

CURRENT CONCEPTS REVIEW

Upper Extremity-Specific Measures of Disability and Outcomes in Orthopaedic Surgery

Matthew V. Smith, MD, Ryan P. Calfee, MD, Keith M. Baumgarten, MD, Robert H. Brophy, MD, and Rick W. Wright, MD

Investigation performed at the Department of Orthopaedic Surgery, Washington University School of Medicine, Saint Louis, Missouri

- ▶ Outcome measures may consist of simple questions or they may be more complex instruments that evaluate multiple interrelated domains that influence patient function.
- ▶ Outcome measures should be relevant to patients, easy to use, reliable, valid, and responsive to clinical changes.
- ▶ The Disabilities of the Arm, Shoulder and Hand score can be used to measure disability for any region of the upper limb.
- ▶ Joint and disease-specific outcome measures have been developed for the shoulder, the elbow, and the wrist and hand. Many of these measures would benefit from further research into their validity, reliability, and optimal applicability.

While outcome measures have long been used in medical research, there is increasing interest in the use of outcome measures in orthopaedic surgery¹. Measures vary from simple questions regarding patient satisfaction or ability to return to work to more complex scoring systems that assess multiple interrelated domains that affect outcomes such as pain, activities of daily living, and objective physical examination measures. Outcome measures may include physician and/or patient-based assessments and can focus on general health or a specific joint or disease process².

In general, outcome measures should be relevant to patients, easy to use, reliable, valid, and responsive to clinical changes³. A reliable outcomes instrument produces consistent results when repeated under similar conditions. An instrument is valid if it truly measures what it is intended to measure. Validity has several facets (construct validity, criterion validity, and content validity). Therefore, the validity of an instrument is based on a body of evidence rather than a single test. An instrument is responsive if it is able to detect a change in a condition when it

occurs. The *minimal clinically important difference* of an outcome measure is the minimum change in a score that indicates a change in disability. For example, if the minimal clinically important difference is 15 points for an outcome measure and a patient has a 10-point change on that measure after treatment, then there may not have been a true clinical change for the patient despite a numerical change in score. Lastly, outcome measures may have *ceiling* or *floor effects*, meaning that the outcome measure is unable to assess improvement (ceiling) or deterioration (floor) in a condition. For example, a ceiling effect may occur if a baseball player with shoulder pain scores so well on an instrument that he cannot score higher on that measure after the pain has resolved. Therefore, a ceiling effect occurs because the outcome measure may not be able to detect improvements because the subjects already approach the maximum score for the measure at their baseline.

The goal of this review is to outline common instruments used to measure outcomes for the upper extremity. We highlight why each measure was developed, the conditions each

Disclosure: None of the authors received payments or services, either directly or indirectly (i.e., via his or her institution), from a third party in support of any aspect of this work. One or more of the authors, or his or her institution, has had a financial relationship, in the thirty-six months prior to submission of this work, with an entity in the biomedical arena that could be perceived to influence or have the potential to influence what is written in this work. No author has had any other relationships, or has engaged in any other activities, that could be perceived to influence or have the potential to influence what is written in this work. The complete **Disclosures of Potential Conflicts of Interest** submitted by authors are always provided with the online version of the article.

TABLE I Upper Extremity Outcome Measures for the Shoulder

Scale*	Anatomic Region	Measures	Validated	Responder Burden		Target Population		MCID†
				Clinician	Patient (no. of questions)	Age (yr)	Disorders	
DASH ⁴	Shoulder, arm, and hand	Function (bilateral)	Yes	None	38	18-65	All upper extremity diagnoses	10-17
ASES ¹³	Shoulder	Pain, instability, activities of daily living	Yes	Rarely used	10	20-81	All shoulder diagnoses	6.4 (12-17 for RTC)
Constant ¹⁶	Shoulder	Function	Yes	6 items	2	14-85	Rotator cuff repair, shoulder arthroplasty, adhesive capsulitis, and proximal humeral fractures	NA
Shoulder activity level ²³	Shoulder	Activity level	Yes	None	7		Rotator cuff, osteoarthritis, and instability	NA
SANE ²⁵	Shoulder	Patient perception	No	None	1		All shoulder diagnoses	NA
UCLA Shoulder Score ²⁶	Shoulder	Pain, function, forward flexion, forward flexion strength, and patient satisfaction	No	2 items	3		All shoulder diagnoses	NA
WOOS ^{32,33}	Shoulder	Pain, physical symptoms, sports, recreation, work, and emotional	Yes	None	19		Osteoarthritis	NA
RCQOL ³⁴	Shoulder	Symptoms, sports, work, lifestyle, and social and emotional	No	None	34	25-83	Rotator cuff pathology	NA
WORC ³⁶	Shoulder	Pain, physical symptoms, sports, recreation, work, and emotional	Yes	None	20	20-84	Rotator cuff pathology	245
WOSI ³⁷	Shoulder	Pain, physical symptoms, sports, recreation, work, and emotional	Yes	None	21		Shoulder instability	220

*DASH = Disabilities of the Arm, Shoulder and Hand, ASES = American Shoulder and Elbow Surgeons, SANE = Single Assessment Numeric Evaluation, UCLA = University of California Los Angeles, WOOS = Western Ontario Osteoarthritis of the Shoulder index, RCQOL = Rotator Cuff Quality of Life, WORC = Western Ontario Rotator Cuff index, and WOSI = Western Ontario Shoulder Instability index. †MCID = minimal clinically important difference, RTC = rotator cuff, and NA = not available.

has been designed to assess, and provide the minimal clinically important difference, if available. The outcome measures are summarized in Table I (shoulder), Table II (elbow), and Table III (hand and wrist).

