Nelissen RG. The Western Ontario rotator cuff index in rotator cuff disease patients: a comprehensive reliability and responsiveness validation study. Am J Sports Med 2012;40:1611–9.
149. Holtby R, Razmjou H. Measurement properties of the Western Ontario rotator cuff outcome measure: a preliminary report. J Shoulder Elbow Surg 2005;14:506–10.
150. Vas J, Ortega C, Olmo V, Perez-Fernandez F, Hernandez L, Medina I, et al. Single-point acupuncture and physiotherapy for the treatment of painful shoulder: a multicentre randomized controlled trial. Rheumatology (Oxford) 2008;47:887–93.
151. O'Connor DA, Chipchase LS, Tomlinson J, Krishnan J. Arthroscopic subacromial decompression: responsiveness of disease-specific and health-related quality of life outcome measures. Arthroscopy 1999;15:836–40.
152. Torrens C, Guirro P, Santana F. The minimal clinically important difference for function and strength in patients undergoing reverse shoulder arthroplasty. J Shoulder Elbow Surg 2016;25:262–8.
153. Simovitch R, Flurin P-H, Wright T, Zuckerman JD, Roche CP. Quantifying success after total shoulder arthroplasty: the minimal clinically important difference. J Shoulder Elbow Surg 2018;27:298–305.
154. Kukkonen J, Kauko T, Vahlberg T, Joukainen A, Äärimaa V. Investigating minimal clinically important difference for Constant score in patients undergoing rotator cuff surgery. J Shoulder Elbow Surg 2013;22:1650–5.
155. Holmgren T, Oberg B, Adolfsson L, Björnsson Hallgren H, Johansson K. Minimal important changes in the Constant-Murley score in patients with subacromial pain. J Shoulder Elbow Surg 2014;23:1083–90.
156. Jenssen KK, Lundgreen K, Madsen JE, Kvakestad R, Pripp AH, Dimmen S. No functional difference between three and six weeks of immobilization after arthroscopic rotator cuff repair: a prospective randomized controlled non-inferiority trial. Arthroscopy 2018;34:2765–74.
157. Launonen AP, Sumrein BO, Reito A, Lepola V, Paloneva J, Jonsson KB, et al. Operative versus non-operative treatment for 2-part proximal humerus fracture: A multicenter randomized controlled trial. PLoS Med 2019;16:e1002855.
158. Rasmussen JV, Olsen BS, Sorensen AK, Hróbjartsson A, Brorson S. Resurfacing hemiarthroplasty compared to stemmed hemiarthroplasty for glenohumeral osteoarthritis: a randomised clinical trial. Int Orthop 2015;39:263–9.
159. Kim G-W, Won YH, Park S-H, Seo J-H, Kim D-H, Lee HN, et al. Effects of a newly developed therapeutic deep heating device using high frequency in patients with shoulder pain and disability: a pilot study. Pain Res Manag 2019;2019:8215371.
160. Coory JA, Parr AF, Wilkinson MP, Gupta A. Efficacy of suprascapular nerve block compared with subacromial injection: a randomized controlled trial in patients with rotator cuff tears. J Shoulder Elbow Surg 2019;28:430–6.
161. Cai YU, Sun Z, Liao B, Song Z, Xiao T, Zhu P. Sodium hyaluronate and platelet-rich plasma for partial-thickness rotator cuff tears. Med Sci Sports Exerc 2019;51:227–33.
162. Sun Y, Liu S, Chen S, Chen J. The effect of corticosteroid injection into rotator interval for early frozen shoulder: a randomized controlled trial. Am J Sports Med 2018;46:663–70.
163. Lippitt S, Harryman D, Matsen F. A practical tool for evaluation of function: the Simple Shoulder Test. In: Matsen F, Fu F, Hawkins

R, editors. The shoulder: a balance of mobility and stability. Rosemont (IL): American Academy of Orthopedic Surgeons; 1993. p. 545–59.

