INTRODUCTION

As with all musculoskeletal conditions, outcome assessment of conditions affecting the shoulder should utilize a patient-centered approach. Careful consideration should be given to the clinical question and the disease state that is being studied in order to use an appropriate outcome measure when choosing which outcome measures to use. For maximum comparability and best overall assessment, a combination of general health with a region specific and disease or population-specific instrument is often used with or without additional metrics to measure activity level, comorbidities, or patient expectations. When choosing outcome tools, it is important to use a metric that has been rigorously validated. When performing research, such consideration is best given early in the study design process as the selected outcome instruments will be used to calculate the study's sample size for a given power, based on an estimate of the desired, or clinically significant, effect of the investigation. When evaluating outcomes for a performance metric, patient satisfaction scales should be included.

When selecting from the outcome instruments available, the individual tools for measuring outcomes can be separated into discrete categories to include measures of general health, measures of upper extremity function, measures of shoulder activity level, shoulder-specific questionnaires, and condition or population-specific shoulder instruments.

The minimum clinically important difference (MCID) represents the smallest improvement considered worthwhile by a patient. This difference has to be determined for each outcome measure. While values less than this may be statistically significant, they are not clinically meaningful. This is of value in interpreting the literature and in power analyses.

Understanding how new measurement tools are created can help physicians to evaluate critically the instruments that are currently available. Major elements for the development of health-related quality of life tools include specifying measurement goals for the population being studied, question generation, item reduction, and questionnaire formatting. Question generation is carried out through reviewing the literature, interviewing experts, and interviewing patients with the condition being studied. The tool that was created then undergoes extensive pretesting, as well as testing for reliability (intra and inter-rater), responsiveness, validation, and interpretability. Interpretability places the magnitude of changes seen in the measurement tool in the context for people interpreting the results. This is best done by defining the MCID for each tool. Outcomes instruments developed using less rigorous methodology are more prone to biases and error.

More than 30 shoulder outcome measures have been described. The most popular and rigorously validated shoulder outcome measures for use in patient evaluation and research efforts are discussed in this chapter.

OUTCOME MEASURES

Measures of General Health

- Short Form-36 (SF-36)
- Short Form-12 (SF-12)
- Veterans RAND 36 Health Item Survey (VR-36)
- EuroQol-5D (EQ-5D)

Short Form-36

The SF-36 is the most widely used and best-established generic instrument for assessing physical, mental, and psychosocial health. It is based on eight health domains (vitality,
physical functioning, bodily pain, general health perceptions, physical role functioning, emotional role functioning, social role functioning, and mental health) that can be aggregated into two component summary scales (Physical Component Summary and Mental Component Summary) containing a total of 36 items.

The form is a registered trademark of the Medical Outcomes Study by RAND Health. Version 2 was copyrighted by QualityMetric Incorporated (2000) for commercial use. A commercial site, www.qualitymetric.com, provides information about obtaining a license to use the SF-36 v. 2 and an online version of the standard form. In addition, this site provides automated scoring that is normalized for the general population of the United States (an average score is 50 and the standard deviation is 10). In general, patients should complete the form at least 4 weeks following an incident that affects their health. An advantage of the SF-36 is the ability to convert to a preference-based measure of utility, the SF-6D, which can be used for cost-utility analyses and comparative effectiveness research.

When a comprehensive overview of the patient’s state of health is required, a general outcome score should be included. For the SF-36, normative data are available in the United States and German populations, allowing for accurate comparisons of the relative degree of disability.10

Short Form-12
The Short Form-12 (version 2) is a 12-question version of the SF-36 that measures the same eight health domains with only one to two questions per domain. This can also be converted to a SF-6D. Information about usage and how to purchase licenses to use can be found at the same website, www.qualitymetric.com.

RAND
The Veterans RAND 36-Item Health Survey 1.0 includes the same items as those in the SF-36, but the recommended scoring algorithm is somewhat different for the general health and pain scales than that of the SF-36.23 The RAND is available in the public domain license free from RAND, which is beneficial, as issues with licensing can be avoided.

There is also a Veterans RAND 12-Item Health Survey (VR-12), similar to the SF-12. A preference-based measure of utility, the SF-6D, can be used for cost utility analyses and comparative effectiveness research. More information on the RAND can be found at http://www.rand.org/health/surveys_tools/mos/mos_core_36item.html.