Disabilities of the Arm, Shoulder and Hand Score

Through the combined efforts of the Council of Musculoskeletal Specialty Societies, the American Academy of Orthopaedic Surgeons, and the Institute for Work and Health in Toronto, Ontario, Canada, the Disabilities of the Arm, Shoulder

and Hand (DASH) score was introduced in 1996⁴. Designed to evaluate single or multiple disorders in the upper limb, this questionnaire can be used to measure disability for any region of the upper limb⁵. The DASH has been shown to be valid and responsive compared with other joint-specific measures of the upper extremity⁵. The DASH has been shown to correlate to general health measures such as the Short Form-36 (SF-36)⁶.

The DASH is a self-administered questionnaire that consists of thirty core questions and an optional additional eight questions assessing work and sports and/or performing

TABLE II Upper Extremity Outcome Measures for the Elbow

Scale*	Anatomic Region	Measures	Validated	Responder Burden		Target Population		MCID†
				Clinician	Patient (no. of questions)	Age (yr)	Disorders	
DASH ⁴	Shoulder, arm, and hand	Function (bilateral)	Yes	None	38	18-65	All upper extremity diagnoses	10-17
ASES ³⁸	Elbow	Pain, function, strength, range of motion, stability, and patient satisfaction	Yes	38 items	19		General elbow disorders	NA
Broberg and Morrey ⁴²	Elbow	Motion, strength, stability, and pain	Yes	5 items	1		General elbow disorders	NA
Mayo Clinic performance index ⁴⁵	Elbow	Pain, motion, stability and daily function	Yes	3 items	5		General elbow disorders	15
HSS assessment scale ⁴⁶	Elbow	Pain, duration of bending activities, activity level for the elbow, strength, and range of motion	Yes	6 items	3		General elbow disorders	12
Liverpool Elbow Score ⁴⁷	Elbow	Range of motion, strength, ulnar nerve sensitivity, pain, activities of daily living, and recreational activities	Yes	6 items	9		General elbow disorders	NA
Conway-Jobe score ⁴⁸	Elbow	Return to sport	No	None			Throwing athletes	NA
Timmerman and Andrews ⁴⁹	Elbow		No	3 items	4		Posttraumatic elbow injury and throwing athletes	NA

*DASH = Disabilities of the Arm, Shoulder and Hand, ASES = American Shoulder and Elbow Surgeons, and HSS = Hospital for Special Surgery. †MCID = minimal clinically important difference, and NA = not available.

arts activities. Each individual item is scored on a 5-point Likert^{7,8} scale, with lower scores correlating to minimal impairment and higher scores indicating more impairment. The cumulative DASH score is scaled from 0 to 100, with higher scores indicating more disability⁵. This score is unique in its instructions for subjects to rate their ability to complete activities regardless of which hand or arm is needed to perform that activity. As such, the questionnaire produces a score of patient function representing the composite abilities of both upper extremities. This feature is both an advantage and limitation of the instrument.

While there is no set age limit for the use of the DASH, it is recommended for patients who are eighteen to sixty-five years old. The validity of the DASH score in intercollegiate athletes may be limited by a ceiling effect because of the high overall function of this population⁹. In one investigation analyzing the minimal clinically important difference for the DASH with use of a variety of upper extremity diagnoses, a minimal clinically important difference of 10 was suggested for the proximal part of the upper extremity (shoulder) diagnoses and 17 was suggested for the distal part of the upper extremity (elbow, wrist, and hand) diagnoses¹⁰.

In 2005, the QuickDASH was developed to minimize time and responder burden¹¹. The QuickDASH consists of eleven questions scored on a 5-point scale similar to the DASH. A lower score indicates less disability. There is a high correlation between the QuickDASH and DASH scores ($r > 0.97$)¹¹. Although the QuickDASH has the advantage of reduced responder burden over the DASH, only one question can be skipped to still have a valid score. Depending on the data being collected and the population being studied, this may limit the usefulness of the QuickDASH. The minimal clinically important difference for the QuickDASH has been estimated to be 19, and the minimal detectable change for the QuickDASH has been estimated to be 11 points¹².

Shoulder Outcome Measures

General Measures

American Shoulder and Elbow Surgeons (ASES)

Shoulder Outcome Score

The research committee of the ASES developed this score in 1994 to be applicable to all shoulder diagnoses¹³. It consists of a physician-rated and a patient-rated questionnaire. The physician-rated outcomes are not commonly reported. The patient-rated

TABLE III Upper Extremity Outcome Measures for the Hand

Scale*	Anatomic Region	Measures	Validated	Responder Burden		Common Target Population		MCID†
				Clinician	Patient (no. of questions)	Age (yr)	Disorders	
DASH ⁴	Shoulder, arm, and hand	Function (bilateral)	Yes	None	38	18-65	All upper extremity diagnoses	10-17
Boston Carpal Tunnel Questionnaire ⁵⁰	Hand	Pain, sensibility, weakness, and function	Yes	None	19	Adult	Carpal tunnel syndrome	NA
Jebsen-Taylor Hand Function Test ⁵⁴	Hand	Timed functional task completion	Yes	Time Tasks	7 tasks	All	General hand and wrist conditions as well as neurologic conditions affecting upper extremity	NA
Michigan Hand Outcomes Questionnaire ^{59,60}	Hand	Hand function, daily activities, work activity, pain, appearance, and satisfaction	Yes	None	71	Adult	General hand and wrist disorders	11-23‡
Patient-Rated Wrist Evaluation ⁶⁴	Hand	Pain, daily activities, recreation, and work activities	Yes	None	15	Adult	General hand and wrist disorders	NA

*DASH = Disabilities of the Arm, Shoulder and Hand. †MCID = minimal clinically important difference, and NA = not available. ‡Values domain and diagnosis-specific.

questionnaire is divided into three domains that assess (1) pain, (2) instability, and (3) activities of daily living. For the patient-rated assessment, there are ten questions rated on a 4-point ordinal scale (from 0 to 3) to give a maximum score of 30 points. This score is then converted to a 100-point scale with higher scores indicating better outcome.

