PROF. RACHELLE BUCHBINDER (Orcid ID : 0000-0002-0597-0933)

DR. SOFIA RAMIRO (Orcid ID : 0000-0002-8899-9087)

DR. JOEL J GAGNIER (Orcid ID : 0000-0002-3162-3935)

DR. SAMUEL L WHITTLE (Orcid ID : 0000-0001-7417-7691)

Article type : Special Article

Corresponding author e-mail address : rachelle.buchbinder@monash.edu

Measures of adult shoulder function

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, Disabilities of the Arm, Shoulder, and Hand Questionnaire (DASH) and Its Short Version (QuickDASH), University of California–Los Angeles (UCLA) Shoulder Rating Scale, Western Ontario Rotator Cuff (WORC) Index, Patient-Reported Outcomes Measurement Information System (PROMIS) Physical Function Upper Extremity Computerized Adaptive Test (CAT)

Buchbinder R, Ramiro S, Huang H, Gagnier J, Jia Y, Whittle S.

Introduction

In the first systematic review of randomized controlled trials evaluating efficacy of interventions for painful shoulder published in 1998, none of the 31 included trials included 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 upon 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 five years identified searching Ovid MEDLINE and Ovid EMBASE from 2015 until 9 December 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 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 summarises the available information about how these measures perform for different patient populations in different settings.

SPADI - Shoulder Pain and Disability Index

Description

Purpose. Patient-reported outcome measure (PROM) 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 due to inadequate test-retest reliability or poor correlation with shoulder range of motion, resulting in a 13-item scale.

Content or domains. Comprises two subscales: pain and disability.

Number of items. 13 items in total, 5 in pain subscale and 8 in disability subscale.

Response options/scale. All items were originally rated using a visual analogue scale (VAS) (5). More recent versions have most commonly used an 11-point (0–10) numerical rating scale (NRS) (6). Anchors for each of the pain items are 0 = 'no pain' and 10 = 'worst pain imaginable', and for the 8 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. Printed in various references, e.g., (7). Free PDF version online at https://www.tac.vic.gov.au/files-to-move/media/upload/spi.pdf, or as an online calculator at https://www.orthotoolkit.com/spadi/.

Practical application

Method of administration. Self-assessment, either using pen and paper or electronic version. May also be administered via telephone (8).

Scoring. Each item scored 0 (best) to 10 (worst). A minimum of 2/3 of 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 x 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 (n=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 three 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 three years ago (mean score 6.1, $p<0.0001$). Women

had a higher (worse) mean score (4.2) than men (2.7) after adjustment for nationality (p=0.026) and scores increased with age.

Respondent time to complete. Median time to complete 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 final score. Administration and scoring takes 5 minutes (5).

Translations/adaptations. 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 alpha scores for the total SPADI scale and the individual subscales in various settings, including shoulder disorders in outpatients (Cronbach's alpha: total = 0.96, disability = 0.95, pain = 0.89) (29); population-based individuals with shoulder pain (Cronbach's alpha: SPADI = 0.92, disability = 0.90, 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% CI 0.42 to 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 to 0.93), pain subscale 0.85 (95% CI 0.76 to 0.91),

disability subscale 0.86 (95% CI 0.78 to 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 RCT 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 a 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 physiotherapy for shoulder pain also demonstrated a bidimensional structure, however there was evidence of differential item functioning for some items in the disability subscale (e.g., 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 Spearmans' correlations to other instruments or measures, and the population it has been measured in are as follows:
Shoulder range of motion: Patients with shoulder pain: r = 0.55 to 0.80 (5).
Disabilities of Arm, Shoulder and Hand (DASH): Following shoulder arthroplasty: r = 0.93 (28); and participants in two clinical trials of treatments for adhesive capsulitis: r = 0.55 (35).
American Shoulder and Elbow Surgeons (ASES) Shoulder Score (ASES): following shoulder arthroplasty r = 0.81 (28); and patients referred to an upper extremity clinic for shoulder problems) r = 0.77 (36).

Simple Shoulder Test (SST): Patients referred to an upper extremity clinic for shoulder problems: r = 0.74 (36).

Constant Murley (CM) Shoulder scale: Following shoulder arthroplasty: r = 0.82 (28); degenerative or inflammatory shoulder disease referred for surgery: r = 0.53 (37); and patients enrolled in physical therapy for shoulder dysfunction: r = 0.56 (38).

Oxford Shoulder Score (OSS): Patients attending a specialist shoulder clinic with subacromial impingement): r = 0.85 (39); degenerative or inflammatory shoulder disease referred for surgery: r = 0.74 (37); shoulder problems in Turkish patients: Cronbach's r = -0.74 (and r = -0.71 and -0.72 for the pain and disability subscales respectively) (40)

Shoulder Disability Questionnaire (SDQ): Patients with shoulder pain in primary care: r = 0.57 (10).

Shoulder Rating Questionnaire (SRQ): Patients with shoulder pain in primary care: r = 0.83 (10).

Sickness Impact Profile (SIP): Patients with shoulder pain in outpatient physiotherapy clinic: r = 0.57 (41); and community volunteers with shoulder pain: r = 0.45 (42).

Short-Form 36 (SF-36) Physical Component Scale: patients with shoulder pain in the general community: r = -0.46 (8).

Health Assessment Questionnaire (HAQ): Participants in two clinical trials of treatments for adhesive capsulitis: r = 0.55 (35).

Responsiveness. The total SPADI score exhibits good responsiveness as measured by Area Under the Curve (AUC) of the Receiver Operating Characteristic (ROC) curve. AUC was 0.87 (95% CI 0.79 to 0.95) in subjects in primary care receiving non-surgical treatments for shoulder pain (10), ranged from 0.81 (95% CI 0.78 to 0.84) to 0.85 (95% CI 0.82 to 0.88) in subjects receiving physical therapy for shoulder pain (43), and from 0.74 (95% CI 0.64 to 0.83) to 0.85 (95% CI 0.76 to 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 (ES) and Standardized Response Means (SRMs) of the SPADI total score and subscales in various shoulder conditions and settings are as follows:

Total shoulder arthroplasty: SPADI: ES = 2.10, 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 to the ASES (ES = 2.13).

Adhesive capsulitis, trials of oral steroids and hydrodilatation: ES = 1.20 to 1.64, SRM = 1.27 to 1.68, and greater responsiveness of the SPADI was observed compared with the Croft Index (ES = 0.87 to 1.21) or DASH (ES = 0.55 to 0.83) (35).

Shoulder pain, physical therapy: ES = 1.26, SRM = 1.38 (41); and ES = 1.26, 1.71, SRM = 1.35, 1.75 at 6-week and 6-months respectively and similar to the QuickDASH (ES = 1.04, 1.40, SRM = 1.26, 1.56 at 6-week and 6-months respectively)(43).

Rotator cuff surgery or total shoulder arthroplasty: SRM = 1.23 (31). This was better than the SRQ (SRM = 0.65) and comparable to the SST (SRM = 1.05).

Various upper extremity disorders, occupational or physical therapy: ES 1.21, SRM 1.08, which was similar to the DASH (ES 1.06, SRM 1.08) and SF-36 Physical Component Summary (PCS) scale (ES 1.20, SRM 1.07) (45).

Spanish patients with shoulder pain/dysfunction after surgery for breast cancer: ES: 0.59, SRM = 0.82 (30).

Minimally important differences. 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 Minimally Clinically Important Difference (MCID) has been estimated to be 13.2 points (musculoskeletal upper extremity problems) (45); 20 points (Norwegian SPADI in patients receiving non-surgical 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 and has been used in populations with various shoulder disorders, primarily non-specific 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 randomised controlled trials of interventions for various shoulder disorders up to the end of 2015 (2). Recent examples include trials of therapeutic deep heat for shoulder pain (46), exercise versus physiotherapy 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, 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 as 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, therefore a change of at least the MDC (17-20 points) is required to be confident of a clinically-important change.

Clinical usability. Useful in a clinical context for assessing both joint-specific pain and function. Brief, easy to administer and score, and responsive to change.

Research usability. Brief, easy to administer, responsive. Valid for intervention and population-level studies.

American Shoulder and Elbow Surgeons (ASES) Society Standardized Shoulder Assessment Form

Description

Purpose. The ASES 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 (ADL) and include a patient self-evaluation section. A draft version was developed based upon a review of all relevant forms that existed at the time, and modified based upon two rounds of feedback from the ASES Society's membership. In 1998, the original ASES was modified to the modified ASES (mASES) 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 three domains: pain, ADLs and instability. The physician assessment comprises four domains: range of motion, signs, strength and instability. Most commonly however only the two domains of patient self-assessment of pain and ADLs is used.

Number of items. The full ASES comprises 18 patient self-evaluation items (6 pain, 10 ADLs, 2 instability) and 29 physician-assessed items (5 range of motion, 11 signs, 5 strength, 8 instability). There is also one item asking the patient to indicate where the pain is on a diagram and one open item asking the physician about physical findings. Most commonly only the 16 items evaluating patient-reported assessment of pain and ADLs is 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, Advil, Tylenol etc.)?; and 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 0-10 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; do usual sport) are 4-point ordinal Likert scales:0 - unable to do, 1 - very difficult, 2 - somewhat difficult, and 3 - not difficult). For the last two items (do usual work, 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 0-10 VAS (how unstable is your shoulder?) with anchors of 0 – very stable at all to 10 – very unstable.

Physician-assessed range of motion. 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 opposite acromion).

Physician-assessed signs. Four items (supraspinatus/greater tuberosity tenderness; AC joint tenderness; biceps tendon tenderness (or rupture); and other tenderness (list)) have 0 to 3 Likert scale responses (0 none, 1 -mild, 2- moderate, 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 is present, the respondent is asked to indicate the location and if 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, 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, 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 rim of glenoid). Four items (reproduces symptoms, voluntary instability, relocation test positive, generalized laxity) the response options are yes/no. The final item asks if there are any other physical signs with an open field.

Recall period for items. Patient-reported outcomes – not stated except for some items when the recall period is 'today'.

Cost to use. Free of charge.

How to obtain. 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
https://www.orthotoolkit.com/ases/

Practical Application

Method of administration. Both patient self-assessment and physician-assessment for the full ASES 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-evaluation 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 multiplied by 5/3. The formula is: [(10 – VAS pain score) x 5] + [5/3 x cumulative ADL score]. This derives a score out of a possible 0 – 100. Some studies also report separate ASES pain and function (ADL) subscale scores. The instability items, 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 to 100 – best). A missing rule or distinct cut-offs to reflect severity have not been published. Normative data are available based upon a sample of 343 patients (aged 6 to 87 years) from an outpatient orthopaedic center being seen for conditions unrelated to the shoulder (patients with prior shoulder problems 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 and above 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 three years) or joint arthroplasty. People without a history of shoulder pathology reported a higher (better) mean ASES compared to those with a history of a shoulder problem (96.7 (range 0-100) versus 93.0, Wilcoxon 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 versus 97 in men, p=0.03) and scores declined with increasing age.

Respondent time to complete. Less than 5 minutes (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 score computation is easy and can be implemented in any database.

Translations/adaptations. 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).

Psychometric Information

Floor and ceiling effects. Many studies have reported acceptable floor and ceiling effects across different patient populations including: planned surgery for rotator cuff disease or glenohumeral arthritis (0% floor and ceiling), instability (0% floor, 1.3% ceiling) (63); following shoulder arthroplasty (1% floor, 8% ceiling) (28, 55); impingement, adhesive capsulitis or glenohumeral arthritis (0% floor and ceiling) (56); rotator cuff disease (2.3% floor and ceiling) (64); and 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. Internal consistency of the ASES has been found to range from Cronbach's α 0.61 to 0.96 across various study populations with different shoulder condition including in mixed population referred for physical therapy including some post-surgery: 0.86 (67); mixed population with impingement, adhesive capsulitis and glenohumeral arthritis: 0.85 (56); instability: 0.61, rotator cuff disease: 0.64 and glenohumeral arthritis 0.62 (63); post shoulder arthroplasty 0.96 (55); various shoulder conditions, non-surgical: 0.89 to 0.92, post-surgery: 0.91 (68); mixed outpatient population: 0.88 (61); rotator cuff disease including tears or adhesive capsulitis: 0.813 (62); rotator cuff disease, glenohumeral or acromioclavicular arthritis and instability: 0.88 (95% CI 0.84 to 0.91) (60); and mixed population mainly rotator cuff tears: 0.91 (59); shoulder dysfunction: 0.794 (58). One study including 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 mixed population referred for physical therapy including some post-surgery (ICC = 0.84, 95% CI 0.75 to

0.91)(67); various shoulder conditions (non-surgical: ICC = 0.84 (95% CI 0.66 to 0.92), post-surgery: ICC = 0.91, 95% CI0.82–0.96)(68); impingement, adhesive capsulitis and glenohumeral arthritis (ICC = 0.91) (56); rotator cuff disease, instability and glenohumeral arthritis (ICC = 0.94, 95% CI 0.88 to 0.97)(63); post shoulder arthroplasty (ICC = 0.93, 95% CI 0.90-0.95)(55); rotator cuff disease including tears or adhesive capsulitis (ICC = 0.96, 95% CI 0.918 to 0.981) (62); mixed outpatient population (ICC = 0.94 (0.95 and 0.86 for pain and ADL subscales respectively) (61); rotator cuff disease, glenohumeral or acromioclavicular arthritis and instability (ICC = 0.83, 95% CI 0.70 to 0.90) (60); rotator cuff repair, arthroplasty or physical therapy impingement or adhesive capsulitis (CC = 0.82) (65); shoulder dysfunction (CC = 0.75) (58); and 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 good person reliability 0.86 (66), while one in patients with rotator cuff disease reported person reliability to be fair (0.48), and inferior to the SST (0.71, moderate) and PROMIS PF CAT (0.93, excellent)(64).

Validity. One study has examined the Spanish version of the pASES using both confirmatory factor and Rasch analysis (59). Factor loadings for confirmatory factor analysis were >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, consistent with its inclusion of both pain and function items (64). This compared with unexplained variance of only 4.5% for the PROMIS PF CAT and 8.4% for the SST. While the SST also includes pain and function item, only 2 out of 12 questions relate to pain in contrast to contributing 50% of the ASES final score.