EuroQol-5D
The EuroQol-5D is a five-question standardized instrument to measure health-related quality of life that can be used for clinical and economic analyses. There are five dimensions (mobility, self-care, usual activities, pain/discomfort, and anxiety/depression) and the EQ VAS—a 20 cm vertical visual analog scale that generates a self-rating of health-related quality of life. The EQ-5D has been well validated and has been shown to be responsive. There are 102 official language versions, and population norms are available for 15 countries.

Written consent of the EuroQol Executive Office is required for use, and there may be licensing fees for usage.

Shoulder Activity Level
Scoring activity level is important, as a patient can decrease symptoms (and falsely elevate scores) by limiting his or her activity level. Use of these scores pre- and postoperatively can help assess any activity level changes. Measuring activity level can be very helpful after sports injuries as well.

The shoulder activity level metric was developed for use 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 (swinging a baseball bat or golf club), and lifting objects weighing ≥25 lb. These five items are scored on a frequency basis from 0 to 4: performing the activity never to 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, and the MCID has not been established

Measure of Upper Extremity Function
• Disabilities of the Arm, Shoulder, and Hand (DASH).

Disabilities of the Arm, Shoulder, and Hand
The DASH outcome measure is a 30-item, self-report questionnaire designed to measure physical function and symptoms in people with any of several musculoskeletal disorders of the upper limb. It was developed in 1996 by the Council of Musculoskeletal Specialty Societies, the American Academy of Orthopaedic Surgeons, and the Institute for Work and Health. The tool gives clinicians and researchers the advantage of having a single, reliable instrument that can be used to assess any or all joints in the upper extremity.27

The DASH is a 30-item questionnaire that measures symptoms (six items, three of which concern pain) and function (24 items) of the upper extremity.7,43 Originally, the DASH was a one-dimensional instrument, but subsequent division into subscales for pain and function has been well described.4 There are two optional modules (work and sports/music) available consisting of four items each, a work module and a sports/performing arts module. All questions reference the prior week, and each individual question is scored on a 5-point Likert scale. Up to three missing responses to items can be replaced by the mean value of the responses to the other items before summing.33 Normative data have been established. While the survey was small (n = 1657), these population-based data are valid within its limitations (age- and sex-specific substrata in the United States).28 The DASH has an MCID of 10.2 points.7

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.17,21,29,36,44,50,52 The DASH has been specifically validated for glenohumeral arthritis and rotator cuff tendinitis, total shoulder arthroplasty, rotator cuff repair, and psoriatic arthritis.3,39,42 Although it has not specifically been validated for other
entities, this score has been used to assess many shoulder-specific conditions.

Relative to other measures, the DASH allows for comparison of different conditions of the upper extremity, including different joints, and constitutes a comprehensive assessment of the whole arm. Further information on the DASH can be found at http://www.dash.iwh.on.ca/index.htm.

A shorter version called the QuickDASH was developed in 2005 to facilitate use, minimize the burden on the respondent, and minimize missing data. Only one missing item is permitted on the QuickDASH. Correlations between QuickDASH and DASH are extremely high ($r > 0.97$).

In the latest edition of the AMA Guides to the Evaluation of Permanent Impairment, the QuickDASH is recommended as an adjunct to assess functional impairment when determining disability ratings related to the upper extremity.

Shoulder-Specific Measures

- American Shoulder and Elbow Surgeons (ASES)
- Constant Score
- Simple Shoulder Test (SST)
- Shoulder Pain and Disability Index (SPADI)
- Single Assessment Numeric Evaluation (SANE)
- University of California at Los Angeles (UCLA) Shoulder Scale
- Oxford Shoulder Score (OSS)

American Shoulder and Elbow Surgeons

The ASES score was developed in 1994 by the research committee of the ASES with the goal of creating a scoring system that could be applied to all patients regardless of diagnosis. It is a composite tool that involves both physician assessment as well as a patient-reported component (the physician portion is not commonly reported in the literature). The clinical assessment section provides scores on symptoms/pain (11 questions) and function (AROM, five questions). The patient self-assessment section is divided into three domains: pain, instability, and activities of daily living. The ASES does not have an established method for handling missing data. The minimal clinically important difference is 6.4 points. The ASES score has been validated in English and German in surgical and nonsurgical patients 20 to 81 years old. The ASES has also been validated in patients with osteoarthritis (OA), shoulder instability, rotator cuff pathology, and shoulder arthroplasty.