This score was validated in a group of patients with a wide range of shoulder diagnoses, treated both operatively and nonoperatively, who ranged from twenty to eighty-one years of age¹⁴. An estimation of the minimal clinically important difference and the minimal detectable change was 6.4 ASES points and 9.7 ASES points, respectively¹⁴. Another recent estimate of the minimal clinically important difference for nonoperative treatment of rotator cuff disease was between 12 and 17 ASES points¹⁵.

The Constant Shoulder Score

The Constant shoulder score was originally developed as a university thesis in 1986¹⁶. It has been used as an outcome measure comparing shoulder function before and after treatment. The Constant score did not undergo formal validation during its development. Recently, the Constant score has been specifically validated for total shoulder arthroplasty¹⁷, rotator cuff repair, adhesive capsulitis of the shoulder¹⁸, and proximal humeral fractures¹⁹. The Constant score is a 100-point scale of which 35 points are based on patient-rated assessment. Higher scores indicate a better outcome. Multiple modifications were made after the

development of the initial scoring system because of concerns regarding reliability, correlation with other shoulder scores, scoring methods that increase the risk for bias, and variability in objective testing measures²⁰⁻²². The minimal clinically important difference is unknown.

Shoulder Activity Level

This instrument was developed to measure the level of shoulder activity in addition to traditional scores that measure pain and function²³. It consists of five activities that are scored from “never or less than once a month” to “daily,” as well as asking two questions about sports participation. The score includes a numeric component ranging from 0 to 20, with 20 being the most active, and an alpha score. It underwent reliability and validation testing during its development and has been evaluated in patients with rotator cuff disease, glenohumeral osteoarthritis, and glenohumeral instability²⁴. The minimal clinically important difference is unknown.

Single Assessment Numeric Evaluation (SANE)

The SANE is a simple, single-question, patient-based shoulder function assessment tool: “How would you rate your shoulder today as a percentage of normal (0% to 100% scale with 100% being normal)?”²⁵ It is popular because of its simplicity and ability to be applied to a wide variety of clinical situations. It has not been validated. No minimal clinically important difference has been reported.

University of California Los Angeles (UCLA) Shoulder Score

The UCLA Shoulder Score, first described in 1981, is one of the earliest available shoulder outcome measures²⁶. It has been used in evaluating essentially all shoulder conditions²⁷⁻³¹. It assesses five domains: (1) pain, (2) function, (3) forward flexion, (4) forward flexion strength, and (5) overall satisfaction. A potential score of 35 is possible, with higher scores indicating a better outcome. It has not been formally validated. The minimal clinically important difference has not been established.

The Western Ontario Osteoarthritis of the Shoulder (WOOS) Index

The WOOS index was developed in 2001 as the primary outcome measure in clinical trials evaluating patients with osteoarthritis of the shoulder^{32,33}. It is a nineteen-item self-administered questionnaire that assesses four areas: (1) pain and physical symptoms; (2) sports, recreation, and work; (3) lifestyle function; and (4) emotional function. A 100-mm visual analog scale (VAS) is used to score each item. A total score ranges from 0 to 1900, with a higher score indicating a worse outcome. Its developers determined that it was valid, reliable, and responsive in patients with osteoarthritis treated with arthroplasty³². The minimal clinically important difference has not been reported.

Rotator Cuff Disease Outcome Measures**Rotator Cuff Quality of Life (RCQOL)**

The RCQOL score was developed to evaluate large and massive rotator cuff tears³⁴. It consists of thirty-four questions that assess (1) symptoms and physical complaints, (2) sports and recreation, (3) work, (4) lifestyle, (5) social issues, and (6) emotional issues. It is scored on a VAS and is converted to a total score ranging from 0 to 100, with higher scores indicating a better outcome. It has been used on patients from twenty-five to eighty-three years old to assess all forms of rotator cuff pathology³⁵. The minimal clinically important difference has not been reported.

Western Ontario Rotator Cuff Index (WORC)

The WORC index was developed in 2003 as a patient-assessed measurement tool for rotator cuff disease³⁶. It is a valid and reliable disease-specific quality-of-life index that assesses (1) pain and physical symptoms, (2) sports and recreation, (3) work function, (4) social function, and (5) emotional function³⁶. The questionnaire consists of twenty-one questions that are scored with use of a 100-mm VAS. The total score ranges from 0 to 2100, with higher scores indicating a worse outcome. It has been validated for patients from twenty to eighty-four years old^{35,36}. The minimal clinically important difference is 245.26 (11.7%)³³.

Shoulder Instability Outcome Measure**The Western Ontario Shoulder Instability (WOSI) Index**

The WOSI index was devised in 1998 as a valid, reliable, and responsive disease-specific quality-of-life measurement tool for patients with shoulder instability³⁷. The WOSI consists of twenty-one questions assessing (1) physical symptoms; (2) sports, recreation, and work; (3) lifestyle; and (4) emotions. It is scored with use of a 100-mm VAS. The score ranges from 0 to

2100, with a higher score indicating a worse outcome. The minimal clinically important difference is 220 (10.4%)³³.