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

SPADI: patients post arthroplasty 0.81 (28), 0.92 (55) and post total –0.942 and reverse – 0.932 hemiarthroplasty (70); mixed outpatient population: 0.82 (61);

Western Ontario Rotator Cuff Index (WORC): patients who were receiving physical therapy for impingement or after rotator cuff repair, acromioplasty or decompression surgery: 0.81 (71)

Western Ontario Stability Index (WOSI): mixed population with rotator cuff disease, isolated superior labral anterior posterior (SLAP) lesions, and shoulder instability after surgery: 0.15 (69)

Rowe, rating sheet for Bankart repair score: mixed population with rotator cuff disease, isolated superior labral anterior posterior (SLAP) lesions, and shoulder instability after surgery: 0.15 (69)

DASH: patients post arthroplasty 0.79 (28), 0.84 (55); patients with impingement, adhesive capsulitis and glenohumeral arthritis: 0.92 (56); shoulder dysfunction: 0.69 (58)

CM Scale: patients post arthroplasty 0.71 (28); mixed population with rotator cuff disease, isolated superior labral anterior posterior (SLAP) lesions, and shoulder instability after surgery 0.36(69); mixed population, the majority with rotator cuff tears: 0.62 (59); rotator cuff tears: 0.871(72)

SST: mixed population with rotator cuff disease, isolated superior labral anterior posterior (SLAP) lesions, and shoulder instability after surgery 0.35(69); mixed outpatient population including rotator cuff disease, glenohumeral or acromioclavicular arthritis and instability: 0.73 (60); rotator cuff disease: 0.536 (64); post total –0.89 and reverse –0.87 hemiarthroplasty (70)

UCLA: mixed population with rotator cuff disease, isolated superior labral anterior posterior (SLAP) lesions, and shoulder instability after surgery 0.38(69); arthroscopic rotator cuff population, pre-, 6, 12 and 24 months post-surgery): Very high correlation overall ($r = 0.91$, $P < .001$); moderate in the preoperative period ($r = 0.67$, $P < .001$); high at 6 months after surgery ($r = 0.87$, $P < .001$) and very high at 12 and 24 months ($r = 0.90$ and 0.92, $P < .001$)(73)

Rotator Cuff QOL: patients who were receiving physical therapy for impingement or after rotator cuff repair, acromioplasty or decompression surgery: 0.70 (71)

Single Assessment (or Alpha) Numerical Evaluation (SANE) score: Following rotator cuff repair: 0.75, revision: 0.88, SLAP repair: 0.78, overall: 0.8 (74); rotator cuff repair: 0.85, total shoulder replacement: 0.72; physical therapy: 0.82 (65)

Patient-Reported Outcomes Measurement Information System (PROMIS) Physical Function Computerized Adaptive Test (PF CAT): rotator cuff disease: 0.581 (64); planned surgery for rotator cuff tear: 0.43(75); impingement: 0.694, ASES function subscale: 0.664 and pain subscale: 0.426 (66)

Patient-Reported Outcome Measurement Information System Physical Function Upper Extremity Computer Adaptive Testing (PROMIS PFUE CAT): shoulder pain (excluding patients with prior rotator cuff surgery, shoulder surgery in last 6 months or partial rotator cuff tear): 0.72 (76); planned surgery for rotator cuff tear: 0.59(75)

Patient-Reported Outcome Measurement Information System Pain Interference Computer Adaptive Testing (PROMIS PI CAT): planned surgery for rotator cuff tear: -0.43(75); impingement: 0.729, ASES function subscale: 0.667 and pain subscale: 0.594 (66)

SF-36 bodily pain: patients with impingement, adhesive capsulitis and glenohumeral arthritis: 0.60 (56) and patients post arthroplasty: 0.65 (55); mixed population: 0.64 (61); shoulder dysfunction: 0.60 (58); mixed population with rotator cuff disease, isolated superior labral anterior posterior (SLAP) lesions, and shoulder instability after surgery 0.05 (69); mixed outpatient population including rotator cuff disease, glenohumeral or acromioclavicular arthritis and instability: 0.68 (60); mixed population, the majority with rotator cuff tears: 0.74 (59)

SF-36 PCS: patients post arthroplasty 0.64 (28, 55); patients with impingement, adhesive capsulitis and glenohumeral arthritis: 0.48 (56); mixed population referred for physical therapy including some after surgery: 0.40 (67); mixed population with rotator cuff disease, isolated superior labral anterior posterior (SLAP) lesions, and shoulder instability after surgery 0.20 (69); mixed outpatient population including rotator cuff disease, glenohumeral or acromioclavicular arthritis and instability: 0.57 (60); mixed population, the majority with rotator cuff tears: 0.65 (59)

SF-36 physical functioning: patients with impingement, adhesive capsulitis and glenohumeral arthritis: 0.47 (56); patients post arthroplasty 0.57 (55); mixed population referred for physical therapy including some after surgery: 0.41 (67); mixed population: 0.35

(61); shoulder dysfunction: 0.50 (58); mixed population with rotator cuff disease, isolated superior labral anterior posterior (SLAP) lesions, and shoulder instability after surgery 0.27 (69); mixed outpatient population including rotator cuff disease, glenohumeral or acromioclavicular arthritis and instability: 0.51 (60); mixed population, the majority with rotator cuff tears: 0.59 (59)
Clinician assessed component of the ASES: patients post arthroplasty 0.48 (28)

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

Responsiveness. The ASES has been found to have good or at least comparable responsiveness compared with other shoulder-specific measures. Reported ES and SRM 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: Standardised ES = 0.80, SRM = 0.75 (ES greater among those who received surgery versus other treatments) (59)
Calcific tendinitis: subacromial steroid: 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, glenohumeral arthritis: SRM = 1.11 (63)
Rotator cuff disease, isolated superior labral anterior posterior (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 among 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 upon 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: ASES function subscale: 12.01 points, pain subscale: 16.92 points (78)
Post arthroplasty (total, reverse or hemi): 20.9 points (79)
Post arthroplasty (total or reverse): 6.3 (95% CI 2.3 to 15.0) to 13.5 (95% CI 4.8 to 22.3) points depending upon 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 outcome from different surgical and non-operative treatments in people with varying shoulder conditions including glenohumeral arthritis, rotator cuff disease, shoulder instability and adhesive capsulitis.

Use in clinical trials. The ASES had been used in 43 randomised 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 five years including 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); 135 versus 155 degrees reverse arthroplasty prosthesis for rotator cuff arthropathy (92); liposomal bupivacaine vs. continuous peripheral nerve block following arthroplasty (93); glucocorticoid injection versus oral NSAIDs for adhesive capsulitis (94);
12-month exercise program versus usual are after rotator cuff repair (95); multimodal analgesia injection combined with glucocorticoid versus saline injection after arthroscopic rotator cuff repair (96); Interference screw versus and 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 intraarticular glucocorticoid for shoulder stiffness (101); triple-loaded single-row or 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 (PITT) (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 fractur (107); arthroscopy-assisted versus standard intramedullary nail fixation for diaphyseal humerus fractures (108); reverse total shoulder arthroplasty as primary versus revision procedure for proximal humerus fractures (109); proprioceptive exercises in addition to conventional physiotherapy for impingement (110); intraoperative platelet rich plasma versus local anesthetic injection after arthroscopic rotator cuff repair (111); arthroscopic suprapectoral versus open sub-pectoral 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); 0.5ml versus 1ml 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 infraspinatus remplissage for anterior shoulder instability (119); and 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. Mix of scales (binary, Likert, VAS). In the case where 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) used to derive an ASES score. This component is easy to understand, use and score.

Research usability. 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) (CS) Score

Description

Purpose. The CS was first described in a university thesis in 1986 and 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 (range of motion 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. There are four domains: pain, ADLs, shoulder range of motion and strength. Pain and ADLs are both patient-reported; range of motion and strength are measured by an examiner.

Number of items. There are 10 items in total (1 pain, 4 ADLs, 4 shoulder range of motion, 1 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 4-item Likert scale (none = 15 points, mild = 10 points, moderate = 5 points, severe = 0 points), but was later modified to a 15cm visual analogue scale (scored in centimeters, 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 0–2 scale (0 = disturbance every night, 1 = occasional disturbance, 2 = undisturbed). Work and recreational activities are each rated on a 15cm VAS in response to the question 'how much of your normal work does your shoulder allow?' and 'how much of your normal recreational activity does your shoulder allow?'. Anchors are 'none' and 'all'. Each VAS response is measured in 3cm quintiles to give a score of 0–4. Functional mobility is rated according to the level that the patient can 'use their hand comfortably' on a 5-item scale scored from 2–10 in 2-point increments ('up to waist' = 2, 'up to xiphoid' =4, 'up to neck' = 6, 'up to top of head' = 8, 'above head' = 10). Some versions include a further response category ('below waist'), allocated zero points (125).

Shoulder ROM is measured by the examiner as the maximum painless active movement in 4 planes (forward elevation, lateral elevation, external rotation, internal rotation). Each movement is allocated up to 10 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, >150 degrees), allocated 2 points for each increment from zero to 10 points. External rotation is measured according to the functional capacity to reach various anatomical landmarks, with 2 points allocated for each of 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', 'full elevation'. Internal rotation is measured by ability to reach anatomical landmarks with the thumb, in 2-point increments from zero to 10 points: 'lateral aspect of thigh', 'behind the buttock', 'sacroiliac joint', 'level of the waist', twelfth thoracic vertebra', interscapular level'.

Shoulder strength was originally described as the maximum resisted force with the arm at 90 degrees 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 maximum force generated in a 5 second effort, with the shoulder at 90 degrees abduction in the scapular plane and a pronated wrist. The score is the force in kilograms: 1 point per 0.5kg, to a maximum of 25 points.

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

Cost to use. Free of charge.

How to obtain. A PDF version that includes a standardised 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. 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 up to 20 points (sleep 0–2, work 0–4, recreation 0–4, functional mobility 0–10), range of motion cup to 40 points and strength up to 25 points) (122). The total score is the sum of these four subscales.

Score interpretation. 0 = worst and 100 = best function. Individual subscales are not reported separately. The original description of the instrument by Constant and Murley was as a composite measure of shoulder function and deliberately included specified that 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, 573 women, age range 11 to 87) who reported one painless shoulder and a further 115 healthy volunteers (56 men, 59 women, age range 20 to 69) 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 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. Requires time to perform, however some of the examination maneuvers (particularly range of motion) may be incorporated in the normal physical examination. Some special equipment is needed, including a goniometer to measure range of motion and a dynamometer or similar device for measuring strength. No special software is required. 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 orthopaedic 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 zero points for strength due to either pain or inability to reach the measurement position (37).

Reliability. Internal consistency (measured by Cronbach's α) has been estimated to range from 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), to 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 (intra-tester ICC 0.95 - 0.96) than when two different experienced clinicians performed the assessment (inter-tester ICC 0.84 - 0.87) (137).

The strength testing component of the CS is a potential source of measurement error due to lack of standardisation of the testing procedure and equipment. In a sample of healthy volunteers, however, both intra-tester and inter-tester reliability was high when assessed by senior physiotherapy students with experience in use of this tool (intra-tester ICC 0.90 - 0.98, inter-tester 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, considerably fewer than other measures including DASH and ASES, but similar to SPADI (140). The seven categories represented two ICF components: "(b) Body Functions" (sleep functions, sensation of pain, mobility of joint functions, muscle power functions) and "(d) Activities and Participation" (hand and arm use,

remunerative employment, 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 measured in) are as follows:
Standardized Index of Shoulder Function (FI2S): Various non-surgical and post-surgical shoulder disorders: r = 0.91 (142).
Shoulder Function Index (SFInX): Proximal humeral fracture - baseline: r = 0.89 (143).
OSS: Degenerative or inflammatory shoulder disease referred for surgery: r = 0.71 (37); patients undergoing surgery for rotator cuff disease: r = 0.65 (144); patients with a proximal humerus fracture: r = 0.53 (at 6 weeks) and 0.79 (at 12 weeks) (145).
ASES: Patients undergoing surgery for various shoulder disorders: r = 0.36 (69); patients undergoing surgery for rotator cuff tears: r = 0.87 (72); patients enrolled in physical therapy for shoulder dysfunction: r = 0.495 (38).
Penn Shoulder Score: Patients undertaking physical therapy for shoulder disorders: r = 0.85 (54).
SST: Patients undergoing surgery for various shoulder disorders: r = 0.52 (69); (after rotator cuff surgery): r = 0.49 (72) to 0.70 (146); patients enrolled in physical therapy for shoulder dysfunction: r = 0.65 (38).
UCLA: Patients with a proximal humerus fracture: r = 0.70 (at 6 weeks) and 0.92 (at 12 weeks) (145); patients undergoing surgery for various shoulder disorders: r = 0.67 (69); after rotator cuff surgery: r = 0.66 (146); patients enrolled in physical therapy for shoulder dysfunction: r = 0.59 (38).
SPADI: Degenerative or inflammatory shoulder disease referred for surgery: r = 0.53 (37); (patients enrolled in physical therapy for shoulder dysfunction: r = 0.56 (38).
DASH: Degenerative or inflammatory shoulder disease referred for surgery: r = 0.50 (37); patients undergoing surgery for rotator cuff tears) r = 0.76 (72); patients with a proximal humerus fracture: r = 0.62 (at 6 weeks) and 0.86 (at 12 weeks) (145); in German-speaking Swiss patients following shoulder arthroplasty r = 0.82 (28).

Western Ontario Shoulder Instability (WOSI) Index: Patients undergoing surgery for various shoulder disorders: r = 0.18 (69); patients with symptomatic shoulder instability r = 0.59 (147).

WORC: Rotator cuff disease: r = 0.56 (148) to 0.65 (149).

Bostrom's shoulder movement impairment scale: Degenerative or inflammatory shoulder disease referred for surgery: r = 0.78 (37).

Shoulder Function Assessment (SFA) scale: Degenerative or inflammatory shoulder disease referred for surgery): r = 0.86 (37).

Short-Form 36 (SF-36) Physical Component Score: Patients undergoing surgery for various shoulder disorders: r = 0.41 (69).