Constant Score

The Constant Shoulder score, also known as the Constant-Murley score, was originally published as a master’s thesis in 1986, with the methodology published in 1987. The Constant score was recommended by the European Society for Surgery of the Shoulder and the Elbow and the Journal of Shoulder and Elbow Surgery as the minimal dataset needed for presentations and communications.

The Constant Score is an aggregate score of examiner and patient self-assessed symptoms and examiner-based performance tests for range of motion and power. Power is tested using a spring balance with the shoulder abducted to 90 degrees. It is important to note that in the evaluation of knee OA and low back pain, examiner-measured parameters, such as ROM, have been found to correlate poorly with the patient’s self-reported health and are not always useful for outcome assessment. Due to problems with examiner bias and variability on some of the ROM and strength parameters which account for 65 of the total 100 points, the Constant score is not recommended as a primary outcome instrument. Since patient-reported outcomes and physical examination findings have different potential biases and limitations, combining them into a single score can be problematic.

In the original paper, validation was carried out in an unknown population aged 14 to 85. The Constant score has been validated for proximal humerus fractures, total shoulder arthroplasty, rotator cuff repair, and adhesive capsulitis.

Simple Shoulder Test

The SST was developed by Matsen at the University of Washington and consists of 12 questions about the function of the involved shoulder. In a study of 1077 patients with shoulder instability and rotator cuff injuries ranging from 14 to 85 years old, the SST was reliable and demonstrated significant construct and content validity. Correlations with the ASES were significant. Advantages are that it includes only 12 questions, is widely used, and is free for use. A copy can be downloaded at http://www.orthop.washington.edu/PatientCare/OurServices/ShoulderElbow/Articles/SimpleShoulderTest.aspx.

Shoulder Pain and Disability Index

The SPADI assesses pain and routine functional skills using 13 items (5 pain items, 8 function items). A 10-point change in the score in either direction accurately distinguishes between patients whose shoulder problems improve or worsen. The SPADI form is available in a publication by Williams, Holleman, and Simel, and various websites provide downloadable versions of the disability scale. Currently, there are no normative data for the SPADI; however, the cross-cultural applicability has been examined. The utility of the SPADI lies in its nature as a short, shoulder-specific metric that is easily used in daily clinical practice where time is of the essence. The MCID comprises 8 to 13 points.

Single Assessment Numeric Evaluation

The SANE asks patients a single question to evaluate their shoulder health. Patients are instructed to provide the SANE as a percentage of normal? (0% to 100% scale with 100% being normal). Patients are instructed to provide the SANE rating as a whole number. This tool is simple and applicable to many diagnoses. While it appears to have face validity, in its original description the authors only examined convergent validity against the Rowe and ASES.

University of California at Los Angeles

The UCLA shoulder scale was described in 1981 by Amstutz et al. as a rating scheme for total shoulder arthroplasty patients. This is one of the first shoulder outcome scores described, and is commonly used throughout the literature. While it was originally designed for shoulder arthroplasty patients, it has subsequently been applied to other shoulder conditions in a modified format. 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. The reasons for weighting the scale with these point values were not described. This score’s weakness are its lack of validation, as well as combining patient-reported outcome data with physical examination findings.

**Oxford Shoulder Score**
The OSS is a validated, patient-reported, shoulder-specific instrument that was originally devised to assess outcomes in randomized trials that involved shoulder surgery, except stabilization for instability for which there is the Oxford shoulder instability score. The OSS was developed using modern psychometric methodology that used patient’s input for item generation. The OSS includes 12 items each having five levels of response, and the scoring has been recently revised such that each item are scored from 0 (worst) to 4 (best), and the items are summed to give an overall score ranging from 0 (worst) to 48 (best). There are rules to handle missing values provided there are no more than two items missing (replace them with the mean of the other answered items); if more than two are missing, it is recommended that an overall score not be calculated. The instrument has been translated and validated in German and Turkish.