Elbow Outcome Measures**General Measures****American Shoulder and Elbow Surgeons (ASES)****Elbow Outcome Score**

The ASES Elbow Outcome Score was developed by the ASES research committee in 1999³⁸ and is designed to measure elbow function independently of the diagnosis. It consists of a nineteen-question patient self-evaluation of pain, function, and satisfaction with surgery and a thirty-eight-item physician assessment of shoulder motion, strength, and stability. Pain has the highest influence (57%) on any variability of the score³⁹. The original description of the measurement did not include information pertaining to global scoring of the assessment³⁷. Its subscales have been converted to a global score in a manner similar to the ASES shoulder score, with a range from 0 to 100 (with 0 indicating the worst outcome to 100, the best)⁴⁰. It has been determined to be reliable and valid for use with general elbow pathology⁴¹, but the minimal clinically important difference has not been reported.

Broberg and Morrey Elbow Scale

The Broberg and Morrey scale was introduced in 1986⁴². This index includes a physician assessment of motion (flexion and forearm pronation and supination), strength, stability, and a patient assessment of pain. Pain carries the highest weight (40%) of the four variables. Pain has the highest influence (59%) on variability of the score³⁹. The scale ranges from 0 to 100, with a higher score indicating a better outcome. The raw score is assigned a categorical rank of poor (0 to 59), fair (60 to 79), good (80 to 94), and excellent (95 to 100). This scale has been validated for general elbow disorders and is more likely to correlate to other elbow measures if raw scores are reported rather than categorical rank⁴³. Its construct validity is intermediate for patient-rated severity and excellent for physician-rated severity⁴³. No minimal clinically important difference has been reported for this scale.

Mayo Clinic Performance Index for the Elbow

Morrey and Adams developed the Mayo Clinic performance index in 1992 to evaluate outcomes after total elbow arthroplasty⁴⁴. It consists of physician assessment of pain, arc of elbow motion, and stability, and a patient rating of daily function. Pain is weighted highest of the four variables (45%). Pain also has the highest influence (66%) on variability of the score³⁹. The scale ranges from 0 to 100, with a higher score indicating a better outcome. The raw score is assigned a categorical rank of poor (0 to 59), fair (60 to 74), good (75 to 89), and excellent (90 to 100). The score is more likely to correlate to other elbow measures if raw scores are reported rather than categorical rank⁴³. It has been validated for general elbow disorders⁴³. Its construct validity is good for patient-rated variables and excellent for physician-rated variables⁴³. A minimal clinically important difference of 15 was reported for patients with rheumatoid arthritis after arthroplasty or synovectomy⁴⁵.

Hospital for Special Surgery (HSS) Assessment Scale

The HSS assessment scale is a scoring system that was devised in 1980 to analyze the preoperative and postoperative data for patients having a total elbow arthroplasty⁴⁶. It assesses pain, duration of bending activities, activity level for the elbow, strength, and elbow motion. It is scored on a scale from 0 to 100, with a higher score indicating a better outcome. It has been validated for general elbow disorders⁴³. Construct validity is intermediate for patient-rated variables and intermediate for physician-rated variables⁴³. A minimal clinically important difference of 12 was reported for patients with rheumatoid arthritis after arthroplasty or synovectomy⁴⁵.

Liverpool Elbow Score

The Liverpool Elbow Score was developed in 2004 to measure outcomes for general elbow disorders⁴⁷. It contains two main components. The first is a six-item physician rating of elbow motion, strength, and ulnar nerve sensitivity. The second is a nine-question patient-rated questionnaire assessing pain, activities of daily living, and recreational activities. The variables are equally weighted. The questions are answered on a five-level Likert scale that is converted to a total score of 0 to 10 with use of a complex formula⁴⁷. Diagnosis does not appear to have an important influence on the results⁴⁷. A higher score indicates a better outcome. It has been validated for general elbow disorders⁴⁷. It is responsive to changes, but no minimal clinically important difference has been reported.

Overhead Athlete Measurements**Scoring System of Conway et al. (or the Conway-Jobe Score)**

The scoring system of Conway et al. was first reported in 1992 to assess overhead athletes' ability to return to their sport after an ulnar collateral ligament reconstruction⁴⁸. An excellent result indicates that the patient was able to compete at the same level or higher than that before the injury for more than twelve months. A good result indicates that the patient was able to compete at a lower level for more than twelve months or was able to throw in daily batting practice. A fair result indicates that the patient was able to return to a recreational level. A poor result indicates that the patient was unable to participate in his or her overhead sport. This measure is used most commonly to assess outcomes in throwing athletes. It has not been validated. No minimal clinically important difference has been reported.

Timmerman-Andrews Elbow Score

The Timmerman and Andrews score was developed in 1994 to evaluate posttraumatic elbow pain and stiffness⁴⁹. This outcome measure evaluates both subjective and objective criteria. The subjective categories include pain, locking, swelling, and activity level. The objective scale includes assessment for elbow flexion contracture, arc of forearm pronation and supination, and total arc of elbow motion. The subjective and objective scores are combined to give a range from 20 to 200. A higher score implies a better outcome. The raw scores are categorized as excellent (180 to 200), good (160 to 179), fair (120 to 159),

and poor (<120). It is most commonly used to assess outcomes after surgery in throwing athletes. This score has not been validated, and no minimal clinically important difference has been reported.

