

# UCSF

## UC San Francisco Previously Published Works

### Title

Measurement Scales in Clinical Research of the Upper Extremity, Part 2: Outcome Measures in Studies of the Hand/Wrist and Shoulder/Elbow

### Permalink

<https://escholarship.org/uc/item/96w715cx>

### Journal

The Journal Of Hand Surgery, 38(2)

### ISSN

0363-5023

### Authors

Badalamente, Marie  
Coffelt, Laureen  
Elfar, John  
[et al.](#)

### Publication Date

2013-02-01

### DOI

10.1016/j.jhsa.2012.11.029

Peer reviewed



Published in final edited form as:

*J Hand Surg Am.* 2013 February ; 38(2): 407–412. doi:10.1016/j.jhsa.2012.11.029.

## Measurement Scales in Clinical Research of the Upper Extremity, Part 2: Outcome Measures in Studies of the Hand/Wrist and Shoulder/Elbow

Marie Badalamente, PhD, Lauren Coffelt, BScOT, John Elfar, MD, Glenn Gaston, MD, Warren Hammert, MD, Jerry Huang, MD, Lisa Lattanza, MD, Joy Macdermid, PhD, Greg Merrell, MD, David Netscher, MD, Zubin Panthaki, MD, Greg Rafijah, MD, Douglas Trczinski, MD, and Brent Graham, MD [Chair] for the American Society for Surgery of the Hand Clinical Trials and Outcomes Committee

American Society for Surgery of the Hand, Chicago, IL

### Abstract

Part 1 of this article outlined the basic characteristics of useful clinical measurement instruments and described scales used to measure general health, pain, and patient satisfaction. Part 2 describes the features of some of the scales most commonly used in clinical research in the hand, wrist, elbow, and shoulder.

### Keywords

Measurement; outcome instruments; upper extremity

## HAND AND WRIST

The scales that dominate the measurement of outcomes related to the hand and wrist are the Disabilities of the Arm, Shoulder, and Hand (DASH)<sup>1</sup> and the Michigan Hand Outcomes Questionnaire (MHQ).<sup>2</sup> The Upper Extremity Function Scale<sup>3</sup> is also a region-specific instrument. The Patient-Related Wrist Evaluation (PRWE)<sup>4</sup> focuses solely on wrist function. A number of disease-specific scales are relevant to the hand, including the Boston Carpal Tunnel Questionnaire,<sup>5</sup> the Health Assessment Questionnaire (HAQ)<sup>6</sup> and the Arthritis Impact Measurement Scale (AIMS),<sup>7</sup> the latter 2 of which measure outcomes related to rheumatoid arthritis,<sup>8</sup> and the Australian/Canadian Hand Osteoarthritis Index<sup>9</sup> which focuses on osteoarthritis.

### Disabilities of the Arm, Shoulder, and Hand, and Hand/Upper Extremity Function Scale

The performance characteristics of the DASH were reviewed in Part 1. Briefly restated, the scale is reliable, valid for a wide variety of conditions and treatments affecting the hand and wrist, and has a known minimally clinically importance difference. The responsiveness of

---

Corresponding author: Brent Graham, MD, Toronto Western Hospital, 399 Bathurst Street, East Wing 2-425, Toronto, Ontario M5T 2S8; Brent.Graham@uhn.on.ca.

No benefits in any form have been received or will be received related directly or indirectly to the subject of this article.

the instrument varies with the condition under evaluation, and this is related both to its focus on disability and the fact that the hand and wrist are components of the entire upper extremity.

The distinction between the constructs of “disability” and “function” is important. *Disability* is reported by patients and can vary substantially for a given amount of clinician-measured impairment in domains such as strength or movement. There has been a substantial and growing body of evidence linking patient-perceived disability to psychosocial factors such as catastrophization. As a result, disability might correlate poorly with impairment and with the idea of function. *Function* is a general term that implicitly includes dimensions of movement, strength, joint stability, pain, and, in the hand, even cosmetic appearance. The multifaceted nature of the term *function* is not usually acknowledged, and because most of these constructs are not actually measured when function is used a result, the report of function as a concept actually has limited usefulness. Disability and function might be related, but they are by no means equated with one another.