164. Rowe C, Zarins B. Chronic unreduced dislocations of the shoulder. Bone Joint Surg Am 1982;64:494–505.
165. Cook KF, Gartsman GM, Roddey TS, Olson SL. The measurement level and trait-specific reliability of 4 scales of shoulder functioning: an empiric investigation. Arch Phys Med Rehabil 2001;82:1558–65.
166. Van Kampen D, van Beers L, Scholtes V, Terwee C, Willems W. Validation of the Dutch version of the Simple Shoulder Test. J Shoulder Elbow Surg 2012;21:808–14.
167. Neto J, Gesser R, Steglich V, Bonilauri Ferreira A, Gandhi M, Vissoci J, et al. Validation of the Simple Shoulder Test in a Portuguese-Brazilian population. Is the latent variable structure and validation of the Simple Shoulder Test Stable across cultures? [measurement study]. PLoS One 2013;8:e62890.
168. Membrilla-Mesa M, Tejero-Fernández V, Cuesta-Vargas A, Arroyo-Morales M. Validation and reliability of a Spanish version of Simple Shoulder Test (SST-Sp). Qual Life Res 2015;24:411–6.
169. Naghdi S, Nakhostin Ansari N, Rustaie N, Akbari M, Ebadi S, Senobari M, et al. Simple shoulder test and Oxford Shoulder Score: Persian translation and cross-cultural validation. Arch Orthop Trauma Surg 2015;135:1707–18.
170. Vascellari A, Venturin D, Ramponi C, Ben G, Poser A, Rossi A, et al. Psychometric properties of three different scales for subjective evaluation of shoulder pain and dysfunction in Italian patients after shoulder surgery for anterior instability. J Shoulder Elbow Surg 2018;27:1497–504.
171. Ryliškis S, Piešina E, Kocius M, Marx R. Cross-cultural adaptation and psychometric properties of the Lithuanian version of the simple shoulder test. Acta med lituanica 2008;15:163–8.
172. Godfrey J, Hamman R, Lowenstein S, Briggs K, Kocher M. Reliability, validity, and responsiveness of the simple shoulder test: psychometric properties by age and injury type. J Shoulder Elbow Surg 2007;16:260–7.
173. Hsu J, Russ S, Somerson J, Tang A, Warme W, Matsen FA. Is the Simple Shoulder Test a valid outcome instrument for shoulder arthroplasty? [measurement study]. J Shoulder Elbow Surg 2017;26:1693–700.
174. Raman J, MacDermid J, Walton D, Athwal G. Rasch analysis indicates that the Simple Shoulder Test is robust, but minor item modifications and attention to gender differences should be considered. J Hand Ther 2017;30:348–58.
175. Van Kampen D, Willems W, van Beers L, Castelein R, Scholtes V, Terwee C. Determination and comparison of the smallest detectable change (SDC) and the minimal important change (MIC) of four-shoulder patient-reported outcome measures (PROMs). J Orthop Surg Res 2013;8:40.
176. 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;19:829–36.
177. 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;40:413–21.
178. Devereaux M, Velanoski KQ, Pennings A, Elmaraghy A. Short-term effectiveness of precut kinesiology tape versus an NSAID as adjuvant treatment to exercise for subacromial impingement: a randomized controlled trial. Clin J Sport Med 2016;26:24–32.
179. Saied A, Sadeghifar A, Okati A. Arthroscopic release versus manipulation under anesthesia for frozen shoulder: a prospective study. Journal of Isfahan Medical School 2016;34:285–92.
180. Mardani-Kivi M, Karimi Mobarakeh M, Keyhani S, Ebrahim-Zadeh MH, Haghparast Ghadim-Limudahi Z. Treatment of long head of biceps tendon lesions together with rotator cuff tears: which method is preferred? Tenotomy or tenodesis. Techniques in Shoulder & Elbow Surgery 2018;19:101–5.
181. Dawson J, Fitzpatrick R, Carr A. Questionnaire on the perceptions of patients about shoulder surgery. J Bone Joint Surg Br 1996;78:593–600.
182. Dawson J, Rogers K, Fitzpatrick R, Carr A. The Oxford Shoulder Score revisited. Arch Orthop Trauma Surg 2009;129:119–23.
183. Roh YH, Noh JH, Kim W, Oh JH, Gong HS, Baek GH. Cross-cultural adaptation and validation of the Korean version of the Oxford shoulder score. Arch Orthop Trauma Surg 2012;132:93–9.
184. Haragus H, Prejbeanu R, Patrascu J, Faur C, Roman M, Melinte R, et al. Cross-cultural adaptation and validation of the Romanian Oxford Shoulder Score. Medicine (Baltimore) 2018;97:e10926.
185. Murena L, Vulcano E, D'Angelo F, Monti M, Cherubino P. Italian cross-cultural adaptation and validation of the Oxford Shoulder Score. J Shoulder Elbow Surg 2010;19:335–41.
186. Gonçalves RS, Caldeira CQ, Rodrigues MV, Felicia SC, Cavalheiro LM, Ferreira PL. Cross-cultural adaptation and validation of the Portuguese version of the Oxford Shoulder Score (OSS). Acta Reumatol Port 2018;43:102–8.
187. Huber W, Hofstaetter JG, Hanslik-Schnabel B, Posch M, Wurnig C. The German version of the Oxford Shoulder Score: cross-cultural adaptation and validation. Arch of Orthop Trauma Surg 2004;124:531–6.
188. Da Silva Lima E, Natour J, Moreira E, Jones A. Translation, cultural adaptation and reproducibility of the Oxford Shoulder Score questionnaire for Brazil, among patients with rheumatoid arthritis. Sao Paulo Med J 2016;134:40–6.
189. Tuton D, Barbe C, Salmon JH, Drame M, Nerot C, Ohl X. Transcultural validation of the Oxford Shoulder Score for the French-speaking population. Orthop Traumatol Surg Res 2016;102:555–8.
190. Xu X, Wang F, Wang X, Wei X, Wang Z. Chinese cross-cultural adaptation and validation of the Oxford shoulder score. Health Qual Life Outcomes 2015;13:193.
191. Frich LH, Noergaard PM, Brorson S. Validation of the Danish version of Oxford Shoulder Score. Dan Med Bull 2011;58:A4335.
192. Bejer A, Szczepanik M, Plocki J, Szymczyk D, Kulczyk M, Pop T. Translation, cross-cultural adaptation and validation of the polish version of the Oxford Shoulder Score in patients undergoing arthroscopic rotator cuff repair. Health Qual Life Outcomes 2019;17:191.
193. Baker P, Nanda R, Goodchild L, Finn P, Rangan A. A comparison of the Constant and Oxford Shoulder Scores in patients with conservatively treated proximal humeral fractures. J Shoulder Elbow Surg 2008;17:37–41.
194. Rees J, Dawson J, Hand G, Cooper C, Judge A, Price A, et al. The use of patient-reported outcome measures and patient satisfaction ratings to assess outcome in hemiarthroplasty of the shoulder. J Bone Joint Surg Br 2010;92:1107–11.
195. Dawson J, Hill G, Fitzpatrick R, Carr A. Comparison of clinical and patient-based measures to assess medium-term outcomes following shoulder surgery for disorders of the rotator cuff. Arthritis Rheum 2002;47:513–9.
196. Christie A, Dagfinrud H, Garratt AM, Ringen Osnes H, Hagen KB. Identification of shoulder-specific patient acceptable symptom state in patients with rheumatic diseases undergoing shoulder surgery. J Hand Ther 2011;24:53–60.
197. Handoll H, Brealey S, Rangan A, Keding A, Corbacho B, Jefferson L, et al. The ProFHER (PROximal Fracture of the Humerus: evaluation by Randomisation) trial: a pragmatic multicentre randomised

controlled trial evaluating the clinical effectiveness and cost-effectiveness of surgical compared with non-surgical treatment for proximal fracture of the humerus in adults. Health Technol Assess 2015;19:1–280.