**Condition or Population-Specific Shoulder Instruments**

- **Shoulder Instability**
  - Western Ontario Shoulder Instability Index (WOSI)
  - Rotator Cuff Tears
  - Western Ontario Rotator Cuff Index (WORC)
  - Rotator Cuff Quality of Life (RCQOL)
  - Glenohumeral Arthritis/Shoulder Arthroplasty
  - Western Ontario Osteoarthritis of the Shoulder (WOOS)
  - Population-Specific Instruments
    - Kerlan Jobe Orthopaedic Clinic Score (KJOC Score)

**Western Ontario Shoulder Instability Index**
The WOSI was established at McMaster in 1998 as a valid, reliable, and responsive measurement tool for patients with shoulder instability. Items were determined based on review of the literature as well as discussions with orthopedic surgeons, sports and family physicians, therapists, and patients. The WOSI consists of 21 questions in four domains, using a 100-mm VAS response. The domains include physical symptoms (10 questions), sports/recreation/work (4 questions), lifestyle (4 questions), and emotional function (3 questions). The WOSI has been shown to have good validity, reliability, and responsiveness and is validated in English and Swedish. The MCID is 220 points.

**Western Ontario Rotator Cuff Index**
The WORC was developed in 2003. As with the WOSI, items were determined based on a literature review and on discussions with orthopedic surgeons, sports and family physicians, therapists, and patients. It consists of 21 questions in five domains, with 100-mm VAS responses for each item that are summed to a total maximum score of 2100; a higher score indicates reduced quality of life. The domains include pain and physical symptoms (six questions), sports/recreation (four questions), work function (four questions), social function (four questions), and emotional function (three questions). The WORC has been shown to be valid, reliable, and responsive. The WORC has been validated with an age range from 20 to 84 years old in English, Turkish, Brazilian Portuguese, German, Persian, French, and Norwegian. The initial validation included patients with surgical and nonsurgical acute rotator cuff tendinitis, tendinosis, partial-thickness tears, full-thickness tears, and rotator cuff arthropathy. Kirkley et al. estimated the MCID to be 245. Ekeberg et al. compared the OSS, SAPI, and WORC in subjects with rotator cuff disease and found them all suitable for measuring change in their study population, and reported MCIDs of 5, 20, and 275, respectively.

**Rotator Cuff Quality of Life**
The RCQOL score was developed in 2000 to assess pathology of the rotator cuff. It consists of 34 questions in five domains, with a 100-mm VAS response.

The five domains are symptoms/physical complaints (16 questions), sport/recreation (4 questions), work (4 questions), lifestyle (5 questions), and social/emotional issues (5 questions). While the age of patients studied during the development of the RCQOL was not specified, it has been used on patients 25 to 83 years old. It has been translated into German. The MCID has not been established.

**Western Ontario Osteoarthritis of the Shoulder**
The WOOS index is a rigorously designed measurement tool for shoulder OA developed in 2001. The WOOS was meant to be used as the primary outcome measure in trials involving patients with symptomatic, primary shoulder OA. As with the WORC and WOSI, questions were determined based on a literature review and on discussions with orthopedic surgeons, sports and family physicians, therapists, and patients. The WOOS consists of 19 questions in four domains, with a 100-mm VAS response. The four domains are physical symptoms, sport/recreation/work, lifestyle function, and emotional function. The WOOS was examined by its developers and was determined to be valid, reliable, and responsive. The initial validation population was composed of patients with OA who were treated with arthroplasty. The WOOS has been validated in English, French, Spanish, and German. The WOOS has also been used to evaluate arthroscopic debridement of shoulder OA, even though it was not specifically validated for this.

**Kerlan Jobe Orthopaedic Clinic Score**
The KJOC score was developed to specifically evaluate the performance of overhead athletes. This is a 10-item format that is similar to the SAS with patients instructed to place a mark on a 10-cm line indicating their current level of performance and/or function. The maximum score is 100. Items are specific to function and athletic performance, symptoms, and interpersonal relationships related to performance. This questionnaire was correlated with the DASH and DASH sports module by 282 competitive overhead athletes and was found to be valid and responsive in this group.
SUMMARY

When deciding which outcome scores to use for shoulder conditions, combining a general health outcome measure, a general shoulder measure, a condition-specific shoulder measure, and an activity measure allows for broad patient assessment. Scores that are valid, reliable, and highly responsive should be used preferentially. Scores with a known MCID are more easily interpretable and allow for power and sample size estimation.

REFERENCES