The DASH will be responsive when the changes being measured are due to changes in patient-reported disability and when these changes are reflected in the upper extremity as a whole. Conditions that are not, in general, associated with much disability for most patients or that have an impact on only a small aspect of overall upper extremity health should not be evaluated using the DASH because it is unlikely to be sufficiently responsive.

A shorter form of the DASH, the *QuickDASH* has also been developed and found to have the same psychometric characteristics as the longer instrument.<sup>10</sup> This scale might be more appropriate in some clinical research settings.

### **The Michigan Hand Outcomes Questionnaire**

The MHQ is in wide use in the evaluation of outcomes related to the hand. It seeks to measure 6 domains of hand function: overall hand function, activities of daily living, pain, work performance, hand appearance, and patient satisfaction. It represents a status instrument that can be used to evaluate the condition of the hand and its related functions and the impact of treatment. The separation of items into a variety of domains is a particularly useful feature that improves the sensibility of the instrument. The MHQ has excellent reliability and has been shown to be valid and responsive<sup>11</sup> for a wide spectrum of conditions affecting the hand, including rheumatoid arthritis and other inflammatory arthropathies, and also for hand injuries. The internal consistency of the instrument is high, meaning that multiple, closely related facets of hand performance are being measured. This probably accounts for its reliability. The minimally clinically important difference has also been established.<sup>12</sup> The instrument has also been translated into a number of languages.

The inclusion in the MHQ of a scale that measures the aesthetic status of the hand is a unique and interesting component because this is often overlooked in patients with injuries and diseases in the hand, but this domain has clear importance. This substantially broadens the scope of the MHQ and increases its appeal and usefulness as a summation of hand status.

The satisfaction scale of the MHQ involves 3 questions that relate to satisfaction with overall functioning of the hand. There have been no studies specifically testing the validity of this set of questions, but it is likely that, as an aggregate, they are more useful than a single item simply asking about satisfaction in general. An abbreviated version of the MHQ, the Brief MHQ has been developed primarily for clinical use.<sup>13</sup> The full instrument is recommended for research purposes.

### **Patient-Related Wrist Evaluation or Patient-Rated Wrist/Hand Evaluation**

The PRWE is a 15-item scale that has been used in more than 70 published wrist/hand studies. The items are scored on a 0 to 10 numeric rating scale and include 3 subscales: a 5-item pain scale, 6 specific functional tasks, and 4 items that address usual activity/role in the areas of self care, household work, occupation, and recreation. The Patient-Rated Wrist/Hand Evaluation (PRWHE) is identical to the PRWE except that “wrist” is replaced with “wrist/hand.” The PRWHE also includes a question relating to hand appearance.<sup>14</sup> The reliability of both of these scales has been established for the full instrument, as well as for the individual subscales.<sup>15</sup> This is an important feature because it suggests that the individual subscales can provide useful and reproducible data if they are used outside the full instrument. The scale has been validated for use in a number of languages.

Reports of the minimally clinically important difference have varied somewhat but appear to be clustered around a difference of approximately 10 points. The responsiveness of both the PRWE and PRWHE is similar to that of the MHQ.

### **Upper Extremity Function Scale**

The Upper Extremity Function Scale also seeks to evaluate integrated upper extremity function.<sup>3</sup> It is a brief, 8-item scale that has been shown to be valid for the evaluation of carpal tunnel syndrome. The scale has not been shown to be well correlated with impairment measures in the hand, suggesting that, like the DASH, it measures a construct that reflects some integrated facet of function that is only partly related to physician-measured impairment.

### **Boston Carpal Tunnel Syndrome Scale**

The Boston Carpal Tunnel Syndrome Scale represents a status instrument for a single condition, carpal tunnel syndrome, although it could conceivably be used in the same capacity for the evaluation of a similar condition such as cubital tunnel syndrome. The instrument comprises 2 subscales, one that evaluates symptoms and one that assesses function. Both subscales comprise a series of questions scored on a Likert scale. Although this instrument is not intended to be a diagnostic instrument, it is often used in this way, presumably on the assumption that a low score indicates little or no symptoms or functional impairment attributable to carpal tunnel syndrome. Conversely, higher scores indicate symptoms and functional limitations related to this diagnosis. The main use of the instrument is in measuring change in status, usually as a result of treatment. The scale is reliable, valid, and responsive.<sup>16</sup> The minimally clinically important difference has been established.<sup>17</sup>