198. Ebadi S, Forogh B, Fallah E, Babaei Ghazani A. Does ultrasound therapy add to the effects of exercise and mobilization in frozen shoulder? A pilot randomized double-blind clinical trial. J Bodyw Mov Ther 2017;21:781–7.
199. Robinson PM, Norris J, Roberts CP. Randomized controlled trial of supervised physiotherapy versus a home exercise program after hydrodilatation for the management of primary frozen shoulder. J Shoulder Elbow Surg 2017;26:757–65.
200. Beard DJ, Rees JL, Cook JA, Rombach I, Cooper C, Merritt N, et al. Arthroscopic subacromial decompression for subacromial shoulder pain (CSAW): a multicentre, pragmatic, parallel group, placebo-controlled, three-group, randomised surgical trial. Lancet 2018;391:329–38.
201. Moosmayer S, Ekeberg OM, Hallgren HB, Heier I, Kvalheim S, Blomquist J, et al. KALK study: ultrasound guided needling and lavage (barbotage) with steroid injection versus sham barbotage with and without steroid injection: protocol for a randomized, double-blinded, controlled, multicenter study. BMC Musculoskelet Dis 2017;18:138.
202. Wang A, McCann P, Colliver J, Koh E, Ackland T, Joss B, et al. Do postoperative platelet-rich plasma injections accelerate early tendon healing and functional recovery after arthroscopic supraspinatus repair? A randomized controlled trial. Am J Sports Med 2015;43:1430–7.
203. Frassanito P, Cavalieri C, Maestri R, Felicetti G. Effectiveness of Extracorporeal Shock Wave Therapy and kinesio taping in calcific tendinopathy of the shoulder: a randomized controlled trial. Eur J Phys Rehabil Med 2018;54:333–40.
204. Piekaar RS, Bouman IC, van Kampen PM, van Eijk F, Huijsmans PE. Early promising outcome following arthroscopic implantation of the subacromial balloon spacer for treating massive rotator cuff tear. Musculoskelet Surg 2018;102:247–55.
205. Issa Z, Rasmussen JV, Petersen JK, Schantz K, Brorson S. Patient-reported outcome after stemmed versus stemless total shoulder arthroplasty for glenohumeral osteoarthritis: a patient-blinded randomized clinical trial. Trials 2019;20:427.
206. Launonen AP, Fjalestad T, Laitinen MK, Lahdeoja T, Ekholm C, Wagle T, et al. Nordic Innovative Trials to Evaluate osteoPorotic Fractures (NITEP) Collaboration: the Nordic DeltaCon Trial protocol: non-operative treatment versus reversed total shoulder arthroplasty in patients 65 years of age and older with a displaced proximal humerus fracture: a prospective, randomised controlled trial. BMJ Open 2019;9:e024916.
207. Kennedy C, Beaton D, Solway S, McConnell S, Bombardier C. The DASH outcome measure user's manual. 3rd ed. Toronto: Institute for Work & Health; 2011.
208. 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;29:602–8.
209. Cano SJ, Barrett LE, Zajicek JP, Hobart JC. Beyond the reach of traditional analyses: using Rasch to evaluate the DASH in people with multiple sclerosis. Mult Scler 2011;17:214–22.
210. 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;91:1370–7.
211. Beaton DE, Wright JG, Katz JN. Upper Extremity Collaborative Group. Development of the QuickDASH: comparison of three item-reduction approaches. J Bone Joint Surg Am 2005;87:1038–46.
212. Gabel CP, Yelland M, Melloh M, Burkett B. A modified QuickDASH-9 provides a valid outcome instrument for upper limb function. BMC Musculoskelet Disord 2009;10:161.
213. Germann G, Harth A, Wind G, Demir E. Standardisation and validation of the German version 2.0 of the Disability of Arm, Shoulder, Hand (DASH) questionnaire. Unfallchirurg 2003;106:13–9. In German.
214. Roy JS, MacDermid JC, Woodhouse LJ. Measuring shoulder function: a systematic review of four questionnaires. Arthritis Rheum 2009;61:623–32.
215. 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;84:208–15.
216. Beaton D, Davis A, Hudak PL, McConnell S. The DASH (Disabilities of the Arm, Shoulder and Hand) outcome measure: what do we know about it now? [review]. Br J Hand Ther 2001;6:109–18.
217. 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;14:128–46.
218. Raven EE, Haverkamp D, Sierevelt IN, van Montfoort DO, Poll RG, Blankevoort L, et al. Construct validity and reliability of the disability of arm, shoulder and hand questionnaire for upper extremity complaints in rheumatoid arthritis. J Rheumatol 2008;35:2334–8.
219. Thoomes-de Graaf M, Scholten-Peeters GG, Schellingerhout JM, Bourne AM, Buchbinder R, Koehorst M, 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;25:2141–60.
220. Haldorsen B, Svege I, Roe Y, Bergland A. Reliability and validity of the Norwegian version of the Disabilities of the Arm, Shoulder and Hand questionnaire in patients with shoulder impingement syndrome. BMC Musculoskelet Disord 2014;15:78.
221. Gummesson C, Atroshi I, Ekdahl C. The disabilities of the arm, shoulder and hand (DASH) outcome questionnaire: longitudinal construct validity and measuring self-rated health change after surgery. BMC Musculoskelet Disord 2003;4:11.
222. 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.
223. Marx RG, Bombardier C, Hogg-Johnson S, Wright JG. Clinimetric and psychometric strategies for development of a health measurement scale. J Clin Epidemiol 1999;52:105–11.
224. Slobogean GP, Noonan VK, O'Brien PJ. The reliability and validity of the Disabilities of Arm, Shoulder, and Hand, EuroQol-5D, Health Utilities Index, and Short Form-6D outcome instruments in patients with proximal humeral fractures. J Shoulder Elbow Surg 2010;19:342–8.
225. 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;18:920–6.
226. Rysstad T, Roe Y, Haldorsen B, Svege I, Strand LI. Responsiveness and minimal important change of the Norwegian version of the Disabilities of the Arm, Shoulder and Hand questionnaire (DASH) in patients with subacromial pain syndrome. BMC Musculoskelet Disord 2017;18:248.