Agreement between the modified CS and patient-reported CS. In patients referred for shoulder surgery, mean scores were similar for clinician-assessed (mean 48, SD 20) and patient-assessed CS (mean 47, SD 19.5). Agreement between clinician assessors and patients was high for ROM (weighted Kappa 0.8-0.9) (124). In French patients with various shoulder disorders, correlation between the total CS score assessed by an orthopaedic surgeon and the 'Auto-Constant' assessed by the patient was ICC = 0.87 (123). Of the 4 subscales, the ICC was lowest for the strength subscale (0.57) but highest for the ROM subscale (0.85).

Responsiveness. Reported ES and SRM in various shoulder conditions and settings are as follows:

Patients with rotator cuff disease seen in a tertiary orthopaedic clinic after 6 months: SRM = 1.38 (149); one week after completion of a course of physical therapy plus acupuncture or mock TENs for painful rotator cuff disease: ES = 0.4 (150); after subacromial decompression surgery SRM = 1.12 and 2.09 at 3 and 6 months, ES = 1.23 and 1.92 at 3 and 6 months (151); six months following surgery for various shoulder disorders: SRM = 0.58, ES = 0.57 (69); two to five years after total shoulder arthroplasty for osteoarthritis: SRM = 2.4, ES = 2.9 (134); six months after total shoulder arthroplasty: SRM = 1.99, ES = 2.23 (44); and at three 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 one 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 three months of physical therapy following arthroscopic decompression surgery in Danish patients (139); and 17 points after three 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 reported up to December 2015 (2). Examples of trials which have used the CS as an outcome measure in the last five years include various surgical interventions (for conditions including rotator cuff tears (156), proximal humeral fracture (157), 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 community 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. 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 or inconsistent application of testing methods have the potential to result in excessive inter-rater variability. There are limited data on 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 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 inter-tester reliability) are an important caveat for use in a research setting.

Simple Shoulder Test

Description

Purpose. To assess functional limitations of the shoulder relative to the patient's activity of daily living before or after treatment, and work (163). Questions were derived from Neer's evaluation (164), the American Shoulder and Elbow Surgeons (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: There are three domains: pain, function/strength and range of motion

Number of items: 12 items (2 pain, 7 function/strength, 3 range of motion).

Response options/scale: All yes/no responses for each item.

Recall period for items: At the moment of assessment.

Cost to use: Free of charge for non-commercial use.

How to obtain. Original publication (163). Free online at https://orthop.washington.edu/patient-care/articles/shoulder/simple-shoulder-test.html

Practical application

Method of administration. Self-assessment.

Scoring. Original score: 0 = worst and 12 = best. Transformed by summing the number of 'yes' answers/number of completed items X 100 = percent 'yes' responses.

Score interpretation. 0 = worst and 100 = best function. A missing rule or distinct cutoffs for severity, and normative data have not been published. There are no subscales.

Respondent time to complete. 2–3 minutes.

Administrative burden. Free online. Score computation is very easy and possible by hand. No software needed. Time to administer and determine 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 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 2% and 9.3% of patients with shoulder instability in one study (172); 21% floor and 6.1% in a study of people with rotator cuff disease (172); and no floor and 15.3% ceiling effects in patients who had undergone shoulder arthroplasty (173). No floor or ceiling effects were found in Iranian patients with a variety of shoulder disorders (169) while 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: 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

to 0.86) (167); Dutch patients with shoulder problems attending an orthopaedic clinic: Cronbach's α = 0.78 and 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); 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 while 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 SST from a sample of patients with shoulder complaints revealed a 2-factor solution which explained 52.6% of the variance (29). Two items relating to 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 it 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, i.e., larger than the ASES and Shoulder Pain and Disability Index (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 3-factor solution, mainly related to activities performed with the arm at shoulder level (arm elevation), shoulder movement and arm strength, and which explained 49.7% of variance, was identified in the Persian STT (169). The Cronbach's α for these three factors were 0.7, 0.53, and 0.6 respectively. For the Brazilian Portugese STT tested in patients with various shoulder problems a 3-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 tested also tested in patients with various shoulder problems a 3-factor structure explained 56.12% of the variance with Cronbach's α = 0.73, 0.72, and 0.66 for the three factors respectively, and 0.793 overall (168).

Test-retest reliability of the SST is high. One study reported ICCs of 1 (95% CI 1 to 1) and 0.97 (95% CI 0.91 to 0.99) measured in patients with instability and rotator cuff disease respectively (172); in a study of patients attending a surgical clinic: ICC = 0.99 (31); in studies of Dutch patients with shoulder complaints: ICC = 0.86 (175) and 0.92 (166); in Italian patients after surgery for instability: ICC = 0.97 (170); in patients with various shoulder problems in Iran: ICC= 0.94 (95% CI 0.86 to 0.97 (169), in Spain: 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 measured in) are as follows:
SPADI: Various shoulder problems: ICC = 0.74 and 0.80 (29, 36); Italian patients after shoulder surgery for anterior instability: ICC = 0.68; pain subscale ICC = 0.63 and function subscale ICC = 0.72 (170)
ASES: Various shoulder problems: 0.73 (36) and 0.536 (64); rotator cuff diseases: 0.76 (172); instability: 0.89 (172); Italian patients after shoulder surgery for anterior instability: ICC = 0.8 (170).
OSS: general shoulder problems in Dutch patients: ICC = 0.61 (176) and ICC = 0.74 (166), and Iranian patients ICC = 0.68 (169); Italian patients after shoulder surgery for anterior instability: ICC = 0.67 (170)
DASH: Rotator cuff surgery: 0.72 (42); various shoulder problems in Spanish patients:ICC = 0.73 (168) and Dutch patients: ICC = 0.74 (166)
CS: Rotator cuff surgery: 0.70 (146)
Western Ontario Rotator Cuff (WORC) index: Rotator cuff surgery: 0.68 (42), various shoulder problems: SF-36 bodily pain: 0.62 (36)
UCLA: Italian patients after shoulder surgery for anterior instability: pain subscale ICC = 0.61, function subscale ICC = 0.51 (170)
WOSI: Italian patients after shoulder surgery for anterior instability: ICC = 0.63 (170).
Patient-Reported Outcomes Measurement Information System (PROMIS) Physical Function Computerized Adaptive Test(PF CAT): Patients been seen by a shoulder and elbow surgeon: 0.635 (64)
SF-36 physical functioning: various shoulder problems: 0.58 (36) and 0.56 (166)
SF-36 PCS: various shoulder problems: 0.40 and rotator cuff surgery: 0.60 (36, 42);

SF-12 PCS: various shoulder problems: 0.44 (172) and in Spanish patients: 0.43 (168)
SF-36 Mental Component Summary (MCS) score: Rotator cuff surgery: 0.16 (42)

Responsiveness. Effect sizes (ES) and standardized response means (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 + 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. Minimally detectable change (MDC95%) across various shoulder problems is 32.3% (0-100% scale)(29), after surgery for shoulder instability in Italian patients 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 to 100% scale) (168); in Iranian patients with shoulder disorders: MDC = 3.7 (0 to 12 scale)(169); and in Dutch patients with various shoulder problems: MDC 3.3 (0 to 12 scale (166).

MCID is reported for the following shoulder conditions:
Rotator cuff disease using fifteen-item function question as anchor question:2.05 and using four-item improvement question as anchor question: 2.33 points (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 17.1–25.0 (0–100 range) (175).

Generalizability. This measure is shoulder-specific but has also been used to measure shoulder function in patients with other conditions, e.g., 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 outcome in the last five 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 and 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. Very short and easy to use. Good construct validity with function subscales and good reliability, and acceptable floor and ceiling effects.

Caveats and cautions. Due to the binary response options, the use of the SST score as a metric measure is questionable. In the case where 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 SST with other shoulder specific measures, rather than with generic measures, SST can measure shoulder function, but cannot infer overall health of the shoulder.

Clinical usability. Easy to use; widespread use in the US and also translated and culturally adapted to many different countries.

Research usability. This measure is widely used due to its convenience and ease of use. However, we recommend it is 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 (OSS)

Description

Purpose. The OSS was developed to be a **s**elf-assessment of pain and function of the shoulder assessing outcome from shoulder surgery 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 upon interviews with patients attending an outpatient shoulder clinic as well as from 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. There are two domains: pain and daily functions. There are no subscales.

Number of items. 12 items (2 pain, 2 interference with pain and 8 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 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. 4 weeks.

Cost to use. Free of charge for non-commercial 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 to 4) but the scores are reversed, with score ranges from 0 = worst to 48 = best (182). There are no subscales.

At least 10 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 gap/s. 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 worst (most severe) response is adopted.

Score interpretation. In the original OSS, a score of 12 indicates no disability while a score of 60 indicates maximal disability. In the revised OSS (and online form), a score of 0 indicates maximal disability while a score of 48 indicates no disability. A score of 0–19 indicates severe arthritis, 20–29 moderate to severe arthritis, 30–39 mild to moderate arthritis, and 40–48 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 three 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 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 genders adjusted for nationality (mean 47 in both). Mean score declines with increased age.

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

Administrative burden. Score computation is easy and needs no explanation. No training is needed to administer and interpret the scores. Can be completed using pen and paper. Time to administer and score ~ 5 minutes.

Translations/adaptations. The OSS has been **t**ranslated and culturally adapted to 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 above in both original and translated versions: pre-operative mixed population excluded patients with instability: Cronbach's α = 0.89; subset of same population 6 months following various types of surgery: Cronbach's α = 0.89; all items correlated with the total score (r > 0. 4) (181); impingement syndrome with or without rotator cuff tear or calcific tendinitis: Cronbach's α = 0.94 (187); mixed population of patients (most with rheumatoid arthritis or osteoarthritis) undergoing glenohumeral or acromioclavicular surgery: Cronbach's α = 0.843 (37); Spanish patients with shoulder pain/dysfunction after surgery for breast cancer: Cronbach's α = 0.947 (30); current shoulder symptoms in Persian patients: Cronbach's α = 0.91 and if an item deleted it ranged from 0.91 to 0.92 (169); impingement or calcific tendinitis in German patients: Cronbach's α = 0.94 and Cronbach's α not below 0.93 when single items were eliminated (187); degenerative or inflammatory shoulder conditions in Italian patients: Cronbach's α = 0.95 (185); shoulder problems in Turkish patients: Cronbach's α = 0.92 (40), French patients: Cronbach's α = 0.93 (189) and Chinese patients (190); post rotator cuff surgery in Polish

patients: Cronbach's α = 0.96 (192); degenerative or inflammatory shoulder conditions in Korean patients: Cronbach's α = 0.91 (183) and Brazilian patients: Cronbach's α = 0.90 (186); Brazilian patients with rheumatoid arthritis: Cronbach's α = 0.957 (188); and in Romanian patients with rotator cuff disorders or proximal humerus fractures: Cronbach's α = 0.954 (184).

Factor analysis of the Persian OSS found a 2-factor solution which explained 61.6% of variance which is consistent with it being a two-dimensional outcome measure assessing pain and function (169). Principal component analysis of the Chinese OSS found a one factor structure which accounted for 54.2% of total variance (190).

Test-retest reliability is also reported to be acceptable: pre-operative 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 kappa = 0.13 for item 1, weighted kappa values for other items = 0.44-0.79 (39). Assessment of the Spanish OSS in patients with shoulder pain/dysfunction after surgery for breast cancer reported an ICC of 0.974 (30); Persian OSS in patients with current shoulder symptoms: ICC = 0.90 (95% CI 0.77 to 0.95)(169); Danish OSS in patients with shoulder symptoms: ICC = 0.98 (191); German OSS in patients with impingement or calcific tendinitis: ICC = 0.98 (187); Dutch and French versions of the OSS in patients with general shoulder problems: ICC = 0.981 (176) and 0.91 (95% CI 0.88 to 0.94) (189) respectively; Polish OSS post rotator cuff surgery: ICC = 0.99 (192); Korean OSS in degenerative or inflammatory shoulder conditions: ICC = 0.95 (183); Chinese OSS in non-specific shoulder pain patients: ICC = 0.97 (190); rotator cuff disorders and proximal humerus fractures in Romanian patients: ICC = 0.953 and ICC = 0.976 as the average ICC of the office visit and two days from the office visit (184); In Turkish patients with mixed shoulder complaints: ICC = 0.99 (40); Brazilian patients with degenerative or inflammatory shoulder conditions: ICC = 0.92 (186) or rheumatoid arthritis: ICC = 0.917)(188).

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

CM: Mixed population of shoulder conditions pre- and post-surgery: -0.74 and -0.75 respectively (181); impingement with or without rotator cuff tear or calcific tendinitis: 0.66 (187), mixed population of patients (most with rheumatoid arthritis or osteoarthritis) undergoing glenohumeral or acromioclavicular surgery: 0.71 (37); proximal humeral fracture treated conservatively: 0.84, 0.77 and 0.87 at baseline and 3 and 12 months follow up (193); impingement or calcific tendinitis in German patients: 0.60 (187); general shoulder problems in Dutch: ICC = 0.64 (176), French: 0.73 (189) and Chinese patients: 0.66 (190); degenerative or inflammatory shoulder conditions in Italian patients: ICC = 0.73 (185); Danish patients with shoulder complaints: ICC = 0.74 (191); rotator cuff disease prior to surgery and 6 months post surgery: ICC = 0.65 and 0.65 (144); and Brazilian patients with rheumatoid arthritis (188). Correlation with the ADL, strength and range of motion subscales of the CM in Korean patients with degenerative or inflammatory shoulder conditions was ICC = 0.68, 0.48 and 0.42 respectively (183).

SPADI: mixed population of patients (most with rheumatoid arthritis or osteoarthritis) undergoing glenohumeral or acromioclavicular surgery: 0.74 (37); subacromial impingement: 0.85 (39); Spanish patients with shoulder pain/dysfunction after surgery for breast cancer: -0.674 (30): mixed shoulder complaints in Turkish patients: total, pain and disability subscales: ICC = -0.74, 0.71 and 0.72 respectively (40); patients attending a specialist shoulder clinic with subacromial impingement): r = 0.85 (39); and Italian patients after surgery for anterior instability: ICC = 0.79 (170).