### **Arthritis Impact Measurement Scale; Health Assessment Questionnaire**

Both the AIMS and the HAQ were developed as measures of disease activity in rheumatoid arthritis. The HAQ has also been used as a measure of disease impact in osteoarthritis. Both evaluate the overall impact of the condition and do not focus on hand function specifically. They are patient-reported measures that reflect disability more than impairment. There are abbreviated versions of both instruments that have been shown to be valid. The use of the AIMS or HAQ in clinical research related to the upper extremity should probably be limited to a complementary role reflecting overall disease impact when the topic is rheumatoid arthritis. In many instances, general health is better measured using a generic health measure such as the SF-36.

### **Australian/Canadian Hand Osteoarthritis Index**

The Australian/Canadian Hand Osteoarthritis Index is similar to the Boston Carpal Tunnel Syndrome Scale in its focus on a single disease, osteoarthritis. The focus on hand symptoms and function makes it an attractive alternative to the AIMS or HAQ. It comprises 15 scored questions that measure symptoms and function. It correlates well with impairment measures in the hand, such as grip and pinch strength, and has been found to be both valid and reliable.<sup>18</sup> The minimally clinically important difference has been established.<sup>19</sup>

## **SHOULDER**

Roy et al<sup>20</sup> have published an extensive review of the 4 main outcome measures used to measure shoulder function, the DASH, the Shoulder Pain and Disability Index (SPADI), the American Shoulder and Elbow Surgeons score (ASES) and the Simple Shoulder Test (SST). They noted that these 4 instruments were the most extensively studied and, as a result, had the most clearly understood psychometric properties among the large number of instruments that have been developed for the study of shoulder function.

Overall, all 4 instruments have similar performance characteristics in the main areas of validity, reliability, and responsiveness. They are approximately equivalent with respect to the burden placed on patients to complete the instrument, requiring 5 or fewer minutes for each scale. They are also highly correlated with one another because they more or less measure the same, or closely related, constructs. In some measure, this might make the choice of instrument somewhat easier because their overall performance is similar. The best choice for a specific study might be the instrument that other studies in the field have used, for example, for the evaluation of a particular intervention. Many papers report the outcomes on more than one of these scales. The usefulness of this approach is dubious, but there is a burden placed on patients in having them complete more than one scale, especially when the additional information relating to outcome is likely to be minimal. All 4 of instruments have been found to have relatively low responsiveness in studies of patients with shoulder instability.<sup>21,22</sup> The minimally clinically important difference has been established for the DASH, SPADI, and ASES.<sup>20</sup>

## Disabilities of the Arm, Shoulder, and Hand

The DASH has been found to have excellent reliability in multiple studies of a variety of shoulder conditions. Reliability has been found to exceed that of the SPADI and ASES in the assessment of shoulder arthroplasty results.<sup>23</sup> The DASH has also been shown to have better reliability than SPADI for the assessment of conditions such as rotator cuff impingement. The DASH has also been found to have excellent construct validity for shoulder-related disability. Although responsiveness of the DASH is satisfactory, it is somewhat less than that of the other 3 instruments commonly used to assess the shoulder. As is true for the assessment of outcomes in the hand and wrist, this might be due to its measurement of a less sharply focused construct, disability, as well as its evaluation of the entire extremity rather than on the shoulder alone. Floor and ceiling effects are less with the DASH than with generic health measures such as the SF-36, consistent with its greater responsiveness to shoulder problems.

