Shoulder Outcomes Measures

Abstract

General health as well as disease- or condition-specific outcome measures have long been used to assess patients with shoulder conditions. Currently, a variety of validated measures is available. Shoulder outcomes measures may be general (eg, American Shoulder and Elbow Surgeons; Constant; Disabilities of the Arm, Shoulder, and Hand), disease-specific (eg, Rotator Cuff Quality of Life, Western Ontario Rotator Cuff Index), or condition-specific (eg, Oxford Shoulder Instability Questionnaire). The results of shoulder arthroplasty and arthritis treatment can be assessed with the Hospital for Special Surgery score and the validated Western Ontario Osteoarthritis of the Shoulder Index. Combining a general health outcome measure, a general shoulder measure, a disease- or condition-specific shoulder measure, and an activity measure allows for broad patient assessment.

Outcomes of shoulder injuries and treatment have been evaluated with rating scales and scoring systems for years. Improvement in these measures and a better understanding of the methodology behind their creation and validation has resulted in an increased and improved number of tools. An understanding of the goals and use of these instruments allows health care personnel to better assess patient outcomes and critically review the medical literature.

Outcomes measures typically fall into two broad categories: general health and disease- or joint-specific. The Medical Outcomes Study 36-Item Short Form (SF-36) is the most popular general health measure in the field of orthopaedics. The SF-36 has been extensively reviewed previously in the *Journal of the American Academy of Orthopaedic Surgeons*.1,2 Disease- or joint-specific measures focus on the outcome of the management of a particular diagnosis or disease entity in a single joint. Most researchers believe that studies should include a general health measure and a disease-specific measure to allow broad comparison across patient populations.

More than 30 shoulder outcome measures have been described. We focus on the most appropriate shoulder outcomes measures for use in patient evaluation and research efforts (Table 1). These measures are the most popular and the most rigorously validated using current psychometric standards.

General and Arthroplasty Outcome Measures

American Shoulder and Elbow Surgeons Shoulder Outcome Score

The American Shoulder and Elbow Surgeons (ASES) shoulder outcome score was developed in 1994 in the United States by the research committee of the ASES.3 The purpose was to measure function so as to facilitate communication between investigators, stimulate the undertaking of multicenter studies, and allow for comparison of results from different surgeons and centers. The ASES shoulder outcome score is a disease-specific measure that consists of 10 questions assessing pain, function, and activities. It has been validated and is widely used in clinical practice and research.

---

From the Department of Orthopaedic Surgery, Washington University School of Medicine, St. Louis, MO (Dr. Wright), and Orthopedic Institute, Sioux Falls, SD (Dr. Baumgarten).

Dr. Wright or an immediate family member has received research or institutional support from Smith & Nephew, Wright Medical Technology, and DJO. Dr. Baumgarten or an immediate family member serves as a board member, owner, officer, or committee member of Sioux Falls Surgical Center and has received nonincome support (such as equipment or services), commercially derived honoraria, or other non-research–related funding (such as paid travel) from Wolters Kluwer Health—Lippincott Williams & Wilkins.

*J Am Acad Orthop Surg* 2010;18:436-444

Copyright 2010 by the American Academy of Orthopaedic Surgeons.
studies, and encourage validity testing of shoulder outcomes measures. The goal was to create a scoring system that could be applied to all patients regardless of diagnosis.

The questionnaire is divided into a physician-determined domain and a self-assessment domain. The former domain is rarely reported in clinical studies. The self-assessment portion is divided into three domains: pain, which includes several yes/no questions and a visual analogue scale (VAS); instability, which includes one yes/no question and a VAS; and activities of daily living (ADLs). Ten questions are rated on a four-point ordinal scale. The maximum score for the self-assessment portion is 30. This score is converted to a 100-point scale (100 points is the best score) using the formula:

\[
\frac{[10 - \text{VAS pain score}] \times 5 + 5}{3 \times \text{ADL score}}
\]