227. Huisstede BM, Feleus A, Bierma-Zeinstra SM, Verhaar JA, Koes BW. Is the disability of arm, shoulder, and hand questionnaire (DASH) also valid and responsive in patients with neck complaints. Spine (Phila Pa 1976) 2009;34:E130–8.
228. Angst F, Goldhahn J, Drerup S, Flury M, Schwyzer HK, Simmen BR. How sharp is the short QuickDASH? A refined content and validity analysis of the short form of the disabilities of the shoulder, arm and hand questionnaire in the strata of symptoms and function and specific joint conditions. Qual Life Res 2009;18:1043–51.
229. Mahabier KC, Den Hartog D, Theyskens N, Verhofstad MH, Van Lieshout EM, HUMMER Trial Investigators. Reliability, validity, responsiveness, and minimal important change of the Disabilities of the Arm, Shoulder and Hand and Constant-Murley scores in patients with a humeral shaft fracture. J Shoulder Elbow Surg 2017;26:e1–12.
230. Hosseini Khameneh SM, Abbasian M, Abrishamkarzadeh H, Bagheri S, Abdollahimajd F, Safdari F, et al. Humeral shaft fracture: a randomized controlled trial of nonoperative versus operative management (plate fixation). Orthop Res Rev 2019;11:141–7.
231. Bruce J, Williamson E, Lait C, Richmond H, Betteley L, Lall R, et al. Randomised controlled trial of exercise to prevent shoulder problems in women undergoing breast cancer treatment: study protocol for the prevention of shoulder problems trial (UK PROSPER). BMJ Open 2018;8:e019078.
232. Wesner M, Defreitas T, Bredy H, Pothier L, Qin Z, McKillop AB, et al. A pilot study evaluating the effectiveness of platelet-rich plasma therapy for treating degenerative tendinopathies: a randomized control trial with synchronous observational cohort. PLoS One 2016;11:e0147842.
233. Amstutz H, Sew Hoy A, Clarke I. UCLA anatomic total shoulder arthroplasty. Clin Orthop Relat Res 1981:7–20.
234. Ellman H, Hanker G, Bayer M. Repair of the rotator cuff. End-result study of factors influencing reconstruction. J Bone Joint Surg Am 1986;68:1136–44.
235. Coghlan JA, Bell SN, Forbes A, Buchbinder R. Comparison of self-administered University of California, Los Angeles, shoulder score with traditional University of California, Los Angeles, shoulder score completed by clinicians in assessing the outcome of rotator cuff surgery. J Shoulder Elbow Surg 2008;17:564–9.
236. Tauro J. Arthroscopic rotator cuff repair: analysis of technique and results at 2- and 3-year follow-up. Arthroscopy 1998;14:45–51.
237. Brinker MR, Cuomo JS, Popham GJ, O'Connor DP, Barrack RL. An examination of bias in shoulder scoring instruments among healthy collegiate and recreational athletes. J Shoulder Elbow Surg 2002;11:463–9.
238. Oku E, Andrade A, Stadiniky S, Carrera E, Tellini G. Translation and cultural Adaptation of the modified-University of California at Los Angeles Shoulder Rating Scale to Portuguese language. Rev Bras Reumatol 2006;46:246–52.
239. Burkhart S, Danaceau S, Pearce CJ. Arthroscopic rotator cuff repair: analysis of results by tear size and by repair technique-margin convergence versus direct tendon-to-bone repair. Arthroscopy 2001;17:905–12.
240. Diniz Lopes A, Ciconelli RM, Carrera EF, Griffin S, Faloppa F, dos Reis FB. Validity and reliability of the Western Ontario Rotator Cuff Index (WORC) for use in Brazil. Clin J Sport Med 2008;18:266–72.
241. Diniz Lopes A, Ciconelli RM, Carrera EF, Griffin S, Faloppa F, dos Reis FB. Comparison of the responsiveness of the Brazilian version of the Western Ontario Rotator Cuff Index (WORC) with DASH, UCLA and SF-36 in patients with rotator cuff disorders. Clin Exp Rheumatol 2009;27:758–64.