### Simple Shoulder Test

The SST is a 12-item patient report that has been shown to cover 2 constructs, functional capacity of the shoulder and shoulder comfort at rest.<sup>24</sup> The reliability of the SST has been found to be excellent and has marginally exceeded the other instruments in this property in most studies. This might be related to the fact that all the items are answered with a yes/no dichotomous response. Like the other instruments, the SST has been shown to be valid and strongly correlated not only to the other heavily used scales but as also to a variety of other, less commonly used shoulder scales.<sup>20</sup> The responsiveness of the SST has been less studied than the other instruments have. When it has been studied, its best responsiveness has been in the evaluation of rotator cuff surgery outcomes, in which it outperformed the DASH for this particular parameter.<sup>25</sup> It has also been found to be responsive in evaluating outcomes after shoulder arthroplasty.<sup>20</sup>

### Shoulder Pain and Disability Index

The SPADI is a 13-item patient report of shoulder pain and disability. The responses are made on a visual analog scale, combined, and then scaled to a 100 point sum, where 0 is the best score possible. Like the other instruments, the SPADI has been shown to have excellent reliability and construct validity.<sup>26</sup> The SPADI has been shown to have its best responsiveness in the evaluation outcomes of rotator cuff surgery. It is less responsive than both the ASES and the DASH in the evaluation of results after shoulder arthroplasty.

### American Shoulder and Elbow Surgeons Score

The main constructs evaluated by the ASES are pain, which is assessed on a visual analog scale, and function, which is surveyed on 10 items, each of which are responded to on a 4-point ordinal scale. The ASES also has been shown to have excellent reliability in a number of studies.<sup>27,28</sup> Highly correlated as it is with the other instruments, ASES also has excellent validity. The responsiveness of the ASES has been found to exceed that of the other scales in the evaluation of outcomes from shoulder arthroplasty.<sup>23</sup>

In addition to these 4 instruments the Constant-Murley score is also in wide use in the evaluation of outcomes from the treatment of shoulder conditions.<sup>29</sup> It combines elements of

patient reports, such as conduct of activities of daily living and pain, as well as impairment measures, such as strength and range of motion. Although it has not been as fully evaluated as the other 4 commonly used measures, it has been validated for a wide spectrum of conditions and treatments, including shoulder arthroplasty, rotator cuff repair, and fractures of the proximal humerus. The scale has good reliability and is strongly correlated with other instruments, including the 4 most heavily studied scales. The Constant-Murley score has also been found to be responsive, especially for the evaluation of results of rotator cuff surgery and shoulder arthroplasty. The minimal clinically important difference has not been established. The combination of both patient report elements and impairment measures might make an overall summation of results difficult to evaluate if this is the only outcome measured; however, in instances in which this is the goal, the Constant-Murley score may be especially useful.

Given that all the measures are relatively highly correlated, the choice of one instrument over another should be dictated by the specific needs of the research question.

## **ELBOW**

### **Mayo Elbow Performance Score**

The Mayo Performance Elbow Score (MEPS) was developed from the anecdotal experience of one surgeon.<sup>30</sup> The MEPS evaluates 4 domains: pain, range of motion, stability, and function. The function component evaluates activities of daily living that involve elbow function. The successful completion of these activities of daily living requires a combination of motion, strength, and stability and, as such, represents an internal validity test of the constructs directly aimed at measuring these constituent domains. The scoring system involves an ad hoc weighting of the various components so that a maximum score of 100 points is possible. The scores have been assigned arbitrary categorical rankings of excellent, good, fair, and poor.

Although it was not created using the standard principles of measurement instrument development, the MEPS has become one of the most widely used scales for assessing outcomes related to the elbow in the literature. De Boer<sup>31</sup> reported that the intra-rater reliability of the MEPS was good to excellent. However, the MEPS suffered from poor internal consistency and poor inter-rater agreement, especially for the physician-reported subscale of stability. De Boer noted that the poor internal agreement of the MEPS was the result of its nature as a clinical outcome tool that seeks to produce the greatest predictive power. Therefore, each item of the scale adds a distinctive contribution to the assessment that is not redundant; hence, its relatively poor internal consistency. This is probably the main factor contributing to its good intra-rater reliability but poor inter-rater reliability.

Several studies have attempted to evaluate the validity of the MEPS, the most comprehensive of which was reported by Turchin.<sup>32</sup> This study compared the validity of 5 elbow scoring systems, including the MEPS. The main conclusion was that the raw scores of the various measures were relatively well correlated; however, the categorical rankings were poorly correlated because the operational definitions for these were arbitrarily set. This study noted that the MEPS had good sensibility and did not require special equipment. The MEPS was

found to be highly discriminating for patient-rated pain severity and demonstrated excellent discriminant validity on the physician-rated variables. The MEPS is valid for the assessment of outcomes for treatment of the elbow, but raw scores rather than the clinical categorizations of excellent, good, fair, and poor are recommended.

