Outcomes Measurement of Upper Extremity Function

Jennifer Stiller, ATC, and Timothy L. Uhl, PhD, ATC, PT • University of Kentucky

A SIMPLE AND COMMONLY used method of objectively assessing shoulder function is a patient self-report questionnaire. There are several instruments available. This column reviews four functional questionnaires in an effort to identify valid and responsive measures for evaluating upper extremity function: the Constant-Murley Shoulder Score (CMS), the American Shoulder and Elbow Surgeons (ASES) Self-Report Form, the University of Pennsylvania Shoulder Score (U-Penn), and the Disabilities of the Arm, Shoulder, and Hand (DASH) outcome measure. For scientific and clinical information obtained from these tools to be of merit, psychometric properties (reliability, validity, and responsiveness) of available assessment tools must be demonstrated.

Reliability implies that test results will be consistent over time, and test–retest reliability refers to the stability of the scale during repeated measures.[1] Reliability can be measured using an intraclass correlation coefficient (ICC). Coefficients range from 0 to 1.0, with values closer to 1.0 indicating a substantial relationship or correlation.[1] A coefficient of .75 is judged to be acceptable.

Construct validity is the validity of the instrument in a specific test situation or to a theoretical concept.[2,3] To determine validity, the internal consistency of test items can be demonstrated through the use of a Cronbach's alpha coefficient.[1,4] This statistic calculates the degree of homogeneity or similarity of the items that make up the total score, with values closer to 1.0 indicating a more significant relationship.

Responsiveness refers to the ability of a measure to detect change when it has occurred and to measure clinically important change.[1,2,5] Responsiveness is primarily expressed with two statistics: standardized response mean and effect size. Furthermore, the standard error of the measure can be used to calculate the minimal detectable change of a measure, which is the minimum number of points by which a patient's score must change in order for the clinician to be 90 % confident that a true change has occurred. Some scores might also report a minimal clinically important difference, which is the amount of change needed to be clinically meaningful to a patient but that is not yet statistically meaningful.

Constant–Murley Shoulder Score

The CMS is a 100-point functional shoulder-assessment tool in which higher scores reflect increased function.[3] It combines four separate subscales: subjective pain (15 points), function (20 points), objective clinician assessment of range of motion (40 points), and strength (25 points).[3,5,6] The CMS system is used internationally as a means of establishing normal levels of shoulder function appropriate for different age groups and to establish what constitutes disability in normal individuals.[3] It has also been used to establish differential rates of progress after injury or treatment.[3] Reliability has been reported, but validity has been questioned based on three concerns: (a) A single pain scale is considered inadequate to gain a true picture of the patient's pain, (b) the report of function is not specific to any particular activity and is therefore left to interpretation by the patient, and (c) the method of measuring strength has not been standardized.[6] A final weakness of this system is that it requires a large amount of objective data collection by the clinician, thus affecting interrater

 - ATT 10(3), pp. 24-25

reliability. Many patients will likely be lost to follow-up, which can lead to incomplete outcome studies.[6] Limited research has been performed on this instrument to document its scientific merit (Table 1).

The ASES developed a 100-point standardized shoulder-assessment self-report form, 50 points of which are derived from patient self-report of pain on a visual analog scale and 50 points of which are computed from a formula using the cumulative score of 10 activities of daily living derived using a four-point ordinal scale.[5,6] Self-assessment questions deal with categories of pain, instability, and medication use. Activities of daily living include such skills as putting on a coat, sleeping on the affected side, and combing one's hair. An optional objective component of this scoring system takes into account select range-of-motion and manual muscle test scores.[4] The ASES focuses on aspects of pain and function (higher scores reflect increased function), it can be administered in under 5 min, and it has the advantage of a 100-point functional score that can be completed by the patient independent of the examiner.[6] The ASES has been demonstrated to be reliable, valid, and responsive (Table 1).

University of Pennsylvania Shoulder Score

The U-Penn consists of two separate 100-point sections, one being a subjective scale and the other an objective assessment.[6] The subjective scale is an assessment of the patient's pain, satisfaction, and function. Higher scores on each scale indicate increased function. Pain is assessed in each of three conditions: with the arm at rest by the side, with normal activities, and with strenuous activities. All subjective statements are based on a 10-point scale, with endpoints of *no pain* and *worst possible pain* or *very satisfied* and *not satisfied*. Self-assessed function is based on a 20-item questionnaire with a four-category Likert scale.[6] The objective evaluation assesses the patient's range of motion and strength, with scoring based on a percentage difference as compared with the opposite, uninvolved side.[6] Statistical analysis of this scoring system reports high reliability and good correlation of total scores to CMS and ASES shoulder scores[6] (Table 1).