DASH: mixed population of patients (most with rheumatoid arthritis or osteoarthritis) undergoing glenohumeral or acromioclavicular surgery: ICC = 0.79 (37); patients with degenerative or inflammatory shoulder conditions in Brazil: ICC = 0.77, DASH work: ICC = 0.76, DASH sport/music: ICC = 0.62 (186) and in Korea DASH disability/symptoms ICC = 0.81,DASH work: ICC = 0.69, and DASH sport/music: ICC = 0.59 (183). QuickDASH: post rotator cuff surgery in Polish patients: -0.92 (192); rotator cuff disorders and proximal humerus fractures in Romanian patients: ICC = 0.633 and 0.672 two days later (184).

SST: current shoulder symptoms in Persian patients: 0.68 (169); general shoulder problems in Dutch patients: ICC = 0.61 (176) and ICC = and 0.74)(166); and in Italian patients after surgery for anterior instability: 0.67 (170).

Shoulder Function Assessment scale: mixed population of patients (most with rheumatoid arthritis or osteoarthritis) undergoing glenohumeral or acromioclavicular surgery: ICC = 0.72 (37)

Subjective Shoulder Value (SSV): general shoulder problems in French patients: ICC = 0.68 (189).

UCLA: degenerative or inflammatory shoulder conditions in Italian patients: ICC = 0.67 (185)

SF-36 bodily pain: Mixed population of shoulder conditions pre- and post-surgery: ICC = -0.66 and -0.68 respectively (181); subacromial impingement: ICC = 0.69 (39); impingement with or without rotator cuff tear or calcific tendinitis: ICC = 0.76 (187); post rotator cuff surgery in Polish patients: 0.81 (192); rotator cuff disease prior to surgery and 6 months post surgery: 0.64 and 0.75 (144); non-specific shoulder pain in Chinese patients: ICC = 0.53 (190); degenerative or inflammatory shoulder conditions in Italian patients: ICC = 0.66 (185); mixed shoulder complaints in Turkish patients: ICC = 0.74 (40); Brazilian patients with rheumatoid arthritis: 0.624 (188); and Italian patients after surgery for anterior instability: pain subscale 0.7 and function subscale 0.54 (170).

SF-36 physical function: Mixed population of shoulder conditions pre- and post-surgery: -0.61 and -0.62 respectively (181); impingement with or without rotator cuff tear or calcific tendinitis: 0.62 (187); subacromial impingement: 0.57 (39); post rotator cuff surgery in Polish patients: 0.82 (192); non-specific shoulder pain in Chinese patients: ICC = 0.65 (190); rotator cuff disease prior to surgery and 6 months post surgery: -0.57 and 0.68 (144); mixed shoulder complaints in Turkish patients: ICC = 0.63 (40); degenerative or inflammatory shoulder conditions in Italian patients: ICC = 0.74 (185); and in Brazilian patients with rheumatoid arthritis: ICC = 0.589 (188).

SF-36 physical component summary: mixed shoulder complaints in Turkish patients: 0.63 (40); post rotator cuff surgery in Polish patients: ICC = 0.82 (192).

Health Assessment Questionnaire (HAQ): Mixed population of shoulder conditions pre- and post-surgery: pain subscale 0.49 and 0.71 and function subscale 0.86 and 0.80 respectively (181); Brazilian patients with rheumatoid arthritis: 0.663 (188).

One study assessed correlation of the OSS with other shoulder measures in people without current shoulder complaints: ASES: r = 0.71; CM: r = 0.44; UCLA: r = 0.8; SPADI: r = 0.78; and Stanmore Percentage of Normal Shoulder Assessment SPONSA): r = 0.53 (9).

Responsiveness. Effect sizes (ES) and standardized response means (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: osteoarthritis with intact rotator cuff: ES 2.9, osteoarthritis with a torn rotator cuff: ES 2.1 and rheumatoid arthritis: ES 3.1 (194); mixed population of patients undergoing subacromial decompression with or without rotator cuff repair: ES ranged from 1.88 (full rotator cuff repair), 0.14 (partial repair), 1.02 (no repair), 1.55 (no tear)(195); subacromial impingement (regardless of treatment): ES: baseline to 6 weeks and 18 weeks: ES 0.24 and 0.96 (39); mixed population of patients (most with rheumatoid arthritis or osteoarthritis) undergoing glenohumeral or acromioclavicular surgery: ES 0.61 (37); and in Spanish patients with shoulder pain/dysfunction after surgery for breast cancer: ES 0.50, 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) while 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 MCID of 6 points (OSS scale 0 to 48 points) was reported in a study including patients with shoulder problems attending an orthopedic outpatient clinic (baseline and 6 months follow up, both surgery and non-operative treatment, excluding fractures and frozen shoulder) (175). Another study also reported an MCID of 6 points that included patients following subacromial decompression (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 0-100 scale was reported in Persian patients with general shoulder problems (169).

Generalizability. Since the previous review there have been many further 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 outcome from different surgical and non-operative 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 randomised controlled trials (2). The OSS continues to be used as an outcome in trials published in the last five years including: surgical versus non-surgical treatment for proximal humerus fractures (197); intra-operative platelet-rich plasma versus local anesthetic injection for arthroscopic repair (111); ultrasound added to exercise and mobilization for frozen shoulder (198); supervised 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 sub-pectoral 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 non-operative 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 to 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. Construct validity to other measures is good and there is no cost for non-commercial purposes. This tool has been translated and culturally adapted to many countries.

Caveats and cautions. As the MDC is similar to the MICD, caution is necessary for measurement 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 assessment of shoulder surgery. Scores are easy to interpret.

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

DISABILITIES OF THE ARM, SHOULDER, AND HAND QUESTIONNAIRE (DASH) AND ITS SHORT VERSION (QUICKDASH)

Description

Purpose. The DASH was developed for self-assessment of symptoms and function of the entire upper extremity in patients with pain in the arm, shoulder or hand (207). Draft items were identified by literature review (821 reduced to 67 (+3 new) due to content overlap or off target by a consensus group). 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 (9 items) in 2009 (212). All relevant modern strategies were used in the development of the QuickDASH comparing 3 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 6 items about symptoms (3 pain, 1 tingling/numbness, 1 weakness, 1 stiffness) and 24 about function (21 physical

function, 3 social/role function). Two optional additional modules for work (4 items) and sports/performing arts (4 items) are more rarely used in patient settings, but are used when assessing manual workers and athletes. The QuickDASH has 11 items (3 symptoms, 8 function)(211). The QuickDASH-9 has 9 items (1 pain, 8 for function) (212), 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 5 (Likert) levels: 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. 1 week.

Cost to use. Free of charge for non-commercial use; license for commercial use available at the Institute for Work and Health (IWH). Details on commercial use licence and costs can be found online (http://www.dash.iwh.on.ca/licences). Manual (3rd edition) online and paper copy (http://www.dash.iwh.on.ca/dash-manual).

How to obtain. Property and copyright 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. Self-assessment, with paper and pencil or computer.

Scoring. The arithmetic mean of at least 27 of the 30 items (missing rule) is transformed by (mean - 1) x 25 into the scale from 0 = no symptoms/full function to 100 = maximal symptoms/no function for the DASH total score (207). Five of 6 items are necessary for determination of the symptoms score and 22 of 24 items for the function score. Similarly, 10 of 11 items are necessary for the QuickDASH total score, 3 of 3 for symptoms, and 7 of 8 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 easier, e.g., on Microsoft Excel or any calculation or statistics program. Scoring program is 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 by (100 - original score) is also often used for comparison with other scores such as the SF-36. Several studies have showed varying distinct cut-off points to reflect severity. One example is cut-off scores of <15 = “no problem,” 16–40 = “problem, but working,” and >40 = “unable to work” (207).

Normative values of 1,706 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. 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 (<10%)(209) other than item 21, on sexual activity, which has high level of missingness (up to 30%) (211, 216, 217).The questionnaire contains instructions on how to complete it. Time to administer (including control of missing data): DASH, 10 minutes; QuickDASH, 8 minutes. Little special training is necessary for these activities.

Translations/adaptations. Available for free for >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, while 2 studies showed 0.5% and 1.8% ceiling effects for the DASH and 3% for the QuickDASH, and one study reported 2% of floor effects (28, 209, 210, 216, 218-220).

Reliability. The DASH has high internal consistency with Cronbach's α = 0.92–0.98 for the DASH (207, 209, 210, 212, 221) and 0.92–0.95 for the QuickDASH (211, 222). High reported Cronbach's α may point towards some redundancy in the DASH and Rasch analyses have shown some problems regarding the unidimensionality of the DASH score, containing 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 provide evidence that respondents were unable to discern the 5 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).

Rash 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, as the patients' responses were not consistent with those predicted by the Rasch model (209). The corresponding results for the QuickDASH were better (211), but 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 measured in, are as follows:

SPADI: Patients who had undergone shoulder arthroplasty: r = 0.93,(28); patients with glenohumeral arthritis or soft tissue disorders around the shoulder, predominantly rotator cuff tendinitis: r = 0.83 (SPADI function, no overall SPADI assessed)(216); patients with adhesive capsulitis, pre-treatment and 3 weeks after arthrographic joint distension: r = 0.55 at baseline and r = 0.50 for 3-week change scores (35); impingement: r = 0.75,(220); subacromial pain syndrome referred to physical therapy: r = 0.82 for the 3-4-month change score (226).

CM: Patients who had undergone shoulder arthroplasty: r = 0.82 (28)

Short Form-36 physical component summary (SF-36 PCS): post shoulder arthroplasty: r = 0.67,(28); post rotator cuff tear repair: r = 0.50 for 6-month change scores (7); impingement: r = -0.59 (220); rheumatoid arthritis patients with shoulder complaints: r = -0.70 (218).

SF-12 PCS: proximal humeral fracture: r =-0.75 (224); 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). r = 0.62 (pain in shoulder, arm and hand) (227).

SF-36 MCS: post shoulder arthroplasty: r = 0.06 (28); post rotator cuff tear repair: r = 0.29 for 6-month change scores (7); impingement: r = -0.17 (220); rheumatoid arthritis patients with shoulder complaints: r = -0.27 (218).

SF-12 PCS: 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), r = 0.15 (pain in shoulder, arm and hand) (227)

ASES: Post shoulder arthroplasty: r = 0.79 and r = 0.59 for the c-ASES (28)

HAQ: Patients with adhesive capsulitis, pre-treatment and 3 weeks after arthrographic joint distension: r = 0.54 at baseline and r = 0.35 for 3-week change scores (35); rheumatoid arthritis patients with shoulder complaints: r = -0.88 (218)
SST: Patients undergoing operative repair of a rotator cuff tear: r = 0.72 for 6-month change scores (7)
WORC: Patients undergoing operative repair of a rotator cuff tear: r = 0.71 for 6-month change scores (7)
EuroQol-5D (EQ-5D): Patients with a proximal humeral fracture: r =-0.75 (224)
28-Joint Disease Activity Score (DAS28): rheumatoid arthritis patients with shoulder complaints: r = 0.42 (218)
Global disability rating: Patients with shoulder problems: r = 0.71, r = 0.67 3-month change after physiotherapy (45)

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

Responsiveness. Reported effect sizes (ES) and standardized response means (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); shoulder problem, physiotherapy: ES 1.06, SRM 1.08 points (45); glenohumeral arthritis or soft tissue disorders around the shoulder, predominantly rotator cuff tendinitis, surgical intervention: ES 0.64, SRM 0.81 points (216); adhesive capsulitis, arthrographic joint distension: patients receiving treatment of known efficacy, ES 0.58, SRM 0.96 points; patients reporting 'marked or moderate' improvement, ES 0.55, SRM 0.91 points; patients reporting 'marked improvement', ES 0.83, SRM 1.45 points (35).

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

Minimally important differences. Reported MDC for the DASH ranges between 12 and 19 points: patients with shoulder problems treated with physiotherapy: 12.2 points (45); glenohumeral arthritis or rotator cuff tendinitis treated surgically: 12.75 points (216); impingement, no treatment: 13.1 points (220) or physical therapy: 11.8 points; humeral shaft fracture treated both operative and non-operatively: 19 points (229); shoulder complaints, surgical and non-surgical interventions: 16.3 points (175).

Reported MCIDs range between 4.4 and 12.4, notably less than reported MDCs: humeral shaft fracture treated both operative and non-operatively: 6.7 points (229); impingement, physical therapy: 4.4 points (226); shoulder complaints, surgical and non-surgical interventions: MIC 12.4 points (175); shoulder problems treated with physiotherapy: 10.2 points (45).

MDC and MCID reported for the QuickDASH include: Shoulder pain, physiotherapy: MDC 11.2 percentage points, MCID 8.0 points (225), shoulder complaints, surgical and non-surgical interventions: SDC 17.1 points, MCID 13.4 points (175).

Generalizability. The DASH has been used to assess outcome from various non-operative 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 up until 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 shoulder and other disorders of the upper extremity. It is particularly

useful where measurement of symptoms and function of the entire upper extremity is wanted. Since 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 to each other.

Caveats and cautions. The DASH covers the whole upper extremity and not only the shoulder. Its specificity and responsiveness is generally lower than shoulder-specific tools, although higher than generic quality of life tools. This limits its usefulness in assessing outcome from shoulder-specific conditions. The DASH includes items relating to the lower part of the limb where (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 disability of the lower extremity. In addition Rasch analysis has revealed scoring issues and item misfit. Compared to other instruments, the strict 90% missing rule produces a relatively high percentage of missing data.

Clinical usability. The DASH is a good tool for comprehensive assessment of upper extremity conditions, i.e., if shoulder problems cannot be differentiated from hand problems. It is easy to apply, analyze and interpret. Comparison of empirical and normative data allows 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 comparison of different settings and different upper extremity instruments, especially for analysis of construct validity compared to other instruments. However concerns about its lack of specificity to shoulder function and its relatively limited responsiveness indicate it is not the instrument of choice for shoulder function.

University of California–Los Angeles (UCLA) 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 outcome of patients with other shoulder conditions undergoing surgery or non-operative 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 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), and 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 includes 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 five domains: pain, function, active forward elevation, strength and patient satisfaction (after surgery). Other versions include one with two additional domains (abduction range of motion and abduction strength)(236); inclusion of patient satisfaction item to the original UCLA (146); exclusion of patient satisfaction (as it asks about satisfaction of treatment so is not suitable as a measure of status)(237); and an Italian version which includes only two pain and function (25).