242. Balci TO, Turk AC, Sahin F, Kotevoglu N, Kuran B. Efficacy of therapeutic ultrasound in treatment of adhesive capsulitis: a prospective double blind placebo-controlled randomized trial. J Back Musculoskelet Rehabil 2018;31:955–61.
243. Baskaya MC, Ercalik C, Karatas Kir O, Ercalik T, Tuncer T. The efficacy of mirror therapy in patients with adhesive capsulitis: a randomized, prospective, controlled study. J Back Musculoskelet Rehabil 2018;31:1177–82.
244. Lin J. Platelet-rich plasma injection in the treatment of frozen shoulder: a randomized controlled trial with 6-month follow-up. Int J Clin Pharmacol Ther 2018;56:366–71.
245. De Roo PJ, Muermans S, Maroy M, Linden P, van den Daelen L. Passive mobilization after arthroscopic rotator cuff repair is not detrimental in the early postoperative period. Acta Orthop Belg 2015;81:485–92.
246. Zhang J, Bai D, Yang J, Y L, Zhao C. Early motion versus immobilization for arthroscopic repair in the treatment of large size rotator cuff tears. Biomedical Research 2017;28:6818–22.
247. Diaz-Sanabria JI, Trueba-Vasavilbaso C, Gil-Orbezo F, Trueba-Davalillo C, Acuna-Tovar M. Doxycycline as an adjuvant in post-surgical recovery of rotator cuff injuries. One-year follow-up. Acta Ortop Mex 2017;31:157–61. In Spanish.
248. Malavolta EA, Gracitelli MEC, Assuncao JH, Ferreira Neto AA, Bordalo-Rodrigues M, de Camargo OP. Clinical and structural evaluations of rotator cuff repair with and without added platelet-rich plasma at 5-year follow-up: a prospective randomized study. Am J Sports Med 2018;46:3134–41.
249. Franceschi F, Papalia R, Franceschetti E, Palumbo A, del Buono A, Paciotti M, et al. Double-row repair lowers the retear risk after accelerated rehabilitation. Am J Sports Med 2016;44:948–56.
250. Yamakado K. Healing rate comparison between suture-bridging and medially based single row repair in medium-sized supraspinatus tear: a randomized prospective trial. Arthroscopy 2019;35:2803–13.
251. Rueda Garrido JC, Vas J, Lopez DR. Acupuncture treatment of shoulder impingement syndrome: a randomized controlled trial. Complement Ther Med 2016;25:92–7.
252. Sahin Onat S, Bicer S, Sahin Z, Kucukali Turkyilmaz A, Kara M, Ozbudak Demir S. Effectiveness of kinesiotaping and subacromial corticosteroid injection in shoulder impingement syndrome. Am J Phys Med Rehabil 2016;95:553–60.
253. Yu K, Zhang DY, Yang J, Liu M, Yang J, Tan L. Clinical efficacy of ultrasound-guided subacromial drug injection in the treatment of subacromial impingement syndrome. Zhonghua wai ke za zhi 2018;56(10):781–5. In Chinese.
254. Uchiyama Y, Handa A, Shimpuku E, Omi H, Hashimoto H, Imai T, et al. Open Bankart repair plus inferior capsular shift versus arthroscopic Bankart repair without augmentations for traumatic anterior shoulder instability: a prospective study. J Orthop Surg (Hong Kong) 2017;25:2309499017727947.
255. Coghlan J, Forbes A, McKenzie D, Bell S, Buchbinder R. Efficacy of subacromial ropivacaine infusion for rotator cuff surgery: a randomized trial. J Bone Joint Surg Am 2009;91:1558–67.
256. Lenza M, Taniguchi L, Ferretti M. Figure-of-eight bandage versus arm sling for treating middle-third clavicle fractures in adults: study protocol for a randomised controlled trial. Trials 2016;17:229.
257. Kirkley A, Alvarez C, Griffin S. The development and evaluation of a disease-specific quality-of-life questionnaire for disorders of the rotator cuff: the Western Ontario Rotator Cuff Index. Clin J Sport Med 2003;13:84–92.
258. Wessel RN, Lim TE, van Mameren H, de Bie RA. Validation of the Western Ontario Rotator Cuff index in patients with arthroscopic rotator cuff repair: a study protocol. BMC Musculoskelet Disord 2011;12:64.