Hand and Wrist Outcome Measures**Boston Carpal Tunnel Questionnaire
(or the Levine and Katz Questionnaire)**

The Boston Carpal Tunnel Questionnaire (BCTQ) is a patient-reported assessment developed to evaluate the severity of symptoms and functional status associated with carpal tunnel syndrome⁵⁰. The questionnaire was designed to be relevant to a range of patients from young individuals who perform manual labor to the elderly and has been applied to patients treated operatively and nonoperatively. This measure consists of two subscales. The first assesses symptom severity through eleven questions querying pain, altered sensibility, and weakness. The second section evaluates functional status by assessing the patient's self-reported ability to perform eight tasks. All questions are scored from 1 to 5, with 1 indicating no difficulty or lack of symptoms, and 5 corresponding to an inability to complete tasks or to the most severe symptoms. The responses on each subscale are summed, and a mean value is reported for each subscale from 1 to 5^{51,52}. The questionnaire has internal consistency and sensitivity to clinical change⁵³. A potential limitation is weak correlation with physical examination measures, including two-point discrimination, Semmes-Weinstein monofilament testing, and median nerve sensory conduction velocity⁵⁰.

Function Test of Jebsen et al.**(or the Jebsen-Taylor Hand Function Test)**

The function test of Jebsen et al. was first introduced in 1969⁵⁴. The test evaluates performance on standardized tasks in an effort to quantify hand function. There are seven unilateral tasks that encompass a variety of hand functions including picking up small objects, stacking items, picking up heavy objects, and writing. This functional assessment has been estimated to take approximately fifteen minutes, requiring staff to administer and record results. Subjects completing this test can be evaluated against established norms stratified by age and sex. As a test of timed task completions, this instrument has been applied to a wide range of conditions beyond extremity trauma or peripheral nerve compression to include conditions such as stroke or central nervous system injury^{55,56}. Investigators have noted excellent correlation between results of this and other functional task tests but have questioned whether the ability of patients to perform functional tasks in this test predicts patient-rated outcomes assessing satisfaction and overall function^{57,58}.

Michigan Hand Outcomes Questionnaire

The Michigan Hand Outcomes Questionnaire (MHQ) was designed as a general survey of hand function and includes twenty-five items for the right and the left hand, as well as questions for both hands, designed for self-administration^{59,60}.

The MHQ provides data for overall hand performance as well as unique scores of separate domains related to hand function, daily activities, work performance, pain, aesthetics, and satisfaction. In contrast to the DASH, each hand is assessed independently. Furthermore, questions regarding function specifically refer to the impact of the involved hand and/or wrist on the activity as opposed to the DASH's more general queries of function, which presumably take all joints of the upper extremity into account. However, this level of detail makes the MHQ longer than the DASH, with the questions duplicated for right and left sides. Several publications have demonstrated the reliability, validity, and responsiveness of the MHQ^{57,59-63}. This questionnaire is scored from 0 to 100, with 100 being the optimal score. The minimal clinically important difference of the MHQ has ranged from 11 to 23 points for chronic, nontraumatic conditions⁶¹. The discriminative ability of the MHQ has been limited by a ceiling effect in traumatic conditions⁶¹.

Patient-Rated Wrist Evaluation

The Patient-Rated Wrist Evaluation (PRWE) consists of a fifteen-item questionnaire measuring wrist pain (five questions) and function (ten questions) during daily activities⁶⁴. It is easily completed in only a few minutes. Pain and function questions are answered from 0 (no pain) to 10 (the worst pain or greatest impairment). Each subscale is scored from 0 to 50 and can be summed to a total score on a scale of 0 to 100, with 0 indicating no disability (pain and function are equally weighted). Multiple publications support the PRWE's reliability, validity, responsiveness, and utilization in various populations (distal radial fractures, osteoarthritis, rheumatoid arthritis, scaphoid fractures, and Kienböck disease)^{10,64-67}. The PRWE stem questions have been modified to allow application for wrist and hand problems, although the content of the scored questions has not changed⁶⁸.

Discussion

This review of outcome measures in the upper extremity is by no means exhaustive but includes commonly used instruments for the shoulder, elbow, and wrist and hand. In general, these measures can be useful, but they vary widely in terms of what is known about their validity, reliability, minimal clinically important difference, and applicability to specific patient populations. We believe that it is important to understand how and why these instruments were developed, what they are validated to assess, and whether these measures are responsive to clinical changes so that these outcome measures are used appropriately in research and in clinical practice and the conclusions drawn from these data can be critically evaluated.

It is important to understand the potential for variation in patient-rated outcomes, particularly since rating of disability and outcomes is often affected by psychological and sociological factors that are not measured well. Assessment of general health with instruments like the SF-36 and the use of activity level measures may help to account for some of these factors but not all. In addition, caution should be used in interpreting

instruments that use physician-based rating scales to measure disability or outcomes as these may have inherent observer bias³³. Furthermore, it is inappropriate to combine physician and patient-based scoring in one measure³³. However, many of the commonly used outcome measures for the upper extremity include physician ratings and are included in this overview. We think that by combining general health measures, activity scales, and condition-specific measures, a rigorous outcome assessment can be performed.

When choosing among outcome measures for clinical research, there is rarely a "right answer" or singularly appropriate rating system. Orthopaedic surgeons must choose among available measures in an effort to balance the utility and burden of collecting and analyzing the desired data. In general, we believe that studies should include a measure of general health, an activity scale, and an instrument that has been validated to assess the outcome of the condition in question. However, responder burden may limit the quality of data being collected. If this is a concern, collecting a DASH score may be useful since the DASH has been shown to correlate with general health measures and joint-specific measures⁵⁶. Also, it is beneficial to use instruments that have an established minimal clinically important difference. Consideration should also be given to including general, joint-specific, measures like the ASES instrument for the shoulder or elbow since they are commonly used in the literature and can allow for comparisons with historic controls or across conditions. Lastly, researchers should consider their specific hypothesis when choosing an outcome instrument. For example, a study of the outcomes following distal radial fractures could utilize nearly all hand outcome measures. The QuickDASH or PRWE may be most suitable if minimizing patient response burden and maximizing ease of scoring are vital. The MHQ is the better choice for the researcher interested in quantifying differential impairment between the patient's uninjured and involved upper extremities. However, if the focus is to assess recovery of dexterity following a distal radial fracture, the Jebsen-Taylor test would be more appropriate.