Number of items. Vary between three items (original UCLA)(233) and seven items (modified version with addition of two new items for abduction (236). The most commonly used version, the modified UCLA scale, as well as the patient-reported UCLA 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 upon categorical responses (233). Pain: Constant, unbearable; strong medication frequently (1 point); Constant, but bearable; strong medication occasionally (2 points); None or little at rest; occurs with light activities; salicylates (4 points); With heavy or particular activities only; salicylates occasionally frequently (5 points); Occasional and slight (8 points); and No pain (10 points). 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, driving (5 points) Slight restriction

only; able to work above shoulder level (8 points); and Normal activities (10 points). Muscle power and motion: Ankylosis with deformity (1 point); Ankylosis with good functional position (2 points); Muscle power poor to fair; elevation less than 60°, internal rotation less than 45° (4 points);Muscle power fair to good; elevation 90°, internal rotation 90° (5 points); Muscle power good or normal; elevation 140°, external rotation 20° (8 points); Normal muscle power; 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: Present always and unbearable, strong medication frequently (1 point); Present always but bearable, strong medication occasionally (2 points); None or little at rest, present during light activities, salicylates used frequently (4 points); Present during heavy or particular activities only, salicylates used occasionally (6 points); Occasional and slight (8 points); and None (10 points). Function options are: Unable to use limb (1 point); Only light activities possible (2 points); Able to do light housework or most activities of daily living (4 points); Most housework, shopping, and driving possible, able to do hair and to dress and undress, including fastening bra (6 points); Slight restriction only, 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 two response options for satisfaction of patient. The six response options for active forward flexion are 150° (5 points); 120°-150° (4 points); 90°-120° (3 points); 45°-90° (2 points); 30°-45° (1 point); and <30° (0 points). Strength of forward flexion (manual muscle testing) 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). Satisfaction of patient: satisfied and better (5 points) or not satisfied and worse (0 points).

In the patient-reported UCLA, the response options for pain, function and satisfaction are the same (235). For active forward flexion the patient is asked how high they can lift their arm forwards, with the degrees shown pictorially. For 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 the active forward flexion and strength of forward flexion items (236).

Recall period for items. Not specified in the original version but the current online modified UCLA version specifies the past 4 weeks (see below).

Cost to use. Free of charge.

How to obtain. The link for the modified UCLA scale can be downloaded for free from the following websites including 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, 170) and a Brazilian Portugese version is published in ref (238).

Practical application

Method of administration. In the original UCLA 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 each of the three items is scored out of a maximum of 10 points (233). It was noted in this version that occasionally analysis of pain, function and range of motion 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: 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; https://www.orthotoolkit.com/UCLA-shoulder/

Score interpretation. In the original UCLA scale a score greater than 8 for pain, function and range of motion was considered to be excellent, >6 good, 2 to 4 fair and <3 poor (233). In the modified and patient-reported versions of the UCLA, a score of 34 or 35 is considered excellent, 28 to 33 good, 21 to 27 fair, and 20 or less points is considered poor (234, 235). Good or excellent scores (above 27) indicate a satisfactory result while fair or poor scores (27 and below) indicate unsatisfactory results. One study used cut offs of 34 or 35, 29 to 33 and <29 to indicate excellent, good and poor results respectively (239). In the 7-item version of the UCLA scale (range of possible scores 2 to 45), scores of 41-45 and 36 to 40 indicate excellent and good scores respectively (236).

Normative data for the modified 4-item UCLA scale (leaving out the patient satisfaction item)(range 2-30) has 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 score compared to those with a history of a shoulder problem (28.9 (range 2-30) versus 28.3, Wilcoxon 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 genders adjusted for nationality (women mean of 28 versus 29 in men, p=0.4768). Mean score declined slightly with increased age after age 50-59. 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 4-item UCLA scale in 120 shoulder throwing and non-throwing 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 normalised score was 98% (95% CI 75 to 100%) and ranged from 97%

(85.7 to 100%) for recreational athletes 40 years or older to 99% (89.3 to 100%) for non-throwing college athletes. Men were found to score higher on the active forward flexion subscale score (0.15 points higher out 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 has not been reported. However it is easy to understand and relatively short (5 items for the modified UCLA), and is likely completed within a few minutes.

Administrative burden. All versions other than the self-reported version of the UCLA, 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-Portugese 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-culture adaptation.

Psychometric Information

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

Reliability. The modified UCLA has single items measuring each domain so internal consistency cannot be measured. However one study found that the Italian UCLA scale version of the pain and function items had high internal consistency with a Cronbach's $\alpha > 0.89$, 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 has been reported in a mixed post shoulder surgery population (N= 31 (mostly RC tears (N=20) and instability (N=9)): ICCs for pain = 0.78 (95% CI 0.58 to 0.89), function = 0.89 (0.78 to 0.94) and satisfaction= 0.79 (0.59 to 0.89 (68). However the same study reported poor test-retest

reliability in a mixed group who had not undergone surgery: pain 0.59 (0.25–0.80), function 0.51 (0.14–0.75) and satisfaction unable to be estimated. The authors suggested that the modified UCLA may not be appropriate for use in a non-surgical population. Another study however found good test-retest reliability in 20 patients who had surgical or conservative management at 6, 12 and 13 weeks post proximal humerus fracture (ICC: 0.93 (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 2-item Italian version of the modified UCLA in a population of patients who had been treated with surgery for anterior shoulder instability (r= 0.93 for UCLA pain subscale, 0.95 for UCLA function subscale (170).

Validity. One study in patients undergoing rehabilitation post-proximal humerus fracture assessed content validity of the modified UCLA score linked to ICF codes (145). Apart from subjective shoulder value (‘‘not definable’’), items linked with two ICF components: “(b)Body Functions” (pain, joint mobility, muscle strength/endurance) and “(d)Activities and participation” (using transport/driving, self-care, dressing, shopping, housework). All items linked with body functions were separate items while 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 measured in, are as follows:
ASES: Arthroscopic rotator cuff population, pre-surgery, 6, 12 and 24 months post-surgery: very high correlation overall (r = 0.91), moderate in the preoperative period (r = 0.67, high at 6 months after surgery (r = 0.87) and very high at 12 and 24 months (r = 0.90 and 0.92) (73); mixed population with rotator cuff disease, isolated superior labral anterior posterior (SLAP) lesions, and shoulder instability after surgery 0.38 (69).
WORC: Mixed population of Brazilian patients with rotator cuff disorders: r = 0.80)(240).
WOSI: Patients having treatment for instability: r = 0.649 at baseline (prior to intervention) and r= 0.694 for change score (baseline to 3 months)(147).

OSS: Proximal humerus fracture: r = 0.77 (at 6 weeks), 0.83 (at 12 weeks), 0.73 (for change scores from 12 to 13 weeks (145); impingement or calcific tendinitis in German patients: r = 0.66 (187); Italian patients after shoulder surgery for anterior instability: r = 0.67)(185)
DASH: Proximal humerus fracture: r = 0.65 (at 6 weeks), 0.84 (at 12 weeks), -0.65 (for change scores from 12 to 13 weeks (145).
SSV: Proximal humerus fracture: r = 0.45 (at 6 weeks), 0.81 (at 12 weeks), 0.46 (for change scores from 12 to 13 weeks (145).
CM: Proximal humerus fracture: r = 0.70 (at 6 weeks), 0.92 (at 12 weeks), 0.83 (for change scores from 12 to 13 weeks (145).

One study assessed correlation of the modified UCLA with other shoulder measures in people without shoulder complaints: ASES: r = 0.65; CM Shoulder Score: r = 0.42; OSS: r = 0.80; Shoulder Pain and Disability Index (SPADI) r = -0.68; and Stanmore Percentage of Normal Shoulder Assessment (SPONSA): r = -0.63 (9). Correlation of the 4-item UCLA has been reported in patients with previous rotator cuff repair: SST: r = 0.76; CM: r = 0.66 (146). The UCLA scores were higher than the CM scores in almost all participants, while 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: SPADI disability subscale: r = 0.64; SST: r = 0.60 (29). Correlations with the UCLA pain subscale score were: SPADI pain subscale: r = 0.63, while 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: SPADI: total SPADI r = -0.63, SPADI pain subscale r = -0.61, SPADI function subscale r = -0.64 subscales and SST: r = 0.61; while correlations with the function item were: SPADI: total SPADI r = -0.57, SPADI pain subscale r = -0.52, SPADI function subscale r = -0.60 subscales and SST: r = 0.51) (170).

Agreement between the modified UCLA and patient-reported UCLA. Acceptable agreement in a post-operative population who have undergone either arthroscopic subacromial

decompression or rotator cuff repair, N=100): ICC = 0.910 (95% confidence interval (CI), 0.87 to 0.94) for whole cohort; ICC = 0.951 (95% CI, 0.92 to 0.97) for cohort who had subacromial decompression (N=46); and ICC = 0.734 (95% CI 0.61 to 0.83) for cohort who had 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, mostly explained by differences in the strength of forward flexion item. This could have arisen due to patient apprehension in pushing against resistance when performing 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 effect sizes (ES) and standardized response means (SRMs) of the modified UCLA according to patient population are as follows: subacromial decompression: ES at 6 weeks, 6 and 12 months: 1.17, 2.0 and 2.73 points; SRM at 6 weeks, 6 and 12 months: 0.83, 1.41 and 1.69 points (241); surgery for instability: SRM 0.385 points (147); rotator cuff disorders three months after treatment (surgery or physiotherapy)(Brazilian Portugese version): ES 1.17; SRM 1.66 points (242);

Minimally important differences. In patients being treated for proximal humerus fractures 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 was found to be 0.90 (1.77) for pain and 0.15 (0.30) for function (170).

Generalizability. The UCLA and later versions has been predominantly used to assess outcome 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 dissection of head and neck cancers.

Use in clinical trials. Up to December 2015, the modified UCLA or other versions has been used in 35 randomised controlled trials (2). Examples from the last five years include therapeutic ultrasound (243), mirror therapy (244), platelet-rich plasma injection (245) for adhesive capsulitis; deep heat versus ultrasound for shoulder pain (46); immediate versus delayed passive mobilization (246, 247) doxycycline (248) and platelet rich plasma (249) following rotator cuff repair; double versus single row rotator cuff repair (250, 251), suture-spanning augmentation of single-row rotator cuff repair (117); rotator cuff repair with or without acromioplasty (87); acupuncture (252), kinesiotaping and subacromial corticosteroid injection (253), ultrasound-guided subacromial NSAID or glucocorticoid injection (254) for impingement syndrome and open versus arthroscopic repair of traumatic anterior shoulder instability (255). 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 (256).

In many trials conducted in countries where English is not the first language it is unclear whether or not the UCLA was cross culturally translated into the local language e.g., trials conducted in Korea (46), Turkey (244, 253), China (245, 247, 254), Belgium (246), Mexico (248), Italy (250). Taiwan (117), Brazil (249), Spain (252) and Japan (251, 255). One ongoing trial comparing figure-of-eight bandage versus arm sling for treating middle-third clavicle fractures in adults is using the Brazilian Portugese version of the modified UCLA scale (257).

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 non-inflammatory shoulder disorders in surgical and non-surgical populations. Normative data are available and the UCLA scale score does not appear to be influenced by gender or handedness but scores do decline slightly with age. A self-administered version, which performs acceptably compared to when an observer assesses forward flexion and strength of forward flexion, is available and may be useful in situations where 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 to similar instruments. It is responsive to change across a range of conditions

but the MID has only been reported for people following proximal humeral fractures.

Caveats and cautions. The modified UCLA 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 are that they combine frequency and severity of pain and 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 which 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 this item 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 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. Combination of self and clinical assessment and solely self-assessment versions are available. Widely used to assess outcomes from shoulder surgery and non-operative 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 Portugese 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 might be preferred when face-to-face assessment is not possible or desirable.

Western Ontario Rotator Cuff (WORC) Index

Description

Purpose. The original version of the WORC Index was published in 2003 and developed to specifically measure disorders of the rotator cuff (258). It has also been used and evaluated for patients with rotator cuff repair (259), shoulder instability (260) and winged scapula (261).

Content or domains. There are five domains: physical symptoms, sports/recreation, work, lifestyles and emotions (258).

Number of items. There are 21 items in total: 6 in Physical symptoms, 4 in Sports/recreation, 4 in Work, 4 in Lifestyles, and 3 in Emotions (258).

Response options/scale. All items are rated using a visual analogue scale (VAS) (258). On the paper version the lines are 100 millimeters long.

Recall period for items. Two weeks.

Cost to use. Free of charge.

How to obtain. Available in various references, e.g., (258).

Practical application

Method of administration. Self-assessment using pen and paper (258) but it has also been delivered electronically in many studies.

Scoring. Each item is scored from 0 (best) to 100 (worst). The total score ranges from 0 to 2100 (258) 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 pretreatment, 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 (260), and it increase to 78.8, 737.7, and 89.8 six months after treatment among patients in the three groups, respectively. 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 six months after treatment (149).

Respondent time to complete. Mean time to complete the WORC Index is 3 minutes.

Administrative burden. Scoring is straightforward and no special software or equipment is required (260). 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 (262), Swedish (263), Canadian French (264), Polish (265), Turkish (266), Brazilian Portuguese (267), Persian (268) (269) and Norwegian (15, 270).

Psychometric information

Floor and ceiling effects. The floor and ceiling effects of the WORC Index overall score was 1.1% and 5.6% respectively in a study of patients with rotator cuff disease and shoulder instability (260). They were low across all domains except for floor effects for the Work domain (12.2%) and ceiling effects for the Emotions domain (14.4%). A substantial ceiling effect (39.1%) of the WORC Index overall score was observed six months after treatment when patients had improved and this was evident across all domains (ranging from 34.5 to 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 domain) to 0.91 (Physical Symptoms), and 0.96 for the total score (258).