Disabilities of the Arm, Shoulder, and Hand

The DASH outcomes measure was developed to evaluate symptoms and upper extremity functional status and to determine the relative impact of disorders.[2,10] The DASH is a 30-item questionnaire with a five-item response option for each item. The test has a maximum score of 100, where higher scores reflect greater disability. It can be used as either a one-time measure or to determine change over time.[2] Discriminative validity has also been determined, indicating that patients who were currently able to work with their condition or who were able to complete activities of daily living to their satisfaction recorded statistically significant differences in DASH scores versus those who were unable to work or complete daily functions. The DASH has been demonstrated to be a valid and reliable tool for both proximal and distal disorders of the upper extremity, therefore confirming its usefulness for multiple joints of the entire upper extremity[2] (Table 1).

Selecting the most appropriate outcomes measure is often a very complex issue in the design of a study. The right measure is critical for success, because it can influence study cost, sample size, time, and burdens placed on subjects, as well as the likelihood that the study can obtain clinically important results.[4] One com-

TABLE 1. ■

Outcome Measure	Reliability (ICC)	Internal Consistency (Cronbach's alpha)	Responsiveness SRM (ES)	SEM (points)	MDC (points)
CMS[1,5,6]	.80, .87	not tested	0.59	not reported	not reported
ASES[1,4,8]	.84, .96	.86	0.93, 1.5 (1.4)	6.7	9.7, 15.5; MCID: 6.4
U-Penn[1,9]	.94	.93	1.27 (1.01)	8.6	12.1
DASH[1,2,10]	.96	.90[1]	1.2 (0.7)	4.6, 7.1	12.75, 12.8

Note. SRM = standardized response mean; ES = effect size; SEM = standard error of measure; MDC = minimal detectable change; CMS = Constant-Murley Shoulder score; ASES = American Shoulder and Elbow Surgeons self-report; MCID = minimal clinically important difference; U-Penn = University of Pennsylvania Shoulder Score; DASH = Disabilities of the Arm, Shoulder, and Hand.

ponent of this decision should be that the instrument is of scientific merit with sound psychometric properties. Another important component in the decision process is the patient population to be studied. The assessment tools described here have been primarily used for general patient populations. Clinicians should be aware of reported reliability, internal consistency, responsiveness, and minimal detectable changes of the self-report measures they are considering, because these might influence which measure is selected.

References

1. Michener L, Leggin B. A review of self-report scales for the assessment of functional limitation and disability of the shoulder. *J Hand Ther*. 2001;14:68-76.
2. Beaton D, Katz J, Fossel A, Wright J, Tarasuk V, Bombardier C. Measuring the whole or the parts? validity, reliability, and responsiveness of the Disabilities of the Arm, Shoulder, and Hand outcome measures in different regions of the upper extremity. *J Hand Ther*. 2001;14: 128-142.
3. Constant C, Murley A. A clinical method of functional assessment of the shoulder. *Clin Orthop Rel Res*. 1985;214:160-164.
4. Michener L, McClure P, Sennett B. American shoulder and elbow surgeons standardized shoulder assessment form, patient self-report section: reliability, validity, and responsiveness. *J Shoulder Elbow Surg*. 2002;11:587-594.
5. Cook K, Roddey T, Olson S, Gartsman G, Valenzuela F, Hanten W. Reliability by surgical status of self-reported outcomes in patients who have shoulder pathologies. *J Orthop Sports Phys Ther*. 2002;32: 336-343.
6. Leggin B, Iannotti J. Shoulder outcome measurement. In: Iannotti JP, Williams GR. *Disorders of the Shoulder: Diagnosis and Management*. Philadelphia, Pa: Lippincott Williams & Wilkins; 1999:1024-1040.
7. Beaton D, Richards R. Assessing the reliability and responsiveness of 5 shoulder questionnaires. *J Shoulder Elbow Surg*. 1998;7:565-572.
8. McClure P, Bialker J, Neff N, Williams G, Karduna A. Shoulder functional and 3-dimensional kinematics in people with shoulder impingement syndrome before and after a 6-week exercise program. *Phys Ther*. 2004;84:832-848.
9. Gummesson C, Atroshi I, Ekdahl C. The Disabilities of the Arm, Shoulder, and Hand (DASH) outcome questionnaire: longitudinal construct validity and measuring self-rated health change after surgery. *BMC Musculoskeletal Disord*. 2003;4(1):11.
10. Gay R, Amadio P, Johnson J. Comparative responsiveness of the disabilities of the arm, shoulder, and hand, the carpal tunnel questionnaire, and the SF-36 to clinical change after carpal tunnel release. *J Hand Surg*. 2003;28:250-254.

Jennifer Stiller is a second-year master's student in the Division of Athletic Training at the University of Kentucky.

Tim Uhl is an assistant professor in the Division of Athletic Training and director of the Musculoskeletal Laboratory in the Department of Rehabilitation Sciences, College of Health Sciences, University of Kentucky.