The projections for future changes in health care suggest that quality outcomes and cost-effective treatments will be an increasing focus. Orthopaedic surgery has lagged behind other medical and surgical specialties in these analyses⁶⁹. Going forward, clinical research should focus on robust outcomes assessment to facilitate meaningful analysis that can demonstrate the value of orthopaedic interventions to patients and society at large.

In summary, there are a number of clinical outcome measures available for the upper extremity, many of which could benefit from further research into their validity, reliability, and optimal applicability. Nevertheless, appropriate outcome measures should be used when studying the treatment of musculoskeletal disorders in the upper extremity. By increasing their familiarity with these instruments, orthopaedic surgeons are better equipped to design studies of upper extremity disorders and to interpret and explain the results of such studies for patients. ■

Matthew V. Smith, MD
Robert H. Brophy, MD
Department of Orthopaedic Surgery,
Washington University School of Medicine,
14532 South Outer Forty Drive, Chesterfield, MO 63017.
E-mail address for M.V. Smith: smithmv@wudosis.wustl.edu

Ryan P. Calfee, MD
Department of Orthopaedic Surgery,
Washington University School of Medicine,
660 South Euclid Avenue, Campus Box 8233,
Saint Louis, MO 63110

Keith M. Baumgarten, MD
Department of Surgery,
Orthopedic Institute,
Sanford School of Medicine,
810 East 23rd Street, Sioux Falls, SD 57117

Rick W. Wright, MD
Department of Orthopaedic Surgery,
Washington University School of Medicine,
One Barnes Hospital Plaza,
Suite 11300 West Pavilion,
Saint Louis, MO 63110

