Upper extremity-specific measures of disability and outcomes in orthopaedic surgery

Matthew V. Smith
Washington University School of Medicine in St. Louis

Ryan P. Calfee
Washington University School of Medicine in St. Louis

Keith M. Baumgarten
Sanford School of Medicine

Robert H. Brophy
Washington University School of Medicine in St. Louis

Rick W. Wright
Washington University School of Medicine in St. Louis

Follow this and additional works at: https://digitalcommons.wustl.edu/open_access_pubs

Part of the Medicine and Health Sciences Commons

Recommended Citation

This Open Access Publication is brought to you for free and open access by Digital Commons@Becker. It has been accepted for inclusion in Open Access Publications by an authorized administrator of Digital Commons@Becker. For more information, please contact engeszer@wustl.edu.
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 surgery. 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 process.

In general, outcome measures should be relevant to patients, easy to use, reliable, valid, and responsive to clinical changes. 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 measure addresses, and the factors that influence their use.
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 1996. 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 limb. The DASH has been shown to be valid and responsive compared with other joint-specific measures of the upper extremity. The DASH has been shown to correlate to general health measures such as the Short Form-36 (SF-36).

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 daily living. 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 daily living.

**TABLE I Upper Extremity Outcome Measures for the Shoulder**

<table>
<thead>
<tr>
<th>Scale*</th>
<th>Anatomic Region</th>
<th>Measures</th>
<th>Validated</th>
<th>Clinician (no. of questions)</th>
<th>Target Population</th>
<th>Age (yr)</th>
<th>Disorders</th>
<th>MCID†</th>
</tr>
</thead>
<tbody>
<tr>
<td>DASH†</td>
<td>Shoulder, arm, and hand</td>
<td>Function (bilateral)</td>
<td>Yes</td>
<td>None</td>
<td>38</td>
<td>All upper extremity diagnoses</td>
<td>18-65</td>
<td>10-17</td>
</tr>
<tr>
<td>ASES‡</td>
<td>Shoulder</td>
<td>Pain, instability, activities of daily living</td>
<td>Yes</td>
<td>Rarely used</td>
<td>10</td>
<td>All shoulder diagnoses</td>
<td>20-81</td>
<td>6.4 (12-17 for RTC)</td>
</tr>
<tr>
<td>Constant§</td>
<td>Shoulder</td>
<td>Function</td>
<td>Yes</td>
<td>6 items</td>
<td>2</td>
<td>Rotator cuff repair, shoulder arthroplasty, adhesive capsulitis, and proximal humeral fractures</td>
<td>14-85</td>
<td>NA</td>
</tr>
<tr>
<td>Shoulder activity level¶</td>
<td>Shoulder</td>
<td>Activity level</td>
<td>Yes</td>
<td>None</td>
<td>7</td>
<td>Rotator cuff, osteoarthritis, and instability</td>
<td>NA</td>
<td></td>
</tr>
<tr>
<td>SANE∥</td>
<td>Shoulder</td>
<td>Patient perception</td>
<td>No</td>
<td>None</td>
<td>1</td>
<td>All shoulder diagnoses</td>
<td>NA</td>
<td></td>
</tr>
<tr>
<td>UCLA Shoulder Score¶¶</td>
<td>Shoulder</td>
<td>Pain, function, forward flexion, forward flexion strength, and patient satisfaction</td>
<td>No</td>
<td>2 items</td>
<td>3</td>
<td>All shoulder diagnoses</td>
<td>NA</td>
<td></td>
</tr>
<tr>
<td>WOOS¶¶¶</td>
<td>Shoulder</td>
<td>Pain, physical symptoms, sports, recreation, work, and emotional</td>
<td>Yes</td>
<td>None</td>
<td>19</td>
<td>Osteoarthritis</td>
<td>NA</td>
<td></td>
</tr>
<tr>
<td>RCQL¶¶¶¶</td>
<td>Shoulder</td>
<td>Symptoms, sports, work, lifestyle, and social and emotional</td>
<td>No</td>
<td>None</td>
<td>34</td>
<td>Rotator cuff pathology</td>
<td>NA</td>
<td></td>
</tr>
<tr>
<td>WORC¶¶¶¶¶</td>
<td>Shoulder</td>
<td>Pain, physical symptoms, sports, recreation, work, and emotional</td>
<td>Yes</td>
<td>None</td>
<td>20</td>
<td>Rotator cuff pathology</td>
<td>20-84</td>
<td>245</td>
</tr>
<tr>
<td>WOSI¶¶¶¶¶¶</td>
<td>Shoulder</td>
<td>Pain, physical symptoms, sports, recreation, work, and emotional</td>
<td>Yes</td>
<td>None</td>
<td>21</td>
<td>Shoulder instability</td>
<td>220</td>
<td></td>
</tr>
</tbody>
</table>

*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, RCQL = 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.
arts activities. Each individual item is scored on a 5-point Likert 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 disability. 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 population. 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) diagnoses.

In 2005, the QuickDASH was developed to minimize time and responder burden. 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 ($r > 0.97$). 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 points.