In contrast, De Boer<sup>31</sup> demonstrated moderate correlation with radiographic findings and no statistical correlation to isolated measures of impairment, including range of motion. Range of motion was found to be a poor indicator of elbow function in general. The MEPS also demonstrated poor validity when used to assess the nondominant elbow. In many instances, a comparison to the opposite elbow or to radiographs would be important, and thus, for these uses, the MEPS has impaired criterion validity. The MEPS has also been found to be less responsive than other similarly designed outcome measures, such as the Elbow Function Assessment (EFA).

Floor and ceiling effects have not been identified, suggesting that the MEPS represents a good overall assessment of the elbow across many different clinical settings. As a clinically derived scale, it also has good sensibility. Several studies have suggested that pain is the strongest predictor of physician-related or patient-related scores. Objective factors are more poorly correlated with overall outcome than is pain. The evaluation of the MEPS by Doornberg<sup>33</sup> found that pain alone accounted for up to 66% of the variance in the final score. They concluded that all the scales measure function consistently, but the dominant factor that is being measured and weighted is pain.

### **Oxford Elbow Score**

The Oxford Elbow Score (OES) is a 12-item, patient-centered outcome questionnaire.<sup>34</sup> The OES has demonstrated excellent internal consistency and reliability. In addition, the scale correlates well with the MEPS, DASH, and SF-36, suggesting satisfactory overall convergent validity. The responsiveness of the OES is comparable to that of the DASH and the SF-36.<sup>35</sup> Most of the questions on the OES relate to disability in upper extremity function secondary to impairment of elbow function, so it would be expected that the performance of the OES would parallel that of the DASH. The OES should be used in the same way as the DASH, in the measurement of self-reported disability, and might be more useful in isolating overall upper extremity disability to the elbow specifically.

### **Liverpool Elbow Score**

The Liverpool Elbow Score (LES) has 2 major components, a physician-derived clinical assessment and patient-centered questions regarding functional outcome.<sup>36</sup> The developers of the scale also report that it has excellent internal consistency and reliability. Like the OES, the LES is well correlated with the DASH, suggesting good construct validity for the measurement of disability. The scale has been found to have good responsiveness to change over time, independent of diagnosis, suggesting that the scale might be applicable across a broad range of clinical scenarios. The developers of the scale report that pain accounts for only 11% of the score compared to other systems that weight pain as 40% to 60% of the total score. It is unclear whether this should be seen as an advantage or disadvantage of this instrument.



When used as a mailed questionnaire, the LES correlated well with the MEPS, a potentially useful attribute for some study designs.

### Elbow Function Assessment

The EFA measures a combination of patient and physician-reported outcomes across 3 main domains: pain, ability to perform activities of daily living, and a motion assessment. The developers of the scale<sup>31</sup> report that the EFA demonstrated better validity than the MEPS for comparisons between dominant and nondominant elbows, a substantial consideration for some study designs. Internal consistency is high. The correlation between EFA scores and radiographic findings was rated as fair.

Later studies<sup>37</sup> have shown that the EFA is more responsive to change than the MEPS and better able to detect clinically meaningful difference. Among the tested instruments, which included the MEPS, the EFA had the best discriminative ability to distinguish improved from unchanged patients.

In conclusion, the choice of an outcome measurement scale should always be dictated by the needs of the research question. In many instances, the use of a series of scales is appropriate. For conditions that have an impact on overall health, generic health measures are informative as a reflection of general health status. Among these, the SF-36 has been the most widely tested. When disability is the focus, the DASH might be the best choice whether the condition of interest most impacts the shoulder or wrist/hand. Disability specifically related to the elbow might be better evaluated with a region-specific instrument such as the OES or LES.

Region-specific scales such as the MHQ and PRWE or any of the well-established shoulder scales are the best choices for studies that are not necessarily focused on the construct of disability. The best instrument for evaluation of the elbow appears to be more dependent on the specific research question because the various scales perform differently for various conditions. This is probably a reflection of the content of each individual scale.