Validity. The convergent validity of the WORC is relatively good. The Pearson correlation coefficient between the WORC and the absolute CM, relative CM (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 and the Physical Summary Score of the SF-36, SST and DASH was r= 0.58, 0.91, and 0.88, respectively in patients following rotator cuff repair (271). However, lower correlations were reported between the WORC and the ASES, UCLA and DASH in another study of patients with rotator cuff disease receiving various treatments (r = 0.68, 0.48, and 0.63, respectively (258).

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 six months after treatment, the correlation between the change in the WORC and the change in the absolute and relative (age-adjusted) CM and ASES was r = 0.77, 0.70, and 0.85, respectively (149). Another study in patients with rotator cuff disease receiving various treatments the reported correlations of the change in scores between the WORC and the ASES, UCLA and DASH of r = 0.76, 0.72, and 0.66 respectively (258)..

The discriminant validity of the WORC is satisfactory, as shown by correlations with objective measurements. For example, the correlation between the WORC and external and internal rotation strength and external and internal rotation range of motion was r = 0.38, 0.45, 0.27, and 0.31, respectively in patients following rotator cuff repair (271). The WORC also has been reported to have low correlations with subscales of the SF-36 ($r < 0.5$) except for Physical function (r = 0.56) in patients with rotator cuff disease (258). 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 criteria for surgery (149).

The longitudinal discriminant validity of the WORC is also satisfactory, as shown by the low correlations between the change of the WORC 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 (258). The correlations between the change in WORC and change of strength of the affected and unaffected arms was $r = 0.37$ and 0.20, respectively, when evaluated 6 months after various treatment in patients with rotator cuff disease in another study (149).

The WORC has good known group validity. For example, the WORC can differentiate between patients whose occupation is affected by shoulder symptoms and the ones who are not (149). The WORC can also detect the difference between patients whose range of motion in external rotation is $>45^{o}$ and the ones whose range of motion is $<45^{o}$, and between patients receiving or not receiving workers' compensation (271). The longitudinal known group validity of the WORC was questioned in one study on the basis that it failed to detect change in patients with less compared with more pathology as determined by the requirement for more extensive surgery at three and six months after treatment (149).

Responsiveness. As a specialized instrument, the WORC 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 (260). In another study, an ANOVA showed the WORC score can differentiate between change from baseline, 3 and 6 months after rotator cuff surgery treatment with an SRM = 1.33 (149).

Minimally important differences. The MDC of the WORC 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 (260). In the same study the MIC of the WORC estimated by ROC cut-off was 34%, 22.9%, and 31.8%, respectively, while the MIC estimated by 95% limit cut-off was 35.3%, 41.9%, and 46%, respectively.

Generalizability. The WORC 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 (261).

Critical appraisal of overall value to the rheumatology community

Strengths. The WORC is relatively short, easy to administer and score, 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 (272) and reliability in patients with rotator cuff disorders. The responsiveness of the WORC has been questioned for some other languages (27, 273).

Clinical usability. Useful in a clinical context for assessing rotator cuff disorders. Relatively brief, easy to administer and score, and responsive to change.

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

Patient-Reported Outcomes Measurement Information System (PROMIS) Physical Function Upper Extremity Computerized Adaptive Test (CAT)

Description

Purpose. PROMIS Physical Function Upper Extremity CAT measures upper extremity function (274). It was originally developed in 2014 as part of the PROMIS initiative which is an NIH Common Fund project involving the dynamic assessment of patient-reported outcomes. 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 (275-279). 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 computerized adaptive test (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 (i.e., short form) in maximizing assessment sensitivity while simultaneously minimizing the number of items needed.

Content or domains. Upper extremity physical function.

Number of items. There is a maximum of 12 of the 16 items when using the CAT. 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. Self-assessment, either using pen and paper for the static version, or 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 standard deviation of 10; higher scores indicate more of the construct (e.g., higher scores indicate greater function).

Score interpretation. This approach allows for 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 (e.g., 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. Takes approximately 70 seconds to complete (range 25 to 307 seconds) according to one study (280), 61.3 seconds on average (sd=28.8) in another study (281), and 96.9 seconds (sd=25.1) in a third study (282). As 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 (e.g., RedCap) that have access to scoring algorithms for arriving at t-scores. The Assessment Center (AC) enables the investigator to create a data collection website including any PROMIS item banks and short forms (283). It supports data collection designs that include multiple time points and multiple treatment arms and enables investigators to monitor enrollment of participants and completeness of data collection during the course of their research.

Translations/adaptations. It has been translated into Dutch-Flemish (284).

Psychometric information

Floor and ceiling effects. No floor or ceiling effects have been found in studies including patients with hand or upper extremity conditions (280), shoulder arthritis (282), shoulder pain (281), patients undergoing primary total shoulder arthroplasty (285), patients with

varying types of rotator cuff disease(286), and patients with upper extremity fractures (287).

A study including patients who underwent operative treatments for shoulder instability reported ceiling effects at 6 months (68.1%) and 2 years (67.0%) (288). In a subgroup aged ≤ 21 years ceiling effects were 71.1% at 6 months and 81% at 2 years. Ceiling effects were found for 3% patients (t-score of 56 or higher) undergoing arthroscopic rotator cuff repair (289), and 7.2% (t-score of 56 or higher) of patients with hand or shoulder conditions (290). 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 (291). A further study showed ceiling effects when inspecting item loadings on lower functional ability levels (292). A study in upper extremity trauma patients found no floor effects but 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)(293).

In a systematic review of floor and ceiling effects across 12 studies including 18,113 patients, Gulledge et al reported floor effects ranging from 0-1.6% and ceiling effects ranging from 0-28% (294).

Reliability. There is some evidence of person and item reliability in patients with shoulder instability: item reliability r = 0.82 to 0.96, person reliability r = 0.84 to 0.85 (291, 292). In patients with upper extremity trauma the average marginal reliability was r = 0.90 (293).

Validity. A Rasch analysis of 1,197 adults patients with varying hand and upper extremity complaints (non-shoulder) showed the item bank fit a unidimensional model (Eigen value of 1.96) with 4.2% unexplained variance (292). 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 the PROMIS physical function upper extremity to load onto one factor only reflecting capability which was separate from another factor reflecting quality of life, although it did not explore differential item functioning (287).

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

Short Musculoskeletal functional assessment (SMFA): Upper extremity trauma: r= -0.76 to -0.69 (293)

Quick-DASH: Upper extremity trauma: r= -0.82 to -0.75 (293); hand and upper extremity conditions: r=-0.81 (280); upper extremity fractures: r=-0.47 to -0.83 (1 week, 9 months)(287)

PROMIS Physical Function- Short Form 8a (PROMIS-PF-SF8a): Upper extremity trauma: r= 0.73 to 0.77 (293)

PROMIS Physical Function CAT (PROMIS PF CAT): Shoulder instability: r=0.63 (291); Rotator cuff pathology: r=0.70 (286); Upper extremity conditions: r=0.69 (baseline), r=0.53 (change)(290); hand and upper extremity conditions: r=0.48 (280)

PROMIS Pain Interference CAT (PROMIS PI): hand and upper extremity conditions: r=-0.60 (280)

PROMIS Mobility CAT (PROMIS M): hand and upper extremity conditions: r=0.41 (280); pediatric and adolescent patients in a ambulatory sports medicine clinic: r=-0.75 (295)

PROMIS Depression CAT (PROMIS D): arthroscopic rotator cuff repair patients: r^2=-0.196 (preoperative), r^2=-0.431 (postoperative)(296); pediatric and adolescent patients in a ambulatory sports medicine clinic: r=-0.21 (295)

ASES: Shoulder instability: r=0.71 (291); rotator cuff pathology: r=0.77 (286); total shoulder arthroplasty: r=0.55 (285); shoulder pain (across conditions): r=0.72 (281); adhesive capsulitis: r=0.86 (281); failed arthroplasty: r=0.49 (281)**;** fracture: 0.87 (281); instability: 0.55 to 0.73 (281, 288); impingement syndrome: 0.88 (281); rotator cuff disease: 0.65 (281);

other shoulder pain: r=0.74 (281); shoulder arthritis: r=0.57 (282); rotator cuff repair: r=0.59 (289).

WOSI: Shoulder instability: r=0.63 (291);

Marx Shoulder Activity Scale (Marx): Shoulder instability: r=0.06 (291); rotator cuff pathology: r=0.23 (286); total shoulder arthroplasty: r=0.06 (285).

Short Form 36-Health Survey Physical Function Subscale: Shoulder instability: r=0.78 (291).

SF36: Rotator cuff pathology: r=0.66 (286); total shoulder arthroplasty: r=0.53 (285).

Short Form 36-Health Survey General Health (SF36 GH): Rotator cuff pathology: r=0.30 (286).

EuroQual 5 dimensions questionnaire (EQ-5D): Shoulder instability: r=0.66 (291); Rotator cuff pathology: r=0.73 (286); total shoulder arthroplasty: r=0.48 (285).

WORC: Rotator cuff pathology: r=0.73 (286).

Western Ontario Osteoarthritis Shoulder index (WOOS): total shoulder arthroplasty: r=0.34 (285).

SST: shoulder pain (across conditions): r=0.82 (281); adhesive capsulitis: r=0.91 (281); failed arthroplasty: r=0.19 (281); fracture: r=0.93 (281); instability: r=0.81(281); impingement syndrome: r=0.93 (281); rotator cuff disease: r=0.78 (281); other shoulder pain: r=0.75 (281); shoulder arthritis: r=0.64 (282); rotator cuff repair: r=0.62 (289).

The 2-item Patient Health Questionnaire (PHQ-2): hand and upper extremity conditions: r=-0.30 (280).

The 2-item Pain Self-Efficacy Questionnaire (PSEQ-2): hand and upper extremity conditions: r=0.47 (280)

Pain (Numerical rating scale; NRS): hand and upper extremity conditions: r=-0.59 (280)

Responsiveness. Several studies report good responsiveness with the SRM, responsiveness to change, effect sizes in a variety of shoulder conditions. In patients with upper extremity trauma one study reported good responsiveness to treatment of the injury (293). This study also reports that effect sizes and SRM were large for the measure for the full sample, a subsample split by occurrence/nonoccurrence of a secondary treatment and for subsample split by fracture severity, although actual values are not reported in the published paper or in supplemental material. In 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 to an absolute mean difference of 0.80 (290). It also reported no differences in magnitude of change in those with hand versus 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 (296). 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 =0.92 (288).

Minimally important differences. The minimally important difference was reported as 2.1 in one study of patients with non-shoulder upper extremity conditions (297).

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 physical function upper extremity CAT.

Critical appraisal of overall value to the rheumatology community

Strengths. The PROMIS physical function upper extremity CAT is short, uses item response theory and has very good evidence of validity, 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 sample of patients with varying "upper extremity" disorders and almost all studies do not delineate shoulder from elbow or hand disorders in any subgroup analyses. In addition, more work should be done to establish the SDC and the MCID/MID in known groups.

Clinical usability. May be useful in a selection of shoulder related disorders. Brief, easy to use and score, and appears to be valid and responsive.

Research usability. Brief, easy to administer and responsive. Valid for intervention and population-level studies.

Summary / recommendations

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 most commonly used instruments in trials in the last five years as well as information for the PROMIS Physical Function Upper Extremity CAT.

Acknowledgement

RB is suppored by an Australian National Health and Medical Research Council (NHMRC) Senior Principal Research Fellowship.

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 M, Huanh H, Verhagen A, Gagnier J, Buchbinder R. Outcome reporting in randomized trials for shoulder disorders: Literature review to inform the development of a Core Outcome Set. Arthritis Care Res. 2018;70(2):252–9.

3. Angst F, Schwyzer H, Aeschlimann A, Simmen B, Goldhahn J. Measures of Adult Shoulder Function. Arthritis Care Res. 2011;63:S174–S88.

4. Terwee C, Jansma E, Riphagen I, de Vet H. 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(4):143-9.

6. Williams JW, Jr., Holleman DR, Jr., Simel DL. Measuring shoulder function with the Shoulder Pain and Disability Index. J Rheumatol. 1995;22(4):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. Journal of Shoulder & Elbow Surgery. 2018;27(2):306-14.

10. Paul A, Lewis M, Shadforth MF, Croft PR, Van Der Windt DAWM, Hay EM. A comparison of four shoulder-specific questionnaires in primary care. Ann Rheum Dis. 2004;63(10):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. 2018;97(26):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(3):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(4):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 Musculoskeletal Disorders. 2008;9:68.

16. Thoomes-de Graaf M, Scholten-Peeters GGM, 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(6):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(2):99-104.

18. Sharma N, Sharma S, Kataria C. Translation and Adaptation of Shoulder Pain and Disability Index (SPADI) into Hindi-Part 1. Ind J Physiother Occup Ther. 2011;5(4):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(4):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(1):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(6):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(1):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(1):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 Otorhinolaryngologica Italica. 2012;32(1):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. 2007;46(1):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. Journal of Clinical Epidemiology. 2010;63(5):575-84.
28. Angst F, Pap G, Mannion A, Herren D, Aeschlimann A, Schwyzer H, 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. Physical Therapy. 2000;80(8):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 & Quality of Life Outcomes. 2015;13:63.
31. Beaton D, Richards RR. Assessing the reliability and responsiveness of 5 shoulder questionnaires. Journal of Shoulder & Elbow Surgery. 1998;7(6):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. 2017.
35. Staples MP, Forbes A, Green S, Buchbinder R. Shoulder-specific disability measures showed acceptable construct validity and responsiveness. J Clin Epidemiol. 2010;63(2):163-70.
36. Beaton DE, Richards RR. Measuring function of the shoulder. A cross-sectional comparison of five questionnaires. Journal of Bone & Joint Surgery - American Volume. 1996;78(6):882-90.
37. Christie A, Hagen K, 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(5):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. Journal of Shoulder & Elbow Surgery. 2005;14(4):380-4.
40. Tuğay U, Tuğay N, Gelecek N, Özkan M. Oxford Shoulder Score: cross-cultural adaptation and validation of the Turkish version. Archives of Orthopaedic & Trauma Surgery. 2011;131(5):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(10):1079-89.
42. MacDermid J, 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(8):538-47.
44. Angst F, Goldhahn J, Drerup S, Aeschlimann A, Schwyzer HK, Simmen BR. Responsiveness of six outcome assessment instruments in total shoulder arthroplasty. Arthritis & Rheumatism. 2008;59(3):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(10):1008-18.
46. Kim G, Won Y, Park S, Seo J, Kim D, Lee Y, 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 research & management. 2019;2019.
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(1):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(8):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(9):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(4):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(3):341-50.
52. Richards R, An K, Bigliani L, Friedman R, Gartsman G, Gristina A, et al. A standardized method for the assessment of shoulder function. J Shoulder Elbow Surg. 1994;3:347–52.