**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 diagnoses. It consists of a physician-rated and a patient-rated questionnaire. The physician-rated outcomes are not commonly reported. The patient-rated

---

### TABLE II Upper Extremity Outcome Measures for the Elbow

<table>
<thead>
<tr>
<th>Scale†</th>
<th>Anatomic Region</th>
<th>Measures</th>
<th>Validated</th>
<th>Responder Burden</th>
<th>Target Population</th>
</tr>
</thead>
<tbody>
<tr>
<td>DASH 1</td>
<td>Shoulder, arm, and hand</td>
<td>Function (bilateral)</td>
<td>Yes</td>
<td>None</td>
<td>38</td>
</tr>
<tr>
<td>ASES 38</td>
<td>Elbow</td>
<td>Pain, function, strength, range of motion, stability, and patient satisfaction</td>
<td>Yes</td>
<td>38 items</td>
<td>19</td>
</tr>
<tr>
<td>Broberg and Morrey 42</td>
<td>Elbow</td>
<td>Motion, strength, stability, and pain</td>
<td>Yes</td>
<td>5 items</td>
<td>1</td>
</tr>
<tr>
<td>Mayo Clinic performance index 45</td>
<td>Elbow</td>
<td>Pain, motion, stability and daily function</td>
<td>Yes</td>
<td>3 items</td>
<td>5</td>
</tr>
<tr>
<td>HSS assessment scale 46</td>
<td>Elbow</td>
<td>Pain, duration of bending activities, activity level for the elbow, strength, and range of motion</td>
<td>Yes</td>
<td>6 items</td>
<td>3</td>
</tr>
<tr>
<td>Liverpool Elbow Score 47</td>
<td>Elbow</td>
<td>Range of motion, strength, ulnar nerve sensitivity, pain, activities of daily living, and recreational activities</td>
<td>Yes</td>
<td>6 items</td>
<td>9</td>
</tr>
<tr>
<td>Conway-Jobe score 48</td>
<td>Elbow</td>
<td>Return to sport</td>
<td>No</td>
<td>None</td>
<td></td>
</tr>
<tr>
<td>Timmerman and Andrews 49</td>
<td>Elbow</td>
<td></td>
<td>No</td>
<td>3 items</td>
<td>4</td>
</tr>
</tbody>
</table>

*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.
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 age. An estimation of the minimal clinically important difference and the minimal detectable change was 6.4 ASES points and 9.7 ASES points, respectively. Another recent estimate of the minimal clinically important difference for nonoperative treatment of rotator cuff disease was between 12 and 17 ASES points.

The Constant Shoulder Score
The Constant shoulder score was originally developed as a university thesis in 1986. 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 arthroplasty, rotator cuff repair, adhesive capsulitis of the shoulder, and proximal humeral fractures. 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 measures. 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 function. 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 instability. 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)?” 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 measures\(^{26}\). It has been used in evaluating essentially all shoulder conditions\(^{27-31}\). 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 100, with a higher score indicating a worse outcome. Its developers determined that it was valid, reliable, and responsive in patients with osteoarthritis treated with arthroplasty\(^{34}\). 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 tears\(^{35}\). 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 a higher score indicating a worse outcome. It has been used on patients from twenty-five to eighty-three years old to assess all forms of rotator cuff pathology\(^{36}\). 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 disease\(^{37}\). 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 function\(^{38}\). The questionnaire 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%)\(^{39}\).

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 1999\(^{40}\) 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 score\(^{41}\). The original description of the measurement did not include information pertaining to global scoring of the assessment\(^{42}\). 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)\(^{43}\). It has been determined to be reliable and valid for use with general elbow pathology\(^{44}\), but the minimal clinically important difference has not been reported.

Broberg and Morrey Elbow Scale
The Broberg and Morrey scale was introduced in 1986\(^{45}\). 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 score\(^{46}\). 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 rank\(^{47}\). Its construct validity is intermediate for patient-rated severity and excellent for physician-rated severity\(^{48}\). 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 arthroplasty\(^{49}\). 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 score\(^{50}\). 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 rank\(^{51}\). It has been validated for general elbow disorders\(^{52}\). Its construct validity is good for patient-rated variables and excellent for physician-rated variables\(^{53}\). A minimal clinically important difference of 15 was reported for patients with rheumatoid arthritis after arthroplasty or synovectomy\(^{54}\).
evaluate posttraumatic elbow pain and stiffness

The Timmerman and Andrews score was developed in 1994 to evaluate posttraumatic elbow pain and stiffness$^{46}$. 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 disorders$^{46}$. Construct validity is intermediate for patient-rated variables and intermediate for physician-rated variables$^{46}$. A minimal clinically important difference of 12 was reported for patients with rheumatoid arthritis after arthroplasty or synovectomy$^{46}$.

Liverpool Elbow Score

The Liverpool Elbow Score was developed in 2004 to measure outcomes for general elbow disorders$^{46}$. 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 formula$^{47}$. Diagnosis does not appear to have an important influence on the results$^{47}$. A higher score indicates a better outcome. It has been validated for general elbow disorders$^{46}$. 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 reconstruction$^{48}$. 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 stiffness$^{46}$. 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 syndrome$^{50}$. 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 change$^{51}$. A potential limitation is weak correlation with physical examination measures, including two-point discrimination, Semmes-Weinstein monofilament testing, and median nerve sensory conduction velocity$^{50}$.

Function Test of Jebsen et al. (or the Jebsen-Taylor Hand Function Test) The function test of Jebsen et al. was first introduced in 1969$^{51}$. 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$^{53,54}$. 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$^{55,56}$.

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$^{57,58}$.
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. 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 conditions. The discriminative ability of the MHQ has been limited by a ceiling effect in traumatic conditions.

**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 activities. 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). The PRWE stem questions have been modified to allow application for wrist and hand problems, although the content of the scored questions has not changed.

**Discussion**

This review of outcome measures in the upper extremity is by no means exhaustive but includes commonly used instruments for the shoulder, elbow, 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 bias. Furthermore, it is inappropriate to combine physician and patient-based scoring in one measure. 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 measures. 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 Jepsen–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 analyses. 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.
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 liga-
53. Amadio PC, Silverstein MD, Istrup DM, Schleck CD, Jensen LM. Outcome as-
56. Alon G, Sunnerhagen KS, Geurts AC, Ohy A. A home-based, self-administered stimulation program to improve selected hand functions of chronic stroke. Neuro-
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 Ques-