A disease-specific scale might be most sensitive to changes related to treatment for the condition for which the instrument was designed. The best example of this would be the Boston Carpal Tunnel Syndrome Scale for the assessment of outcomes after treatment of carpal tunnel syndrome.

### References

1. Hudak PL, Amadio PC, Bombardier C. Development of an upper extremity outcome measure: the DASH (disabilities of the arm, shoulder and hand). The Upper Extremity Collaborative Group (UECG). *Am J Ind Med.* 1996; 29(6):602–608. [PubMed: 8773720]
2. Chung KC, Pillsbury MS, Walters MR, et al. Reliability and validity testing of the Michigan Hand Outcomes Questionnaire. *J Hand Surg Am.* 1998; 23(4):575–587. [PubMed: 9708370]
3. Pransky G, Feuerstein M, Himmelstein J, et al. Measuring functional outcomes in work-related upper extremity disorders. Development and validation of the Upper Extremity Function Scale. *J Occup Environ Med.* 1997; 39(12):1195–1202. [PubMed: 9429173]
4. MacDermid JC. Development of a scale for patient rating of wrist pain and disability. *J Hand Ther.* 1996; 9(2):178–183. [PubMed: 8784681]

5. Levine DW, Simmons BP, Koris MJ, et al. 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(11):1585–1592. [PubMed: 8245050]
6. Reeback J, Silman A. Predictors of outcome at two years in patients with rheumatoid arthritis. *J R Soc Med.* 1984; 77(12):1002–1005. [PubMed: 6512801]
7. Mason JH, Anderson JJ, Meenan RF. A model of health status for rheumatoid arthritis. A factor analysis of the Arthritis Impact Measurement Scales. *Arthritis Rheum.* 1988; 31(6):714–720. [PubMed: 3382446]
8. Spitz PW, Fries JF. The present and future of comprehensive outcome measures for rheumatic diseases. *Clin Rheumatol.* 1987; 6(Suppl 2):105–111. [PubMed: 3690982]
9. Bellamy N, Campbell J, Haraoui B, et al. Dimensionality and clinical importance of pain and disability in hand osteoarthritis: Development of the Australian/Canadian (AUSCAN) Osteoarthritis Hand Index. *Osteoarthritis Cartilage.* 2002; 10(11):855–862. [PubMed: 12435330]
10. Beaton DE, Wright JG, Katz JN, et al. Development of the Quick-DASH: comparison of three item-reduction approaches. *J Bone Joint Surg Am.* 2005; 87(5):1038–1046. [PubMed: 15866967]
11. Chung KC, Hamill JB, Walters MR, et al. The Michigan Hand Outcomes Questionnaire (MHQ): assessment of responsiveness to clinical change. *Ann Plast Surg.* 1999; 42(6):619–622. [PubMed: 10382797]
12. Shauver MJ, Chung KC. The minimal clinically important difference of the Michigan hand outcomes questionnaire. *J Hand Surg.* 2009; 34(3):509–514.
13. Waljee JF, Kim HM, Burns PB, et al. Development of a brief, 12-item version of the Michigan Hand Questionnaire. *Plast Reconstr Surg.* 2011; 128(1):208–220. [PubMed: 21701336]
14. 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(1):18–23. [PubMed: 14770134]
15. MacDermid JC, Turgeon T, Richards RS, et al. Patient rating of wrist pain and disability: a reliable and valid measurement tool. *J Orthop Trauma.* 1998; 12(8):577–586. [PubMed: 9840793]
16. Leite JC, Jerosch-Herold C, Song F. A systematic review of the psychometric properties of the Boston Carpal Tunnel Questionnaire. *BMC Musculoskelet Disord.* 2006; 7:78. [PubMed: 17054773]
17. Amirfeyz R, Pentlow A, Foote J, et al. Assessing the clinical significance of change scores following carpal tunnel surgery. *Int Orthop.* 2009; 33(1):181–185. [PubMed: 17972075]
18. Bellamy N, Campbell J, Haraoui B, et al. Clinimetric properties of the AUSCAN Osteoarthritis Hand Index: an evaluation of reliability, validity and responsiveness. *Osteoarthritis Cartilage.* 2002; 10(11):863–869. [PubMed: 12435331]
19. Allen KD, Jordan JM, Renner JB, et al. Validity, factor structure, and clinical relevance of the AUSCAN Osteoarthritis Hand Index. *Arthritis Rheum.* 2006; 54(2):551–556. [PubMed: 16447231]
20. Roy JS, MacDermid JC, Woodhouse LJ. Measuring shoulder function: a systematic review of four questionnaires. *Arthritis Rheum.* 2009; 61(5):623–632. [PubMed: 19405008]
21. Godfrey J, Hamman R, Lowenstein S, et al. Reliability, validity, and responsiveness of the simple shoulder test: psychometric properties by age and injury type. *J Shoulder Elbow Surg.* 2007; 16(3):260–267. [PubMed: 17188906]
22. Kocher MS, Horan MP, Briggs KK, et al. Reliability, validity, and responsiveness of the American Shoulder and Elbow Surgeons subjective shoulder scale in patients with shoulder instability, rotator cuff disease, and glenohumeral arthritis. *J Bone Joint Surg Am.* 2005; 87(9):2006–2011. [PubMed: 16140816]
23. Angst F, Goldhahn J, Drerup S, et al. Responsiveness of six outcome assessment instruments in total shoulder arthroplasty. *Arthritis Rheum.* 2008; 59(3):391–398. [PubMed: 18311752]
24. Roddey TS, Olson SL, Cook KF, et al. Comparison of the University of California-Los Angeles Shoulder Scale and the Simple Shoulder Test with the shoulder pain and disability index: single-administration reliability and validity. *Phys Ther.* 2000; 80(8):759–768. [PubMed: 10911414]