53. Sallay P, Reed L. The measurement of normative American Shoulder and Elbow Surgeons scores. J Shoulder Elbow Surg. 2003;12:622-7.

54. Leggin B, Michener L, Shaffer M, Brenneman S, Iannotti J, Williams Jr G. The Penn Shoulder Score: Reliability and validity. Journal Orthopaedic & Sports Physical Therapy. 2006;36:138-51.

55. Goldhahn J, Angst F, Drerup S, Pap G, Simmen B, Mannion A. 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:S85–90.

57. Knaut L, Moser A, Melo S, Richards R. 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(2):176–89.

58. Moser A, Knaut L, Zotz T, Scharan K. Validity and reliability of the Portuguese version of the American Shoulder and Elbow Surgeons Standardized Shoulder Assessment Form. Revista Brasileira de Reumatologia. 2012;52(3):348-56.

59. Vrotsou K, Cuéllar R, Silió F, Ángel Rodriguez M, 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 and Quality of Life Outcomes. 2016;14.

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 Musculoskel Dis. 2014;15.

61. Çelik D, Atalar A, Demirhan M, Dirican A. Translation, cultural adaptation, validity and reliability of the Turkish ASES questionnaire. Knee Surg Sports Traumatol Arthrosc. 2013;21(9):2184-9.

62. Yahia A, Guermazi M, Khmekhem M, Ghroubi S, Ayedi K, Elleuch M. Translation into Arabic and validation of the ASES index in assessment of shoulder disabilities. Ann Phys Rehab Med. 2011;54:59-72.

63. Kocher M, Horan M, Briggs K, Richardson T, O'Holleran J, Hawkins R. 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. Journal of Bone & Joint Surgery 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. Journal of Shoulder & Elbow Surgery. 2015;24(12):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. American Journal of Sports Medicine. 2018;46(14):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. Journal of Shoulder & Elbow Surgery. 2019;28(2):324-9.
67. Michener L, McClure P, Sennett B. 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. Journal of Orthopaedic & Sports Physical Therapy. 2002;32(7):336-46.
69. Oh J, Jo K, Kim W, Gong Y, Han S, Kim Y. 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. Journal of Shoulder & Elbow Surgery. 2019;28(10):1869-76.
71. Razmjou H, Bean A, van Osnabrugge V, MacDermid J, Holtby R. Cross-sectional and longitudinal construct validity of two rotator cuff disease-specific outcome measures. BMC Musculoskeletal Disorders. 2006;7.

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. Archives of Orthopaedic & Trauma Surgery. 2000;120(7-8):432-6.
73. Assuncao JH, Malavolta EA, Gracitelli MEC, 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. Journal of Shoulder & Elbow Surgery. 2017;26(7):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(9):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. Journal of Shoulder & Elbow Surgery. 2018;27(6S):S17-S23.
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(5):1430-6.
77. Yoo J, Koh K, Park W, Park J, Kim S, Yoon Y. The outcome of ultrasound-guided needle decompression and steroid injection in calcific tendinitis. J Shoulder Elbow Surg. 2010;19:596–600.
78. Tashjian R, Deloach J, Green A, Porucznik C, Powell A. Minimal clinically important differences in ASES and simple shoulder test scores after nonoperative treatment of rotator cuff disease. J Bone Joint
Surg Am. 2010;92:96–303.
79. Tashjian R, Hung M, Keener J, Bowen R, 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(1):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? Clinical Orthopaedics & Related Research. 2016;474(12):2672-81.
81. Wong SE, Zhang AL, Berliner JL, Ma CB, Feeley BT. Preoperative patient-reported scores can predict postoperative outcomes after shoulder arthroplasty. Journal of Shoulder & Elbow Surgery. 2016;25(6):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. Journal of Shoulder & Elbow Surgery. 2019;28(5):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. Journal of Shoulder & Elbow Surgery. 2018;27(5):e160-e6.
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. Orthopaedic journal of sports medicine. 2016;4(10).
85. Zhu Q, 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(1):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. Prospective randomized trial. Arthroscopy - journal of arthroscopic and related surgery. 2015;31(6):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 - journal of arthroscopic and related surgery. 2018;34(12):e3-e4.
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. American journal of sports medicine. 2015;43(4):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. Orthopaedic journal of sports medicine. 2019;7(7).

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 Surgery, Sports Traumatology, Arthroscopy. 2016;24(12):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. European journal of orthopaedic surgery & traumatologie. 2016;26(8):837-42.
92. Shishani Y, Lederman ES, Denard PJ, Gobezie R. Can a functional difference be detected in reverse arthroplasty with 135[degrees] vs. 155[degrees] prosthesis for the treatment of rotator cuff arthropathy: a prospective randomized study. Journal of shoulder and elbow surgery. 2018;27(4):e121-.
93. Sabesan VJ, Shahriar R, Petersen-Fitts G, Bou-Akl T. 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(5):e150-.
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. American journal of sports medicine. 2016;44(2):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. Clinical rehabilitation. 2015;29(5):447-56.
96. Perdreau A, Joudet T. Efficacy of multimodal analgesia injection combined with corticosteroids after arthroscopic rotator cuff repair. Orthopaedics & traumatology, surgery & research : OTSR. 2015;101(8 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. American journal of sports medicine. 2017;45(2):440-8.
98. Park JH, Park J, Rhee SM, Suk H, Tae-Yon K, Joo R, et al. Optimum tension of bridging suture in transosseous equivalent rotator cuff repair: a clinical prospective randomized comparative study. American academy of orthopaedic surgeons. 2019.

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. Journal of shoulder and elbow surgery. 2016;25(8):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. Orthopaedic journal of sports medicine. 2015;3(1):1-4.
101. Bahk Ji H, Kim Y-S, Lee H-J, Kim Jong H, Cho H-L, Yoon S-J, et al. Comparison of high- and low-dose intra-articular steroid injection for treatment of primary shoulder stiffness: a prospective randomized controlled trial. American academy of orthopaedic surgeons. 2019.
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(5):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. Journal of Shoulder & Elbow Surgery. 2019;28(6):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. American journal of sports medicine. 2016;44(7):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. Archives of Orthopaedic & Trauma Surgery. 2017;137(12):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 surgery, sports traumatology, arthroscopy. 2019.
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.
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. Journal of orthopaedic surgery. 2017;25(3).

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. Journal of Shoulder & Elbow Surgery. 2016;25(7):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. American journal of physical medicine & rehabilitation. 2016;95(3):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. American journal of sports medicine. 2016;44(8):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. Arthroscopy - journal of arthroscopic and related surgery. 2017;33(6):e12-.

113. Hart R, Smid P. Tears of the supraspinatus muscle sutured with added mesenchymal stem cells. Osteoporosis international. 2018;29(1):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[TM] vs. proximal humerus nail MultiLoc(R). BMC musculoskeletal disorders. 2016;17(1):448.

115. Hwang JT, Yang CJ, Noh KC, Yoo YS, 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? Arthroscopy. 2016;32(4):560-7.

116. Lee HJ, Kim YS, Cho HL, Yoon SJ, Han SY, Kim JH, et al. Is atelocollagen injection effective in partial thickness rotator cuff tears at clinic? A prospective randomized clinical evaluation. American academy of orthopaedic surgeons. 2019.

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? American journal of sports medicine. 2019;47(6):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 - journal of arthroscopic and related surgery. 2017;33(6):e12-e3.

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. Orthopaedic journal of sports medicine Conference. 2018;6(7 Supplement 4).

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 surgery, sports traumatology, arthroscopy. 2016;24(2):414-21.

121. Constant CR, Murley AH. A clinical method of functional assessment of the shoulder. Clin Orthop Relat Res. 1987(214):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(2):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(2):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(1):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(4):A4608.

126. Yian EH, Ramappa AJ, Arneberg O, Gerber C. The Constant score in normal shoulders. J Shoulder Elbow Surg. 2005;14(2):128-33.

127. Constant CR. Schulterfunktionsbeurteilung. Der Orthopäde. 1991;20(5):289-94.

128. Moeller AD, Thorsen RR, Torabi TP, Bjoerkman A-SD, 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(5):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(1):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(5):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(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(1):178.
133. Henseler JF, Kolk A, van der Zwaal P, Nagels J, Vliet Vlieland TPM, Nelissen RGHH. The minimal detectable change of the Constant score in impingement, full-thickness tears, and massive rotator cuff tears. J Shoulder Elbow Surg. 2015;24(3):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(3):e513-e9.
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(1):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(2):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(12):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(3):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. Journal of Orthopaedic & Sports Physical Therapy. 2015;45(8):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 Musculoskeletal Disorders. 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? BMC Musculoskelet Disord. 2019;20(1):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(10):967-72.
143. van de Water ATM, 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. Journal of Bone & Joint Surgery - British Volume. 2001;83(6):877-82.
145. van de Water A, Shields N, Davidson M, Evans M, Taylor N. Reliability and validity of shoulder function outcome measures in people with a proximal humeral fracture. Disability and Rehabilitation. 2014;36(13):1072-9.
146. Romeo AA, Mazzocca A, Hang DW, Shott S, Bach BR, Jr. Shoulder scoring scales for the evaluation of rotator cuff repair. Clinical Orthopaedics & Related Research. 2004(427):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(6):764-72.
148. de Witte PB, Henseler JF, Nagels J, Vliet Vlieland TPM, Nelissen RGHH. The Western Ontario rotator cuff index in rotator cuff disease patients: a comprehensive reliability and responsiveness validation study. Am J Sports Med. 2012;40(7):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(5):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. 2008;47(6):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(8):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(2):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(2):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(12):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(8):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(10):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(7):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(2):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(3):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(2):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(3):663-70.
163. Lippitt S, Harryman D, Matsen Fr. 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(4):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. Archives of Physical Medicine & Rehabilitation. 2001;82(11):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(6):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? PLoS One. 2013;8(5).
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(2):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. Archives of Orthopaedic & Trauma Surgery. 2015;135(12):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. Journal of Shoulder & Elbow Surgery. 2018;27(8):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(3):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. Journal of Shoulder & Elbow Surgery. 2007;16(3):260-7.
173. Hsu J, Russ S, Somerson J, Tang A, Warme W, Matsen Fr. Is the Simple Shoulder Test a valid outcome instrument for shoulder arthroplasty? J Shoulder Elbow Surg. 2017:1693-700.
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(3):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.
176. Berendes T, Pilot P, Willems J, Verburg H, te Slaa R. Validation of the Dutch version of the Oxford Shoulder Score. Journal of Shoulder & Elbow Surgery. 2010;19(6):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. Journal of Orthopaedic & Sports Physical Therapy. 2010;40(7):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. Clinical journal of sport medicine. 2016;26(1):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(376):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 and elbow surgery. 2018;19(3):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. Archives of Orthopaedic & Trauma Surgery. 2012;132(1):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. 2018;97(23):e10926.

185. Murena L, Vulcano E, D'Angelo F, Monti M, Cherubino P. Italian cross-cultural adaptation and validation of the Oxford Shoulder Score. Journal of Shoulder & Elbow Surgery. 2010;19(3):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 Reumatologica Portuguesa. 2018;43(2):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. Archives of Orthopaedic & Trauma Surgery. 2004;124(8):531-6.

188. Lima Eda S, 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 Medical Journal. 2016;134(1):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. Orthopaedics & traumatology, surgery & research. 2016;102(5):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 & Quality of Life Outcomes. 2015;13:193.

191. Frich LH, Noergaard PM, Brorson S. Validation of the Danish version of Oxford Shoulder Score. Danish Medical Bulletin. 2011;58(11):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 & Quality of Life Outcomes. 2019;17(1).

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 & Rheumatism. 2002;47(5):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. Journal of Hand Therapy. 2011;24(1):53-60; quiz 1.

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 technology assessment (winchester, england). 2015;19(24):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. Journal of bodywork and movement therapies. 2016;16.

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. Journal of shoulder and elbow surgery. 2017;26(5):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. 2017(pagination).

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 musculoskeletal disorders. 2017;18(1).
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. American journal of sports medicine. 2015;43(6):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. European journal of physical and rehabilitation medicine. 2018;54(3):333-40.
204. Piekaar RSM, Bouman ICE, 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. Musculoskeletal Surgery. 2018;102(3):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(1).
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(1).
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(6):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(2):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(9):1370-7.

211. Beaton DE, Wright JG, Katz JN, Upper Extremity Collaborative G. Development of the QuickDASH: comparison of three item-reduction approaches. J Bone Joint Surg Am. 2005;87(5):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(1):13-9.

214. Roy JS, MacDermid JC, Woodhouse LJ. Measuring shoulder function: a systematic review of four questionnaires. Arthritis & Rheumatism. 2009;61(5):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(2):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? 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(2):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(12):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. Quality of Life Research. 2016;25(9):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(2):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(3):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(6):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(1):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(4):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(8):1043-51.
229. Mahabier KC, Den Hartog D, Theyskens N, Verhofstad MHJ, Van Lieshout EMM, Investigators HT. 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(1):e1-e12.

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(3):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(2):e0147842.
233. Amstutz H, Sew Hoy A, Clarke I. UCLA anatomic total shoulder arthroplasty. Clinical Orthopaedics & Related Research. 1981;155:7-20.
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. Journal of Shoulder & Elbow Surgery. 2008;17(4):564-9.
236. Tauro J. Arthroscopic rotator cuff repair: Analysis of technique and results at 2- and 3-year follow-up. The Journal of Arthroscopic and Related Surgery. 1998;14(1):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. Journal of Shoulder & Elbow Surgery. 2002;11(5):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(4):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(9):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. Clinical Journal of Sport Medicine. 2008;18(3):266-72.