The pain score and the ADL score are weighted equally via this formula. No rationale has been presented for the weighting scheme of this instrument.\(^4\) Normative scores range from 92 to 99.\(^5\) The time to administer the test is 3 to 5 minutes, and scoring takes approximately 2 minutes.\(^6,7\) The ASES score is considered to be difficult to calculate because of the score conversion formula.\(^8\)

The ASES score has been validated in English and German.\(^7\) The latter required translation into German and validation in that language and population. Initial validation was performed in a population with varying shoulder diagnoses managed both surgically and nonsurgically in patients aged 20 to 81 years.\(^7\) The ASES has also been validated in patients with rotator cuff disease, glenohumeral osteoarthritis (OA), shoulder instability, and shoulder arthroplasty.\(^9,10\)

Construct validity, internal consistency, and reliability have in general been noted to be good, although they have been questioned in some situations.\(^7,8,10,11\) The minimal clinically important difference (MCID) is 6.4 ASES points, and the minimal detectable change is 9.7 ASES points (90% CI ≈ 16).\(^5\) Correlation to other scores depends on whether the ASES was compared with general shoulder scores or disease-specific scores.\(^12-14\)

### Constant Shoulder Score

The Constant Shoulder score, also known as the Constant score and the Constant-Murley score, was originally published as a master’s thesis in 1986.\(^15\) The methodology was published in 1987.\(^16,17\) This score was proposed by the European Society for Surgery of the Shoulder and the Elbow (SECEC) as an outcome measure for comparing shoulder function before and after treatment. The Constant score was recommended by the SECEC and the Journal of Shoulder and Elbow Surgery as the minimal dataset needed for presentations and communications.\(^17\)

This measure has a 100-point scoring system, with 100 points being the best score; 35 points are derived from the patient self-assessment portion. The patient-assessed pain domain is weighted at 15 points. The patient ranks her or his pain as none,
mild, moderate, or severe. The function domain is divided into ADLs and an objective domain that measures active range of motion (ROM) and power. Power was tested using a spring balance with the shoulder abducted to 90°. No rationale has been provided for the weighting of the individual items. Multiple modifications were made after the development of the initial scoring system. It may be inappropriate to combine patient-based and physical examination domains. These are fundamentally different issues and probably should not be combined in one total score. Age-related declines in scores and strength exist for both sexes. This test takes 10 to 15 minutes to administer and 2 minutes to score.

The Constant score was one of the original shoulder outcomes scores and therefore was not validated by comparison with other shoulder scores. In the original paper, validation was limited to study of the interobserver error in an unknown population. Validated ages ranged from 14 to 85 years. The first published validation study examined a heterogeneous population presenting with arthritis, dislocation, and impingement. The Constant score has been specifically validated for total shoulder arthroplasty, rotator cuff repair, adhesive capsulitis, and treatment of proximal humerus fractures. Although it has not been specifically validated for other entities, the Constant score has been used to assess many shoulder-specific conditions.

The reliability of this measure and its correlation with other shoulder scores are questionable. Multiple studies have suggested weaknesses and deficiencies with the Constant score. The original publication has been criticized for not sufficiently describing the precise information needed for exact application of the test and interpretation of the results.

Disabilities of the Arm, Shoulder and Hand Score

The Disabilities of the Arm, Shoulder, and Hand score (DASH) was developed in 1996 by the Council of Musculoskeletal Specialty Societies (now the Board of Specialty Societies), the American Academy of Orthopaedic Surgeons, and the Institute for Work and Health. This questionnaire was designed to be used for single or multiple disorders in the upper limb, with the intent that it might serve as a single questionnaire for measuring disability for any upper limb region.