References

- Swiontkowski MF, Buckwalter JA, Keller RB, Haralson R. The outcomes movement in orthopaedic surgery: where we are and where we should go. *J Bone Joint Surg Am.* 1999;81:732-40.
- Wright RW, Baumgarten KM. Shoulder outcomes measures. *J Am Acad Orthop Surg.* 2010;18:436-44.
- Wright RW. Knee injury outcomes measures. *J Am Acad Orthop Surg.* 2009;17:31-9.
- Hudak PL, Amadio PC, Bombardier C. Development of an upper extremity outcome measure: the DASH (disabilities of the arm, shoulder and hand) [corrected]. The Upper Extremity Collaborative Group (UECG). *Am J Ind Med.* 1996;29:602-8.
- Beaton DE, Katz JN, Fossel AH, Wright JG, Tarasuk V, Bombardier C. Measuring the whole or the parts? Validity, reliability, and responsiveness of the Disabilities of the Arm, Shoulder and Hand outcome measure in different regions of the upper extremity. *J Hand Ther.* 2001;14:128-46.
- SooHoo NF, McDonald AP, Seiler JG 3rd, McGillivray GR. Evaluation of the construct validity of the DASH questionnaire by correlation to the SF-36. *J Hand Surg Am.* 2002;27:537-41.
- Likert R. A technique for the measurement of attitudes. *Arch Psychol.* 1932;140:1-55.
- Guyatt GH, Townsend M, Berman LB, Keller JL. A comparison of Likert and visual analogue scales for measuring change in function. *J Chronic Dis.* 1987;40:1129-33.
- Hsu JE, Nacke E, Park MJ, Sennett BJ, Huffman GR. The Disabilities of the Arm, Shoulder, and Hand questionnaire in intercollegiate athletes: validity limited by ceiling effect. *J Shoulder Elbow Surg.* 2010;19:349-54.
- Schmitt JS, Di Fabio RP. Reliable change and minimum important difference (MID) proportions facilitated group responsiveness comparisons using individual threshold criteria. *J Clin Epidemiol.* 2004;57:1008-18.
- Beaton DE, Wright JG, Katz JN; Upper Extremity Collaborative Group. Development of the QuickDASH: comparison of three item-reduction approaches. *J Bone Joint Surg Am.* 2005;87:1038-46.
- Polson K, Reid D, McNair PJ, Larmer P. Responsiveness, minimal importance difference and minimal detectable change scores of the shortened disability arm shoulder hand (QuickDASH) questionnaire. *Man Ther.* 2010;15:404-7.
- Richards R, An KN, Bigliani L, Friedman J. A standardized method of shoulder function. *J Shoulder Elbow Surg.* 1994;3:347-52.
- Michener LA, McClure PW, Sennett BJ. American Shoulder and Elbow Surgeons Standardized Shoulder Assessment Form, patient self-report section: reliability, validity, and responsiveness. *J Shoulder Elbow Surg.* 2002;11:587-94.
- Tashjian RZ, Deloach J, Green A, Porucznik CA, Powell AP. Minimal clinically important differences in ASES and simple shoulder test scores after nonoperative treatment of rotator cuff disease. *J Bone Joint Surg Am.* 2010;92:296-303.
- Constant CR. Age related recovery of shoulder function after injury [Thesis]. Cork, Ireland: University College; 1986.
- Angst F, Pap G, Mannion AF, Herren DB, Aeschlimann A, Schwyzer HK, Simmen BR. 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.
- Othman A, Taylor G. Is the constant score reliable in assessing patients with frozen shoulder? 60 shoulders scored 3 years after manipulation under anaesthesia. *Acta Orthop Scand.* 2004;75:114-6.
- 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.
- Constant CR, Gerber C, Emery RJ, Sjøbjerg JO, Gohlke F, Boileau P. A review of the Constant score: modifications and guidelines for its use. *J Shoulder Elbow Surg.* 2008;17:355-61.
- Banks MJ, Crossman JE, Emery RJ. A standard method of shoulder strength measurement for the Constant score with a spring balance. *J Shoulder Elbow Surg.* 1998;7:116-21.
- Fialka C, Oberleitner G, Stampf P, Brannath W, Hexel M, Vécsei V. Modification of the Constant-Murley shoulder score-introduction of the individual relative Constant score Individual shoulder assessment. *Injury.* 2005;36:1159-65.
- Brophy RH, Beauvais RL, Jones EC, Cordasco FA, Marx RG. Measurement of shoulder activity level. *Clin Orthop Relat Res.* 2005;439:101-8.
- Brophy RH, Levy B, Chu S, Dahm DL, Sperling JW, Marx RG. Shoulder activity level varies by diagnosis. *Knee Surg Sports Traumatol Arthrosc.* 2009;17:1516-21.
- Williams GN, Gangel TJ, Arciero RA, Uhorchak JM, Taylor DC. Comparison of the Single Assessment Numeric Evaluation method and two shoulder rating scales. Outcomes measures after shoulder surgery. *Am J Sports Med.* 1999;27:214-21.
- Amstutz HC, Sew Hoy AL, Clarke IC. UCLA anatomic total shoulder arthroplasty. *Clin Orthop Relat Res.* 1981;155:7-20.
- Fealy S, Kingham TP, Altchek DW. Mini-open rotator cuff repair using a two-row fixation technique: outcomes analysis in patients with small, moderate, and large rotator cuff tears. *Arthroscopy.* 2002;18:665-70.
- Krepler P, Wanivenhaus AH, Wurnig C. Outcome assessment of hemiarthroplasty of the shoulder: a 5-year follow-up with 4 evaluation tools. *Acta Orthop.* 2006;77:778-84.
- O'Connor DA, Chipchase LS, Tomlinson J, Krishnan J. Arthroscopic subacromial decompression: responsiveness of disease-specific and health-related quality of life outcome measures. *Arthroscopy.* 1999;15:836-40.
- Wright RW, Heller MA, Quick DC, Buss DD. Arthroscopic decompression for impingement syndrome secondary to an unstable os acromiale. *Arthroscopy.* 2000;16:595-9.
- Romeo AA, Mazzocca A, Hang DW, Shott S, Bach BR Jr. Shoulder scoring scales for the evaluation of rotator cuff repair. *Clin Orthop Relat Res.* 2004;427:107-14.
- Lo IK, Griffin S, Kirkley A. The development of a disease-specific quality of life measurement tool for osteoarthritis of the shoulder: The Western Ontario Osteoarthritis of the Shoulder (WOOS) index. *Osteoarthritis Cartilage.* 2001;9:771-8.
- Kirkley A, Griffin S, Dainty K. Scoring systems for the functional assessment of the shoulder. *Arthroscopy.* 2003;19:1109-20.
- Hollinshead RM, Mohtadi NG, Vande Guchte RA, Wadey VM. Two 6-year follow-up studies of large and massive rotator cuff tears: comparison of outcome measures. *J Shoulder Elbow Surg.* 2000;9:373-81.
- Mohtadi NG, Hollinshead RM, Sasyniuk TM, Fletcher JA, Chan DS, Li FX. A randomized clinical trial comparing open to arthroscopic acromioplasty with mini-open rotator cuff repair for full-thickness rotator cuff tears: disease-specific quality of life outcome at an average 2-year follow-up. *Am J Sports Med.* 2008;36:1043-51.
- Kirkley A, Alvarez C, Griffin S. The development and evaluation of a disease-specific quality-of-life questionnaire for disorders of the rotator cuff: The Western Ontario Rotator Cuff Index. *Clin J Sport Med.* 2003;13:84-92.
- Kirkley A, Griffin S, McLintock H, Ng L. The development and evaluation of a disease-specific quality of life measurement tool for shoulder instability. The Western Ontario Shoulder Instability Index (WOSI). *Am J Sports Med.* 1998;26:764-72.
- King GJ, Richards RR, Zuckerman JD, Blasler R, Dillman C, Friedman RJ, Gartsman GM, Iannotti JP, Mumahan JP, Mow VC, Woo SL. A standardized method for assessment of elbow function. Research Committee, American Shoulder and Elbow Surgeons. *J Shoulder Elbow Surg.* 1999;8:351-4.
- Doornberg JN, Ring D, Fabian LM, Malhotra L, Zurakowski D, Jupiter JB. Pain dominates measurements of elbow function and health status. *J Bone Joint Surg Am.* 2005;87:1725-31.
- John M, Angst F, Awiszus F, King GJ, MacDermid JC, Simmen BR. The American Shoulder and Elbow Surgeons Elbow Questionnaire: cross-cultural adaptation into German and evaluation of its psychometric properties. *J Hand Ther.* 2010;23:301-13.