25. MacDermid JC, Drosdowech D, Faber K. Responsiveness of self-report scales in patients recovering from rotator cuff surgery. *J Shoulder Elbow Surg.* 2006; 15(4):407–414. [PubMed: 16831642]
26. MacDermid JC, Solomon P, Prkachin K. The Shoulder Pain and Disability Index demonstrates factor, construct and longitudinal validity. *BMC Musculoskelet Disord.* 2006; 7:12. [PubMed: 16472394]
27. 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(6):587–594. [PubMed: 12469084]
28. Sallay PI, Reed L. The measurement of normative American Shoulder and Elbow Surgeons scores. *J Shoulder Elbow Surg.* 2003; 12(6):622–627. [PubMed: 14671531]
29. Constant CR, Gerber C, Emery RJ, et al. A review of the Constant score: modifications and guidelines for its use. *J Shoulder Elbow Surg.* 2008; 17(2):355–361. [PubMed: 18218327]
30. Morrey BF. *Orthopedics.* JAMA. 1993; 270(2):240–242. [PubMed: 8123078]
31. de Boer YA, van den Ende CH, Eygendaal D, et al. Clinical reliability and validity of elbow functional assessment in rheumatoid arthritis. *J Rheumatol.* 1999; 26(9):1909–1917. [PubMed: 10493668]
32. 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(2):154–162. [PubMed: 9486721]
33. Doornberg JN, Ring D, Fabian LM, et al. Pain dominates measurements of elbow function and health status. *J Bone Joint Surg Am.* 2005; 87(8):1725–1731. [PubMed: 16085611]
34. Dawson J, Doll H, Boller I, et al. The development and validation of a patient-reported questionnaire to assess outcomes of elbow surgery. *J Bone Joint Surg Br.* 2008; 90(4):466–473. [PubMed: 18378921]
35. Dawson J, Doll H, Boller I, et al. Comparative responsiveness and minimal change for the Oxford Elbow Score following surgery. *Qual Life Res.* 2008; 17(10):1257–1267. [PubMed: 18958582]
36. Sathyamoorthy P, Kemp GJ, Rawal A, et al. Development and validation of an elbow score. *Rheumatology (Oxford).* 2004; 43(11):1434–1440. [PubMed: 15304676]
37. de Boer YA, Hazes JM, Winia PC, et al. Comparative responsiveness of four elbow scoring instruments in patients with rheumatoid arthritis. *J Rheumatol.* 2001; 28(12):2616–2623. [PubMed: 11764206]