The DASH is a patient-administered 30-item questionnaire that evaluates symptoms and physical function. It is divided into three domains: physical (21 items), symptoms (6 items), and social or role functions (3 items). It also contains two optional sections that produce scores for participation with regard to sports/music (4 items) and work activities (4 items). The response to each individual item is scored on a 5-point Likert scale, ranging from “no difficulty” to “unable,” from “none” to “extreme,” or from “no impact” to “high impact.” All items reference situations in the week leading up to completion of the questionnaire. The circled responses are summed, and 30 is subtracted from that total. (Subtracting 30 anchors the score with a base of 0, a correction that is required because the response scale is 1 to 5 and needs to be changed to a 0 to 4 equivalent.) The figure is then divided by 1.2 to get a DASH function/symptom score ranging from 0 to 100; a higher score reflects greater disability. Missing responses to items are replaced by the mean value of the responses to the other items before summing. The overall score cannot be calculated when responses to more than three items are missing. Normative data have been established. The DASH has an MCID of 10.2 points and a minimal detectable change of 12 points. The time to administer the test is 5 to 7 minutes, with a time to score of approximately 3 minutes. The ease of scoring the DASH is considered to be moderate because of the conversion formula needed to calculate the final score.

The DASH has been validated in many languages, including English, Swedish, Dutch, Chinese, Canadian French, German, Spanish, Brazilian Portuguese, Italian, Greek, Hungarian, Japanese, French, and Korean. There is no set age limit for the DASH; however, the recommended age range is 18 to 65 years. The DASH has been specifically validated for glenohumeral arthritis and rotator cuff tendinitis, total shoulder arthroplasty, rotator cuff repair, and psoriatic arthritis. Although it has not specifically been validated for other entities, this score has been used to assess many shoulder-specific conditions.

The quickDASH was developed in 2005, with the intent of saving time, facilitating use, minimizing the burden on the respondent, and minimizing missing data. Correlations between quickDASH and DASH have been reported to be extremely high (r > 0.97). Only one missing item can be tolerated on the quickDASH.

Shoulder Activity Level

Activity level scoring is important, especially for outcomes assessment in ath-
A patient can decrease symptoms by limiting activity postoperatively. Use of these scores pre- and postoperatively will help assess these activity aspects of outcome. The shoulder activity level was developed to be used in addition to traditional scores that measure pain and function. It consists of five activities: carrying objects weighing ≥8 lb by hand, handling objects overhead, weight lifting or weight training with the arms, executing a swinging motion (eg, hitting a golf ball, swinging a baseball bat), and lifting objects weighing ≥25 lb. These five items are scored on a frequency basis, from 0 to 4: performing the activity never or once per month (0 points), once per month (1 point), once per week (2 points), more than once per week (3 points), or daily (4 points). In addition, two questions determine whether the patient participates in contact sports or overhead throwing sports. The tool underwent reliability and validation testing during its development. Scores range from 0 (least active) to 20 (most active). Because the score is so new, it has been infrequently used in published studies.

Single Assessment Numeric Evaluation

The Single Assessment Numeric Evaluation (SANE) was designed as a simple one-question, patient-based shoulder function assessment tool. The question reads, “How would you rate your shoulder today as a percentage of normal (0% to 100%, with 100% being normal)?” This tool has not undergone formal validation. However, it has been used in a variety of clinical diagnoses. It was used in one study to measure shoulder instability outcome in patients aged 11 to 18 years. The advantages of this tool are its simplicity and its applicability to a wide variety of clinical situations. Formal reliability and validation testing would strengthen the use of this instrument.

UCLA Shoulder Score

The University of California Los Angeles (UCLA) shoulder score was first described in 1981, making it one of the earliest shoulder outcome measures. Although it lacks formal validation, we have included it because of its historic standing and continued popularity. Originally described to assess shoulder arthroplasty, it has subsequently been used to evaluate nearly all shoulder conditions. The UCLA shoulder score evaluates five domains, including pain, function, forward flexion, forward flexion strength, and overall satisfaction. Ten possible points were assigned to pain and function, with five possible points for each of the other domains, resulting in a potential score of 35 (best score). The reasons for weighting the scale with these point values were not described. The score combines patient-based subjective impressions with physical examination findings, which is probably not appropriate. Comparison studies have brought into question the psychometric quality of the UCLA score.