241. O'Connor D, Chipchase L, Tomlinson J, Krishnan J. Arthroscopic subacromial decompression: Responsiveness of disease-specific and health-related quality of life outcome measures. Arthroscopy. 1999;15(8):836-40.

242. Diniz Lopes A, Ciconelli RM, Carrera EF, Griffin S, Faloppa F, Baldy dos Reis F. 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. Clinical & Experimental Rheumatology. 2009;27(5):758-64.

243. 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. Journal of back and musculoskeletal rehabilitation. 2018;31(5):955-61.

244. 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. Journal of Back & Musculoskeletal Rehabilitation. 2018;31(6):1177-82.

245. Lin J. Platelet-rich plasma injection in the treatment of frozen shoulder: a randomized controlled trial with 6-month follow-up . International journal of clinical pharmacology and therapeutics. 2018;56(8):366-71.

246. 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 orthopaedica belgica. 2015;81(3):485-92.

247. 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 (india). 2017;28(15):6818-22.

248. Diaz-Sanabria JI, Trueba-Vasavilbaso C, Gil-Orbezo F, Trueba-Davalillo C, Acuna-Tovar M. [Doxycycline as an adjuvant in postsurgical recovery of rotator cuff injuries. One-year follow-up]. Acta Ortopedica Mexicana. 2017;31(4):157-61.

249. 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. American journal of sports medicine. 2018;46(13):3134-41.

250. 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. American journal of sports medicine. 2016;44(4):948-56.
251. Yamakado K. Healing Rate Comparison Between Suture-Bridging and Medially Based Single Row Repair in Medium-Sized Supraspinatus Tear: a Randomized Prospective Trial. Arthroscopy - journal of arthroscopic and related surgery. 2018;34(12):e4-.
252. Rueda Garrido JC, Vas J, Lopez DR. Acupuncture treatment of shoulder impingement syndrome: a randomized controlled trial. Complementary therapies in medicine. 2016;25:92-7.
253. 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. American Journal of Physical Medicine & Rehabilitation. 2016;95(8):553-60.
254. 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 [chinese journal of surgery]. 2018;56(10):781-5.
255. 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. Journal of Orthopaedic Surgery. 2017;25(3):2309499017727947.
256. 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(7):1558-67.
257. 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.
258. 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(2):84-92.
259. 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.

260. Wessel RN, Wolterbeek N, Fermont AJM, Lavrijsen L, van Mameren H, de Bie RA. Responsiveness and disease specificity of the Western Ontario Rotator Cuff index. J Orthop. 2018;15(2):337-42.
261. 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(1):29-34.
262. 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(1):178.
263. 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 musculoskeletal disorders. 2016;17(1):165.
264. 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(3):292-8; quiz 9.
265. 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(1):333.
266. 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(12):1101-8.
267. Lopes AD, 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(3):266-72.
268. 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(3):293-9.
269. Lopes AD, 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. Clinical Journal of Sport Medicine. 2008;18(3):266-72.

270. 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(5):575-84.
271. 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(02):119-27.
272. 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 Surgery, Sports Traumatology, Arthroscopy. 2012;20(10):1961-70.
273. 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(2):121-5.
274. 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(11):2291-6.
275. DeWalt D, Rothrock N, Yount S, Stone A. Evaluation of item candidates: The PROMIS Qualitative Item Review. Medical Care. 2007;45(5):S12-S21.
276. 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(5):S22-S31.
277. Liu H, Cella D, Gershon G, Ahen J, Morales L, Riley W. Representativeness of the PROMIS internet panel. J Clin Epidemiol. 2010;63(3):1169-78.
278. 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(11):1195-204.
279. 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(11):1179-94.

280. Doring A, Nota A, Hageman M, Ring D. Measurement of upper extremity disability using the patient reported outcomes measurement information system. J Hand Surg Am. 2014;39:1160-5.

281. 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.

282. 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(5):1430-6.

283. 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 Applied Measurement. 2010;11(3):304-14.

284. 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;Epub ahead of print:1-9.

285. Dowdle S, Glass N, Anthony C, Hettrich C. Use PROMIS for patients undergoing primary total shoulder arthroplasty. Orthop J Sports Med 2017;5(9).

286. 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. **J Arthroscopic Rel Surg**. 2017;33(10):1770-4.

287. Jayakumar P, Teunis T, Vranceanu A, Lamb S, Willim 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.

288. 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(5).

289. 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-S23.

290. Beleckas C, Padovano A, Guattery J, Chamberlain A, Keener J, Calfee R. Performance of PROMIS upper extremity verses physical function CATs in upper extremity clinics. J Hand Surg Am. 2017;42(11):867-74.

291. 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(2):449-53.

292. Hung M, Voss M, Bounsanga J, Crum A, Tyser A. Examination of the PROMIS upper extremity item bank. J Hand Therapy. 2017;30(4):485-90.

293. 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(10):e321-6.

294. 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. 2019;Epub ahead of print:1-12.

295. 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. Orthopaedic Journal of Sports Medicine. 2019;7(6).

296. 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.

297. 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. 2019;1:e1-e15.

Table 1: Practical applications

Measure	Number of items	Content/ Domains (Number of items)	Method of administration	Recall period	Response format	Range of scores	Score interpretation	Availability of normative data	Cross-cultural validation
Oxford Shoulder Score (revised version)	12	Pain (4), daily functions (8)	Self	4 weeks	5 point Likert scale, 0 (worst) – 4 (best)	0-48	0–19 severe arthritis, 20–29 moderate to severe arthritis, 30–39 mild to moderate arthritis, 40–48 satisfactory joint function	No	Dutch, Italian, German, Portuguese, Korean, French, Turkish, Chinese, Persian, Spanish, Romanian
Simple Shoulder Test	12	Pain (2), function/strength (7), range of motion (3)	Self	Current	Yes/no	0-12	Percent 'yes' responses; 0 worst to 100 best	No	Dutch, Portuguese, Spanish, Persian, Italian
University of California–Los Angeles	5	Pain (1), function (1), patient satisfaction (1),	Self and observer; self version	4 weeks	Weighted categorical scales:	2-35	34-35 excellent, 28-33 good, both indicate	Yes	Brazilian Portugese, Italian

(UCLA) Shoulder Rating Scale (modified)		forward flexion (1), strength of forward flexion (1)			Pain and function 1-10; Other items 0-5		satisfactory result; 21-27 fair, 20 or less poor, unsatisfactory		
American Shoulder and Elbow Surgeons (ASES) Society Form	45 (12 items in score)	Score: pain (1), ADLs (11), [other domains not included in score: pain (5), stability (2); physician-assessed ROM (5), signs (11), strength (5), instability (8)	Only self (pain, ADLs) in score	Current	0-10 VAS (pain) and 4-point ordinal Likert scales: 0 - unable to do, to 3 - not difficult (ADL)	0-100	Pain and ADL subscales each transformed 0 to 50; 0 worst to 100 best	Yes	German, Italian, Brazilian Portuguese, Spanish, Finnish, Turkish, Tunisian Arabic
Shoulder Pain and Disability Index (SPADI)	13	Pain (5), disability (8)	Self	Past week	0 to 10 NRS	0-100	Each subscale 0-100; total score = mean of 2 subscales; 0 best to 100 worst	Yes	Spanish, Italian, Chinese, Nepali, Thai, Arabic, Persian, Brazilian Portuguese, Danish, German, Dutch, Turkish, Greek, Indian Tamil, Indian Hindi

DASH, QuickDASH	30; 11	Symptoms (6), function (24); Symptoms (3), function (8)	Self	Past week	5-point ordinal Likert scales: 1 – no symptoms/difficulty to 5 – extreme difficulty (unable to do)/symptoms	0-100	Scores transformed by (mean-1) x 25; 0 best to 100 worst	Yes	Available in over 50 languages
Constant-Murley Score	10	Pain (1), ADLs (4), ROM (4), strength (1)	Self (pain, ADLs); observer (ROM, strength)	Past week (ADLs), past 24 hours (pain)	Pain: 0-15 VAS; ADLs: sleep: 2-item ordinal scale, work and Recreation 0-15 VAS, functional Mobility 5-item ordinal scale	0-100	Total score = sum of subscales (pain 15, ADL 20, ROM 40, strength 25); 0 - worst to 100 – best	Yes	Danish, Turkish, Brazilian Portuguese, Greek, Chinese
WORC (western Ontario rotator cuff)	21	Physical Symptoms (6), Sports/Recreation (4), Work (4), in	Self	2 weeks	0-100 VAS	0-100 (normalized)	Higher score indicates greater function on the normalized score	No	Chinese, Swedish, French, Polish, Turkish, Portuguese,

index		Lifestyles (3), and Emotions (3)							Persian, and Norwegian
PROMIS	Up to 12	Upper extremity physical function	Self	Current	5 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)	Higher score indicates greater function: score of $\geq$ 60 and >70 indicates physical function exceeds 68.27% and 95.45% of general population respectively	Yes	Dutch-Flemish

Table 2: Psychometrics

Measure	Floor, ceiling effects	Reliability	Validity	Responsiveness	Minimally important differences	Generalizability	Used in RCTs
Oxford Shoulder Score	Both low	Internal consistency: Cronbach's 0.94; Test–retest:	Moderate to high correlation with other measures	Osteoarthritis and rheumatoid arthritis hemiarthroplasty	MCID 6 to 6.9 on 48-point scale	Widely translated and adapted into different languages and cultural	Yes

		Pearson's correlation 0.98	(range: 0.37 to 0.87)	ES 2.3; Impingement rotator cuff: surgery: ES 1.10–1.88, SRM 1.10–1.14 Rotator cuff: decompression (cuff repair): ES 0.97 (ref. 77) Impingement: no treatment described: ES 0.96 (ref. 43) Degenerative, inflammatory: surgery: ES 0.61 (ref. 64)		settings used for patients with various shoulder problems	
Simple Shoulder Test	Both below 21%	Unidimensional scale Good test-retest reliability (0.97-0.99)	High construct validity (>0.68) with shoulder specific measures such as SPADI, ASES, and DASH, CS, and WORC.	ES considered moderate (>0.63) to large (>0.87) in patients with osteoarthritis, shoulder instability, rotator cuff diseases, and total arthroplasty	MCID for the range 0–12: > 2.05 for rotator cuff disease; > 2.4 points for shoulder arthroplasty, 2.2 points for	Shoulder specific; also been used to assess shoulder function impacted by other diseases e.g., surgery for breast cancer. Translated and culturally adapted to	Yes

					various shoulder problems. Corresponds to MCID >17 for the range 0–100	various countries	
University of California–Los Angeles (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, 2.73 at 6 weeks, 6 and 12 months post subacromial decompression: SRM 0.83, 1.41 and 1.69 points respectively SRM post instability surgery 0.385 points ES 1.17; SRM 1.66 3 months post rotator cuff disorder treatment (Brazilian Portugese version)	MCID 2.0 to 2.4 (2-25 scale) (depending on method) for proximal humerus fracture	Translated and culturally adapted into 2 languages but used in many trials in other countries	Yes

American Shoulder and Elbow Surgeons (ASES) Society Standardized Shoulder Assessment Form	Low floor and ceiling effects	Internal consistency: Cronbach's α ranging from 0.61 to 0.96; fair to excellent test-retest reliability	Moderate to high correlation with other measures	ES 0.8 to 2.13 and SRM 0.75 to 1.81 across various shoulder conditions	MCID 6.3 to 26.9 (0-100 scale)	Translated and culturally adapted to various countries	Yes
Shoulder Pain and Disability Index (SPADI)	Low floor and ceiling effects	Internal consistency: Cronbach's alpha 0.92-0.96; Test-retest: ICC 0.84-0.95	Moderate to high correlation with other measures (range 0.45 – 0.93)	Total shoulder arthroplasty: ES 2.1; Adhesive capsulitis: ES 1.20-1.64; Shoulder pain (physiotherapy): ES 1.26; Rotator cuff surgery: SRM 1.23; Upper extremity disorders (occupational therapy or physiotherapy): ES 1.21	MCID 8-20 (0-100 scale)	Shoulder-specific and used in a wide range of clinical and research settings	Yes

DASH, QuickDASH	Very low or low floor and ceiling effects	Internal consistency: Cronbach's α = 0.92–0.98 DASH and 0.92–0.95 Quick DASH; high test–retest reliability: 0.89 to 0.98 DASH and 0.90 to 0.94 QuickDASH	Moderate to high correlation with shoulder-specific measures (range 0.50 – 0.93)	ES 0.55 to 1.26 and SRM 0. 5 to 1.45 across various shoulder conditions	MCID 4.4 to 12.4 DASH and 8 to 13.4 QuickDASH (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
Constant-Murley Score	Low floor and ceiling effects (possible floor effect for strength subscale)	Internal consistency: Cronbach's alpha 0.37 – 0.8; Test-retest: ICC 0.82-0.98	Moderate to high correlation with other shoulder measures (range 0.49 – 0.91)	Rotator cuff disease: SRM 1.38, ES 0.4; Subacromial decompression surgery SRM 1.12 – 2.09, ES 1.23 – 1.92; Shoulder arthroplasty SRM 1.99 – 2.4, ES 2.23 – 2.9	MCID 8-17 (0-100 scale)	Used in a wide range of clinical and research settings, for a broad range of shoulder conditions	Yes
WORC (western Ontario rotator cuff) index	Lower floor and ceiling effects. Possible ceiling effect after	Test-retest reliability: ICC 0.96 for the total score, 0.54-0.91 for subscales.	Moderate to high correlation with other shoulder measure, low to moderate correlation with	SEM is 1.33 among patients with rotator cuff pathology	MCID 22.9 to 34 (0-100 scale)	Widely translated and adapted in many languages but limited to diseases related to rotator cuff disorders	Yes

	treatment.		non-shoulder specific measures.				
PROMIS	0 to 1.6% floor and 0 to 28% ceiling effects	item reliability r = 0.82 to 0.96, person reliability r = 0.84 to 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 to 0.91)	Good responsiveness to change reported. SRM 0.92 after surgery for shoulder instability	MCID not reported for shoulder conditions; 2.1 in 1 study of non-shoulder upper extremity disorders	Has been used to assess various shoulder disorders, not yet widely translated	No