259. Wessel RN, Wolterbeek N, Fermont AJ, Lavrijsen L, van Mameren H, de Bie RA. Responsiveness and disease specificity of the Western Ontario Rotator Cuff index. J Orthop 2018;15:337–42.
260. Gadsboell J, Tibaek S. Validity of a shoulder-specific quality of life questionnaire, the Western Ontario Rotator Cuff Index, for patients with scapula alata. JSES Open Access 2017;1:29–34.
261. Wang W, Xie QY, Jia ZY, Cui L, Liu D, Wang CR, et al. Cross-cultural translation of the Western Ontario Cuff Index in Chinese and its validation in patients with rotator cuff disorders. BMC Musculoskelet Disord 2017;18:178.
262. Zhaeentan S, Legeby M, Ahlström S, Stark A, Salomonsson B. A validation of the Swedish version of the WORC index in the assessment of patients treated by surgery for subacromial disease including rotator cuff syndrome. BMC Musculoskelet Disord 2016;17:165.
263. St-Pierre C, Dionne CE, Desmeules F, Roy JS. Reliability, validity, and responsiveness of a Canadian French adaptation of the Western Ontario Rotator Cuff (WORC) index. J Hand Ther 2015;28:292–8.
264. Bejer A, Probachta M, Kulczyk M, Griffin S, Domka-Jopek E, Plocki J. Validation of the Polish version of the Western Ontario Rotator Cuff Index in patients following arthroscopic rotator cuff repair. BMC Musculoskelet Disord 2018;19:333.
265. El O, Bircan C, Gulbahar S, Demiral Y, Sahin E, Baydar M, et al. The reliability and validity of the Turkish version of the Western Ontario Rotator Cuff Index. Rheumatol Int 2006;26:1101–8.
266. Mousavi SJ, Hadian MR, Abedi M, Montazeri A. Translation and validation study of the Persian version of the Western Ontario Rotator Cuff Index. Clin Rheumatol 2009;28:293–9.
267. Getahun TY, MacDermid JC, Patterson SD. Concurrent validity of patient rating scales in assessment of outcome after rotator cuff repair. Journal of Musculoskeletal Research 2000;4:119–27.
268. Longo UG, Saris D, Poolman RW, Berton A, Denaro V. Instruments to assess patients with rotator cuff pathology: a systematic review of measurement properties. Knee Surg Sports Traumatol Arthrosc 2012;20:1961–70.
269. Lopes AD, Vilar e Furtado R, Silva CA, Yi LC, Malfatti CA, Araujo SA. Comparison of self-report and interview administration methods based on the Brazilian versions of the Western Ontario Rotator Cuff Index and Disabilities of the Arm, Shoulder and Hand Questionnaire in patients with rotator cuff disorders. Clinics (Sao Paulo) 2009;64:121–5.
270. Hays R, Spritzer K, Amtmann D, Lai J, Dewitt E, Rothrock N, et al. Upper extremity and mobility subdomains from the patient-reported outcome measurement information system (PROMIS) adult physical functioning item bank. Arch Phys Med Rehabil 2013;94:2291–6.
271. DeWalt D, Rothrock N, Yount S, Stone A. Evaluation of item candidates: the PROMIS Qualitative Item Review. Med Care 2007;45 Suppl 1:S12–21.
272. Reeve B, Hays R, Bjorner J, Cook K, Crane P, Terisi J, et al. Psychometric evaluation and calibration of health-related quality of life item banks: plans for the patient-reported outcomes measurement Information system (PROMIS). Med Care 2007;45 Suppl 1:S22–31.
273. Liu H, Cella D, Gershon G, Ahen J, Morales L, Riley W. Representativeness of the Patient-Reported Outcomes Measurement Information System internet panel. J Clin Epidemiol 2010;63:1169–78.
274. Rothrock N, Hays R, Spritzer K, Yount S, Riley W, Cella D. Relative to the general US population, chronic diseases are associated with poorer health-related quality of life as measured by the Patient-Reported Outcomes Measurement Information System (PROMIS). J Clin Epidemiol 2010;63:1195–204.
275. Cella D, Riley W, Stone A, Rothrock N, Reeve B, Yount S, et al. The Patient-Reported Outcomes Measurement Information System (PROMIS) developed and tested its first wave of adult self-reported health outcome item banks: 2005–2008. J Clin Epidemiol 2010;63:1179–94.
276. Doring A, Nota SP, Hageman MG, Ring DC. Measurement of upper extremity disability using the patient reported outcomes measurement information system. J Hand Surg Am 2014;39:1160–5.
277. Minoughan C, Schumaier A, Frithc J, Grawe B. Correlation of patient reported outcome measurement information system physical function upper extremity computer adaptive testing, with the American shoulder and elbow surgeons shoulder assessment form and simple shoulder test in patients with shoulder arthritis. J Shoulder Elbow Surg 2018;27:585–91.
278. Minoughan C, Schumaier A, Frithc J, Grawe B. Correlation of patient reported outcome measurement information system physical function upper extremity computer adaptive testing, with the American shoulder and elbow surgeons shoulder assessment form and simple shoulder test in patients with shoulder pain. Arthroscopy 2018;34:1430–6.
279. Gershon R, Rothrock N, Hanrahan R, Bass M, Cella D. The use of PROMIS and Assessment Center to deliver patient-reported outcomes in clinical research. J Appl Meas 2010;11:304–14.
280. van Bruggen S, Lameijer C, Terwee C. Structural validity and construct validity of the Dutch-Flemish PROMIS physical function-upper extremity version 2.0 item bank in Dutch patients with upper extremity injuries. Disabil Rehabil 2019:1–9.
281. Dowdle S, Glass N, Anthony C, Hettrich C. Use of PROMIS for patients undergoing primary total shoulder arthroplasty. Orthop J Sports Med 2017;5:2325967117726044.
282. Anthony C, Glass N, Hancock K, Bollier M, Hettrich C, Wolf B. Preoperative performance of the patient reported outcome measurement information system in patients with rotator cuff pathology. Arthroscopy 2017;33:1770–4.
283. Jayakumar P, Teunis T, Vranceanu A, Lamb S, Williams M, Ring D, et al. Construct validity and precision of different patient-reported outcome measures during recovery after upper extremity fractures. Clin Orthop Rel Res 2019;477:2521–30.
284. Hajewski C, Glass N, Westermann R, Bollier M, Wolf B, Hettrich C. Performance of the PROMIS after operative interventions for shoulder instability. Orthop J Sports Med 2019;7:2325967119846920.
285. Peterson B, Orvets N, Aleem A, Keener J, Calfee R, Nixon D, et al. Correlation of PROMIS scores with legacy patient-reported outcome scores in patients undergoing rotator cuff repair. J Shoulder Elbow Surg 2018;27:S17–23.
286. Beleckas C, Padovano A, Guattery J, Chamberlain A, Keener J, Calfee R. Performance of Patient-Reported Outcomes Measurement Information System (PROMIS) upper extremity (UE) verses physical function (PF) computer adaptive tests (CATs) in upper extremity clinics. J Hand Surg Am 2017;42:867–74.
287. Anthony C, Glass N, Hancock K, Bollier M, Wolf B, Hettrich C. Performance of PROMIS instruments in patients with shoulder instability. Am J Sports Med 2016;45:449–53.
288. Hung M, Voss M, Bounsanga J, Crum A, Tyser A. Examination of the PROMIS upper extremity item bank. J Hand Ther 2017;30:485–90.
289. Kaat A, Rothrock N, Vrahas M, O'Toole R, Buono S, Zerhusen T, et al. Longitudinal validation of the PROMIS physical function in upper extremity trauma. J Orthop Trauma 2017;31:e321–6.
290. Gulledge C, Lizzio V, Smith G, Guo E, Makhni E. What are the floor and ceiling effects of patient-reported outcomes measurement information system computer adaptive test domains in orthopaedic patients? A systematic review. Arthroscopy 2020;36:901–12.