Western Ontario Osteoarthritis of the Shoulder Index

The Western Ontario Osteoarthritis of the Shoulder index (WOOS) was developed in Canada in 2001. The WOOS was meant to be used as the primary outcome measure in clinical trials involving patients with symptomatic primary OA of the shoulder. The WOOS is a 19-item self-administered questionnaire that was generated after reviewing the literature and preexisting scoring systems, interviewing orthopaedic surgeons and physical therapists, and interviewing patients with OA. The WOOS includes four domains: pain and physical symptoms; sport, recreation, and work; lifestyle function; and emotional function. Each item uses a 100-mm VAS. The score ranges from 0 to 1,900, with a higher raw score equating with worse function. The time to administer the test is 10 minutes, and the ease of scoring has been rated as moderate. The WOOS has been validated in English, French, Spanish, and German. The initial validation population was composed of patients with OA who were treated with arthroplasty. Although not specifically validated for assessment of arthroscopic débridement of OA, the WOOS has been used to assess this treatment. The WOOS was examined by its developers and was determined to be valid, reliable, and responsive.

Additional general shoulder outcomes measures and OA measures are listed in Table 2. These include the Croft measurement of shoulder-related disability (ie, The Disability Questionnaire), United Kingdom Shoulder Disability Questionnaire (SDQ-UK), the Flexilevel Scale of Shoulder Function (FLEX-SF), the Hospital for Special Surgery shoulder score, the L’Insalata Shoulder Rating Questionnaire (SRQ), the Neer shoulder score, the Oxford shoulder score, the Penn Shoulder Score (PSS), the Shoulder Disability Questionnaire (ie, Dutch Shoulder Disability Questionnaire [SDQ-NL]), the Shoulder Pain and Disability Index, the shoulder pain score, the Simple Shoulder Test (SST), and the Subjective Shoulder Rating System.

Rotator Cuff Disease Outcomes Measures

Rotator Cuff Quality of Life

The Rotator Cuff Quality of Life score was first reported in 2000 as
part of the evaluation of large and massive rotator cuff tears. This 34-question measurement uses a VAS, with 0 mm being the worst and 100 mm the best possible score for each response. It assesses five domains, including symptoms and physical complaints (16 items), sport and recreation (4 items), work concerns (4 items), lifestyle issues (5 items), and social and emotional issues (5 items). It is converted to a final score of 0 to 2,100, with a higher raw score indicating worse function. The raw score can be calculated into a percentage score using the following formula:

\[
\frac{2100 - \text{raw score}}{2100 \times 100} = \% \text{ score}
\]

The WORC has been validated in English, Turkish, Brazilian Portuguese, German, Persian, French, and Norwegian, with a validated age range from 20 to 84 years. The initial validation population included patients with both surgical and nonsurgical acute rotator cuff tendinitis, tendinosis, partial-thickness tears, full-thickness tears, and rotator cuff arthropathy. The MCID was 245.26 (11.7%). MacDermid et al recommended the WORC as the outcomes measure for clinical trials of patients undergoing rotator cuff repair. With this measure, the evaluators were able to discriminate among levels of severity of ROM impairment. The WORC has been shown to be valid, reliable, and responsive.

### Shoulder Instability Outcomes

The Western Ontario Shoulder Instability Index (WOSI) was established in Canada in 1998. The purpose was to develop a valid, reliable, and responsive disease-specific quality-of-life measurement tool for patients with shoulder instability. Items were determined based on a literature review and on discussions with patients, orthopaedic surgeons, sports and family physicians, and physiotherapists. There are five domains incorporating 21 questions, with a 100-mm VAS response. The domains include pain and physical symptoms (six questions), sports and recreation (four questions), work function (four questions), social function (four questions), and emotional function (three questions). The score ranges from 0 to 2,100, with a higher raw score indicating worse function. The raw score can be calculated into a percentage score using the following formula:

\[
\frac{2100 - \text{raw score}}{2100 \times 100} = \% \text{ score}
\]
on review of the literature as well as discussions with patients, orthopaedic surgeons, sports and family physicians, and physiotherapists. The WOSI consists of four domains incorporating 21 self-administered questions using a 100-mm VAS response. The domains include physical symptoms (10 items); sports, recreation, and work (4 items); lifestyle (4 items); and emotions (3 items). The score ranges from 0 to 2,100, with a higher raw score indicating worse function. The raw score can be calculated into a percentage score using the following formula:

\[
\frac{2100 - \text{raw score}}{2100 \times 100} = \% \text{ score}
\]

The WOSI has been validated in English and Swedish.\(^7^9\)\(^8^0\) The initial validation study consisted of patients undergoing treatment of shoulder instability; the initial validation study showed good validity, reliability, and responsiveness.\(^7^9\) The MCID was 220 (10.4%).\(^4\)

**Discussion**

A general health outcome measure, such as the SF-36, should always be included as part of the patient evaluation to allow for comparison with other musculoskeletal and systemic diseases. When possible, the surgeon should also use outcome measures that are validated for the specific disease condition being studied. A general cross-sectional shoulder-specific measure allows comparison for different diagnoses.

Several general shoulder measures are available, including the ASES, Constant, DASH, L’Insalata Shoulder Rating Questionnaire, simple shoulder test, SANE, and, for surgical intervention, the Oxford shoulder score. We believe that given the established popularity of the ASES with shoulder surgeons in American and European shoulder societies, it should be used for research purposes as a general shoulder measure. The Constant score is not recommended because of problems with examiner bias and variability on some of the ROM and strength parameters. Clinicians who want to quickly and simply evaluate outcomes for a variety of diagnoses will find the ASES, UCLA, and SANE to be helpful. The DASH has become popular, and in some situations it is the standard for the assessment of patients for workers’ compensation claims.

Activity level measures are not as popular for the shoulder as they are for assessment of knee functional outcome. Low activity levels can falsely elevate many of the outcomes measure scores. This is especially true for assessment of sports injuries, such as shoulder instability and labral injury. Thus, an activity level score is critical for comparing outcomes among patients. A validated shoulder activity level was recently developed, and it should be included in all future shoulder outcome studies.\(^46\)

Specifically, for rotator cuff disease, including nonsurgical and surgical management of rotator cuff tears and impingement syndrome, the validated Rotator Cuff Quality of Life and WORC scores should be used. Both are validated and were developed using appropriate psychometric techniques. The WORC has the advantage of an established MCID. Neither outcome measure appears to be more user- or patient-friendly than the other.

When evaluating the results of shoulder instability treatment, the Oxford Shoulder Instability Questionnaire and the WOSI are appropriate outcome measures to use. If comparison to a historical group is critical, the Rowe score can be used.\(^8^1\) This unvalidated measure has been frequently used in the past. We prefer the WOSI. It has a reasonable responder burden, with only 21 questions, and previous studies have established the MCID change in scores.

The validated WOOS is the best tool for assessing the results of shoulder arthroplasty and arthritis treatment. This measure has undergone rigorous psychometric evaluation and has a reasonable responder burden of 19 questions. The Neer and UCLA scores can be used to offer historical perspective, but neither has been formally validated.

Combining a general health outcome measure, a general shoulder measure, a disease- or condition-specific shoulder measure, and an activity measure allows for broad patient assessment. Such evaluation can be accomplished using validated, patient-based outcome measures representing the highest level of assessment of outcome.

**References**

Evidence-based Medicine: Levels of evidence are described in the table of contents. In this article, reference 73 is a level I study.

Citation numbers printed in bold type indicate references published within the past 5 years.


56. Rick W. Wright, MD, and Keith M. Baumgarten, MD

July 2010, Vol 18, No 7