- 41.** MacDermid JC. Outcome evaluation in patients with elbow pathology: issues in instrument development and evaluation. *J Hand Ther.* 2001;14:105-14.
- 42.** Broberg MA, Morrey BF. Results of delayed excision of the radial head after fracture. *J Bone Joint Surg Am.* 1986;68:669-74.
- 43.** Turchin DC, Beaton DE, Richards RR. Validity of observer-based aggregate scoring systems as descriptors of elbow pain, function, and disability. *J Bone Joint Surg Am.* 1998;80:154-62.
- 44.** Morrey BF, Adams RA. Semiconstrained arthroplasty for the treatment of rheumatoid arthritis of the elbow. *J Bone Joint Surg Am.* 1992;74:479-90.
- 45.** de Boer YA, Hazes JM, Winia PC, Brand R, Rozing PM. Comparative responsiveness of four elbow scoring instruments in patients with rheumatoid arthritis. *J Rheumatol.* 2001;28:2616-23.
- 46.** Inglis AE, Pellicci PM. Total elbow replacement. *J Bone Joint Surg Am.* 1980;62:1252-8.
- 47.** Sathyamoorthy P, Kemp GJ, Rawal A, Rayner V, Frostick SP. Development and validation of an elbow score. *Rheumatology (Oxford).* 2004;43:1434-40.
- 48.** Conway JE, Jobe FW, Glousman RE, Pink M. Medial instability of the elbow in throwing athletes. Treatment by repair or reconstruction of the ulnar collateral ligament. *J Bone Joint Surg Am.* 1992;74:67-83.
- 49.** Timmerman LA, Andrews JR. Arthroscopic treatment of posttraumatic elbow pain and stiffness. *Am J Sports Med.* 1994;22:230-5.
- 50.** Levine DW, Simmons BP, Koris MJ, Daltroy LH, Hohl GG, Fossel AH, Katz JN. A self-administered questionnaire for the assessment of severity of symptoms and functional status in carpal tunnel syndrome. *J Bone Joint Surg Am.* 1993;75:1585-92.
- 51.** Guyette TM, Wilgis EF. Timing of improvement after carpal tunnel release. *J Surg Orthop Adv.* 2004;13:206-9.
- 52.** Burke FD, Wilgis EF, Dubin NH, Bradley MJ, Sinha S. Relationship between the duration and severity of symptoms and the outcome of carpal tunnel surgery. *J Hand Surg Am.* 2006;31:1478-82.
- 53.** Amadio PC, Silverstein MD, Ilstrup DM, Schleck CD, Jensen LM. Outcome assessment for carpal tunnel surgery: the relative responsiveness of generic, arthritis-specific, disease-specific, and physical examination measures. *J Hand Surg.* 1996;21:338-46.
- 54.** Jebsen RH, Taylor N, Trieschmann RB, Trotter MJ, Howard LA. An objective and standardized test of hand function. *Arch Phys Med Rehabil.* 1969;50:311-9.
- 55.** Beekhuizen KS. New perspectives on improving upper extremity function after spinal cord injury. *J Neurol Phys Ther.* 2005;29:157-62.
- 56.** Alon G, Sunnerhagen KS, Geurts AC, Ohry A. A home-based, self-administered stimulation program to improve selected hand functions of chronic stroke. *Neuro-Rehabilitation.* 2003;18:215-25.
- 57.** Davis Sears E, Chung KC. Validity and responsiveness of the Jebsen-Taylor Hand Function Test. *J Hand Surg Am.* 2010;35:30-7.
- 58.** Bovend'Eerdt TJ, Dawes H, Johansen-Berg H, Wade DT. Evaluation of the Modified Jebsen Test of Hand Function and the University of Maryland Arm Questionnaire for Stroke. *Clin Rehabil.* 2004;18:195-202.
- 59.** Chung KC, Pillsbury MS, Walters MR, Hayward RA. Reliability and validity testing of the Michigan Hand Outcomes Questionnaire. *J Hand Surg Am.* 1998;23:575-87.
- 60.** Chung KC, Hamill JB, Walters MR, Hayward RA. The Michigan Hand Outcomes Questionnaire (MHQ): assessment of responsiveness to clinical change. *Ann Plast Surg.* 1999;42:619-22.
- 61.** Shauver MJ, Chung KC. The minimal clinically important difference of the Michigan hand outcomes questionnaire. *J Hand Surg Am.* 2009;34:509-14.
- 62.** Waljee JF, Chung KC, Kim HM, Burns PB, Burke FD, Wilgis EF, Fox DA. Validity and responsiveness of the Michigan Hand Questionnaire in patients with rheumatoid arthritis: a multicenter, international study. *Arthritis Care Res (Hoboken).* 2010;62:1569-77.
- 63.** Kotsis SV, Chung KC. Responsiveness of the Michigan Hand Outcomes Questionnaire and the Disabilities of the Arm, Shoulder and Hand questionnaire in carpal tunnel surgery. *J Hand Surg Am.* 2005;30:81-6.
- 64.** MacDermid JC, Turgeon T, Richards RS, Beadle M, Roth JH. Patient rating of wrist pain and disability: a reliable and valid measurement tool. *J Orthop Trauma.* 1998;12:577-86.
- 65.** Murphy DM, Khoury JG, Imbriglia JE, Adams BD. Comparison of arthroplasty and arthrodesis for the rheumatoid wrist. *J Hand Surg Am.* 2003;28:570-6.
- 66.** Angst F, John M, Goldhahn J, Herren DB, Pap G, Aeschlimann A, Schwyzer HK, Simmen BR. Comprehensive assessment of clinical outcome and quality of life after resection interposition arthroplasty of the thumb saddle joint. *Arthritis Rheum.* 2005;53:205-13.
- 67.** De Smet L, Robijns F, Degreef I. Outcome of proximal row carpectomy. *Scand J Plast Reconstr Surg Hand Surg.* 2006;40:302-6.
- 68.** MacDermid JC, Tottenham V. Responsiveness of the disability of the arm, shoulder, and hand (DASH) and patient-rated wrist/hand evaluation (PRWHE) in evaluating change after hand therapy. *J Hand Ther.* 2004;17:18-23.
- 69.** Brauer CA, Neumann PJ, Rosen AB. Trends in cost effectiveness analyses in orthopaedic surgery. *Clin Orthop Relat Res.* 2007;457:42-8.