291. Makhni E, Meldau J, Blanchett J, Borowsky P, Stephens J, Muh S, et al. Correlation of PROMIS physical function, pain interference, and depression in pediatric and adolescent patients in the ambulatory sports medicine clinic. Orthop J Sports Med 2019;7:2325967119851100.

292. Fisk F, Francovic S, Tramer J, Gulledge C, Kuhlmann N, Chen C, et al. PROMIS CAT forms demonstrate responsiveness in patients following arthroscopic rotator cuff repair across numerous health domains. J Shoulder Elbow Surg 2019;28:2427–32.

293. Kazmers N, Qiu Y, Yoo M, Stephens A, Tyser A, Zhang Y. The minimal clinically important difference of the PROMIS and QuickDASH instruments in a nonshoulder hand and upper extremity patient population. J Hand Surg Am 2020;45:399–407.

Table 1. Practical applications*

Measure	Number of Items	Content/Domains	Method of Administration	Recall Period	Response Format	Range of Scores	Score Interpretation	Availability of Normative Data	Cross-Cultural Validation
OSS (revised version)	12	Pain (four items) and daily functions (eight items)	Self	4 wk	Five-point Likert scale ranging from 0-4 (worst to best)	0-48	0-19 = severe arthritis, 20-29 = moderate to severe arthritis, 30-39 = mild to moderate arthritis, and 40-48 = satisfactory joint function	No	Dutch, Italian, German, Portuguese, Korean, French, Turkish, Chinese, Persian, Spanish, and Romanian
SST	12	Pain (two items), function/strength (seven items), and ROM (three items)	Self	Current	Yes/no	0-12	Percentage of "yes" responses ranging from 0-100 (worst to best)	No	Dutch, Portuguese, Spanish, Persian, and Italian
UCLA Shoulder Rating Scale (modified)	5	Pain (one item), function (one item), patient satisfaction (one item), forward flexion (one item), and strength of forward flexion (one item)	Self and observer (with a self-completed version)	4 wk	Weighted categorical scales; pain and function: 1-10; other items: 0-5	2-35	34-35 = excellent, 28-33 = good (both excellent and good indicate a satisfactory result), 21-27 = fair, and 20 or less = poor (both fair and poor are unsatisfactory)	Yes	Brazilian Portuguese and Italian
ASES Society Form	45 (12 items in score)	Score: pain (one item) and ADLs (11 items); other domains not included in score: pain (five items), stability (two items), physician-assessed ROM (five items), signs (11 items), strength (five items), and instability (eight items)	Only self-assessment (pain and ADLs) in score	Current	0-10 VAS for pain and four-point ordinal Likert scales ranging from 0-3 (unable to do to not difficult) for ADLs	0-100	Pain and ADL subscales are each transformed to 0-50 (0 = worst; 100 = best).	Yes	German, Italian, Brazilian Portuguese, Spanish, Finnish, Turkish, and Tunisian Arabic
SPADI	13	Pain (five items) and disability (eight items)	Self	1 wk	0-10 NRS	0-100	Each subscale is from 0-100; the total score is the mean of the two subscales and ranges from 0-100 (best to worst)	Yes	Spanish, Italian, Chinese, Nepali, Thai, Arabic, Persian, Brazilian Portuguese, Danish, German, Dutch, Turkish, Greek, Indian Tamil, and Indian Hindi

(Continued)

Table 1. (*Cont'd*)

Measure	Number of Items	Content/Domains	Method of Administration	Recall Period	Response Format	Range of Scores	Score Interpretation	Availability of Normative Data	Cross-Cultural Validation
DASH/ QuickDASH	DASH: 30; QuickDASH: 11	DASH: symptoms (six items) and function (24 items); QuickDASH: symptoms (three items) and function (eight items)	Self	1 wk	Five-point ordinal Likert scales: 1 = no symptoms/difficulty and 5 =extreme difficulty (unable to do)/symptoms	0-100	Scores are transformed by (mean − 1) × 25. 0-100 (best to worst)	Yes	Available in over 50 languages
CS	10	Pain (one item), ADLs (four items), ROM (four items), and strength (one item)	Self (pain and ADLs) and observer (ROM and strength)	1 wk (ADLs) and past 24 h (pain)	Pain: 0-15 VAS; ADLs: sleep: two-item ordinal scale (sleep) and 0-15 VAS (work and recreation); functional mobility: five-item ordinal scale	0-100	Total score = sum of subscales (pain = 15, ADL = 20, ROM = 40, and strength = 25) and ranges from 0-100 (worst to best)	Yes	Danish, Turkish, Brazilian Portuguese, Greek, and Chinese
WORC	21	Physical symptoms (six items), sports/ recreation (four items), work (four items), lifestyles (three items), and emotions (three items)	Self	2 wk	0-100 VAS	0-100 (normalized)	A higher score indicates greater function on the normalized score.	No	Chinese, Swedish, French, Polish, Turkish, Portuguese, Persian, and Norwegian
PROMIS	≤12	Upper extremity physical function	Self	Current	Five-point ordinal Likert scale (1 = unable to do, 2 = with much difficulty, 3 = with some difficulty, 4 = with a little difficulty, and 5 = without any difficulty)	*t* metric, mean 50 (SD 10)	A higher score indicates greater function. Scores of 60 and >70 indicate that physical function exceeds 68.27% and 95.45% of general population, respectively.	Yes	Dutch-Flemish

* ADL = activity of daily living; ASES = American Shoulder and Elbow Surgeons; CS = Constant Murley Shoulder Scale; DASH = Disabilities of the Arm, Shoulder, and Hand Questionnaire; OSS = Oxford Shoulder Score; PROMIS = Patient-Reported Outcomes Measurement Information System; ROM = range of motion; SPADI = Shoulder Pain and Disability Index; SST = Simple Shoulder Test; UCLA = University of California, Los Angeles; VAS = visual analog scale; WORC = Western Ontario Rotator Cuff Index.

Table 2. Psychometrics*

Measure	Floor and Ceiling Effects	Reliability	Validity	Responsiveness	Minimally Important Differences	Generalizability	Used in RCTs
OSS	Both low	Internal consistency: Cronbach's α = 0.94; test-retest reliability: Pearson's correlation = 0.98.	Moderate to high correlation with other measures (range: 0.37-0.87)	ES = 2.3 for osteoarthritis and rheumatoid arthritis hemiarthroplasty; ES = 1.10-1.88 and SRM = 1.10-1.14 for impingement rotator cuff surgery; ES 0.97 for rotator cuff decompression (cuff repair) (77); ES = 0.96 for impingement (no treatment described) (43); and ES 0.61 for degenerative inflammatory surgery (64)	MCID = 6-6.9 on a 48-point scale	Widely translated and adapted into different languages and cultural settings used for patients with various shoulder problems	Yes
SST	Both below 21%	Unidimensional scale with good test-retest reliability (0.97-0.99)	High construct validity (>0.68) with shoulder-specific measures such as the SPADI, ASES, DASH, CS, and WORC	The ES is considered moderate (>0.63) to large (>0.87) in patients with osteoarthritis, shoulder instability, rotator cuff diseases, and total arthroplasty.	For the 0-12 scale, MCID > 2.05 for rotator cuff disease; MCID > 2.4 for shoulder arthroplasty; and MCID = 2.2 points for various shoulder problems (corresponds with MCID >17 for the 0-100 scale)	It is shoulder specific, has also been used to assess shoulder function impacted by other diseases (eg, surgery for breast cancer) and has been translated and culturally adapted to various countries.	Yes
UCLA Shoulder Rating Scale	Not reported	Single items within domains; moderate to good test-retest reliability (ICC = 0.93, SEM = 1.5)	Moderate to high correlation with other measures	ES = 1.17, 2.0, and 2.73 at 6 wk, 6 mo, and 12 mo after subacromial decompression, respectively (SRM 0.83, 1.41, and 1.69 points respectively); SRM = 0.385 points after instability surgery; and ES = 1.17 and SRM = 1.66 3 mo after rotator cuff disorder treatment (Brazilian Portuguese version)	MCID = 2.0-2.4 (on a 2-25 scale) (depending on method) for proximal humerus fracture	Translated and culturally adapted into two languages but used in many trials in other countries	Yes
ASES Society Standardized Shoulder Assessment Form	Low floor and ceiling effects	Internal consistency: Cronbach's α= 0.61-0.96; fair to excellent test-retest reliability	Moderate to high correlation with other measures	ES = 0.8-2.13 and SRM = 0.75-1.81 across various shoulder conditions	MCID = 6.3-26.9 (on a 0-100 scale)	Translated and culturally adapted to various countries	Yes
SPADI	Low floor and ceiling effects	Internal consistency: Cronbach's α = 0.92-0.96; test-retest reliability: ICC = 0.84-0.95	Moderate to high correlation with other measures (range: 0.45-0.93)	ES = 2.1 for total shoulder arthroplasty; ES = 1.20-1.64 for adhesive capsulitis; ES = 1.26 for shoulder pain (physical therapy); SRM = 1.23 for rotator cuff surgery:; and ES = 1.21for upper extremity disorders (occupational therapy or physical therapy)	MCID = 8-20 (on a 0-100 scale)	Shoulder-specific and used in a wide range of clinical and research settings	Yes

(Continued)

Table 2. (*Cont'd*)

Measure	Floor and Ceiling Effects	Reliability	Validity	Responsiveness	Minimally Important Differences	Generalizability	Used in RCTs
DASH/ QuickDASH	Very low or low floor and ceiling effects	Internal consistency: Cronbach's α = 0.92-0.98 for the DASH and 0.92-0.95 for the QuickDASH; high test-retest reliability: 0.89-0.98 for the DASH and 0.90-0.94 for the QuickDASH	Moderate to high correlation with shoulder-specific measures (range: 0.50-0.93)	ES = 0.55-1.26 and SRM = 0.5-1.45 across various shoulder conditions	MCID = 4.4-12.4 for the DASH and MCID = 8-13.4 QuickDASH (on 0-100 scales)	Has been used in a wide range of clinical and research settings for a broad range of upper extremity (including shoulder) conditions	Yes
CS	Low floor and ceiling effects (possible floor effect for strength subscale)	Internal consistency: Cronbach's α = 0.37-0.8; test-retest reliability: ICC = 0.82-0.98	Moderate to high correlation with other shoulder measures (range: 0.49-0.91)	ES = 0.4 and SRM = 1.38 for rotator cuff disease; ES =1.23-1.92 and SRM = 1.12-2.09 for subacromial decompression surgery; and ES = 2.23-2.9 and SRM = 1.99-2.4 for shoulder arthroplasty	MCID = 8-17 (on a 0-100 scale)	Used in a wide range of clinical and research settings for a broad range of shoulder conditions	Yes
WORC	Lower floor and ceiling effects. Possible ceiling effect after treatment.	Test-retest reliability: ICC = 0.96 for the total score and 0.54-0.91 for subscales	Moderate to high correlation with other shoulder measures, low to moderate correlation with non–shoulder specific measures	SEM = 1.33 among patients with rotator cuff pathology	MCID = 22.9-34 (on a 0-100 scale)	Widely translated and adapted in many languages but limited to diseases related to rotator cuff disorders	Yes
PROMIS	0 to 1.6% floor and 0 to 28% ceiling effects	Item reliability: r = 0.82-0.96; person reliability: r = 0.84-0.85 in patients with shoulder instability; average marginal reliability: r = 0.90 in patients with upper extremity trauma	Low to high correlation with shoulder measures (range: 0.34-0.91)	Good responsiveness to change reported; SRM = 0.92 after surgery for shoulder instability	MCID not reported for shoulder conditions; MCID = 2.1 in 1 study of nonshoulder upper extremity disorders	Has been used to assess various shoulder disorders but is not yet widely translated	No

* ASES = American Shoulder and Elbow Surgeons; CS = Constant Murley Shoulder Scale; DASH = Disabilities of the Arm, Shoulder, and Hand Questionnaire; ICC = intraclass correlation coefficient; MCID = minimal clinically important difference; OSS = Oxford Shoulder Score; PROMIS = Patient-Reported Outcomes Measurement Information System; RCT = randomized controlled trial; SPADI = Shoulder Pain and Disability Index; SRM = standardized response mean; SST = Simple Shoulder Test; UCLA = University of California, Los Angeles; WORC = Western Ontario Rotator Cuff Index.