

Outcome evaluation measures for wrist and hand – which one to choose?

Manish Changulani · Ugochuku Okonkwo ·
Tulsi Keswani · Yegappan Kalairajah

Received: 15 February 2007 / Accepted: 21 March 2007 / Published online: 30 May 2007
© Springer-Verlag 2007

Abstract The aim of this study was to critically analyse the various outcome measures available for assessing wrist and hand function. To this end, an extensive literature search was performed on Medline, PubMed and the Science Citation Index, focusing on terms associated with the method of development of the outcome measures item generation, item reduction, validity, reliability, internal consistency and their strengths and weaknesses. The most commonly used outcome measures described in literature were the DASH score (disability of shoulder, arm and hand questionnaire), the PRWE score (patient-rated wrist evaluation questionnaire), the Brigham and Women's carpal tunnel questionnaire and the Gartland and Werley score. Our study provides very useful evidence to suggest that the PRWE score is the most responsive instrument for evaluating the outcome in patients with distal radius fractures, while the DASH score is the best instrument for evaluating patients with disorders involving multiple joints of the upper limb. The Brigham and Women's score is a disease-specific outcome instrument for carpal tunnel syndrome; it has been validated and demonstrated to show good responsiveness and reliability in evaluating outcome in patients with carpal tunnel release. The Gartland and Werley score, although the most commonly described

instrument in the literature for evaluating outcome after wrist surgery, has not been validated so to date.

Résumé Le but de ce travail est d'analyser de façon critique les différents scores fonctions du poignet et de la main. Une analyse importante de la littérature a été réalisée sur Medline, PubMed et l'Index Science Citation. Cette analyse a été complète prenant en compte les différents scores. La méthode de mesure la plus habituelle a été le score DASH pour les lésions de l'épaule, du bras et de la main, le score PRWE (pour le poignet), celui de Brigham pour le canal carpien et surtout le poignet, le score de Gartland et celui de Werley. Notre étude permet de mettre en évidence que le score le plus adapté pour les fractures de l'extrémité distale du radius est le score PRWE. Le score DASH est le meilleur instrument pour mesurer les atteintes multiples des articulations des membres supérieurs. Le score de Brigham est spécialement adapté au syndrome du canal carpien, le score de Gartland et Werley est le score le plus habituellement utilisé dans la littérature pour évaluer la chirurgie du poignet.

Introduction

Outcome assessment has become important in evaluating the efficacy of surgical procedures. In accordance with this, most orthopaedic surgeons are now of the opinion that a proper outcome assessment should be performed after any form of surgery. Such an assessment facilitates surgeons in distinguishing between various treatment methods and helps to identify effective treatment options which, in turn, improves patient care.

A wide variety of outcome measures have been proposed for upper limb extremity disorders, including those for the evaluation of wrist and hand function. Some of these are

M. Changulani (✉) · U. Okonkwo · Y. Kalairajah
Orthopaedic Research Unit,
South West London Elective Orthopaedic Centre (SWLEOC),
Dorking Road,
Epsom, Surrey, KT18 7EG, UK
e-mail: changulani8@aol.com

T. Keswani
Flat 2, Windsor House,
Bellgarth Square,
Carlisle, CA2 7PH, UK

generic instruments, such as the Short Form (SF)-36 [14] and sickness impact profile [3]. These generic measures assess the impact of musculoskeletal problems on the overall health and well being of patients, and they were designed for broad use in a variety of disorders. However, more specific outcome instruments have been designed for specific use in musculoskeletal problems, including those specific for anatomical regions, such as the patient-rated wrist evaluation score (PRWE) [12], and those for outcome measures for specific diseases, such as carpal tunnel syndrome (CTS) [9].

The traditional methods for evaluating wrist and hand function following an intervention consist of measuring grip strength and assessing the range of motion, both which provide a good, objective analysis of outcome. However, these methods do not take into account other aspects related to an analysis of outcome, such as the patient's ability to carry out activities of daily living, the ability to return to previous occupations and pain. Hudak et al. [7] emphasised that to evaluate outcome following hand surgery, appropriate, reliable and validated outcome measures are required that take into account all aspects of patient life that may be affected.

This aim of this article is to critically analyse the outcome measures commonly used in the evaluation of wrist and hand function.

Materials and methods

An extensive literature search was carried out on Medline, PubMed and other search engines available online. The outcome instruments described in the literature for evaluating wrist and hand function are the disability of shoulder, arm and hand questionnaire (DASH), Brigham and Women's Hospital carpal tunnel questionnaire (CTQ), the patient-rated wrist evaluation questionnaire (PRWE), the Gartland and Werley score, the hospital for special surgery wrist scoring system (HSS), the Lamberta and Clayton wrist score and the Wrightington wrist function score. We selected the four most commonly used instruments – the DASH questionnaire, PRWE score, Gartland and Werley score and Brigham and Women's Hospital CTQ – for further evaluation in terms of their development, validity, reliability, consistency and strengths and weaknesses.

Analysis of outcome measures

DASH questionnaire

The DASH score was first described in 1996 by Hudak et al. [7]. The main objective behind its development was to develop a regional outcome measure which conceptualises

the upper extremity as a single functional unit. This would allow greater uniformity in research and would give greater relevance to the input from the patient himself rather than relying on other factors, such as radiographs, range of motion and grip strength.

DASH claims to assess both symptoms and functional status with a focus on physical function in populations with upper extremity musculoskeletal conditions. DASH is self administered by patients and aims to capture the patient's own perception of upper extremity function.

Development, item generation and item reduction

Thirteen scales and 821 items were chosen after extensive review of the literature, and these were used in measuring the outcomes of various upper extremity conditions. The initial item reduction was done on the basis of judgement from experts, with a subsequent reduction to 177 items. These were later reduced to 75 by content experts and finally reduced to 30 items after preliminary testing on patients.

Scoring

The DASH questionnaire comprises 30 items that evaluate symptoms and physical function with five response options for each item. The final score can be calculated using a simple formula:

$$\text{DASH score} = \left\{ \frac{\text{sum of } n \text{ responses}}{n} - 1 \right\} \times 25,$$

n = number of completed responses

The questionnaire takes the patient 10–15 min to complete and the administrator takes another 10 min to calculate the final score, which makes this a time-consuming outcome instrument. The reliability, as reported by Cronbach's alpha, is 0.9615, and test–retest reliability is 0.9219.

Construct validity

Convergent construct validity was demonstrated through the correlation between the DASH score and other joint specific instruments, such as the Brigham CTQ (0.73) and SPADI (shoulder pain and disability index) (0.72). The correlation of DASH with severity of pain in the wrist joint was weak (0.67), making it less valid for use in patients with wrist disorders.

Test–retest reliability

A group of 86 patients was asked to complete DASH at baseline and then 3–5 days later. The Pearson correlation

between the baseline and retest scores was 0.96, suggesting an excellent reproducibility for the DASH score.

Responsiveness

The change in DASH was found to correlate well with changes in the patient's condition. The DASH questionnaire demonstrated a change in all situations in which change presumably occurred. The standardised response mean (SRM) of 0.74 for DASH was comparable to 0.76 for such joint specific outcome measures as the Brigham score. This demonstrates the ability of DASH to differ in accordance with alterations in the patient's condition and its ability to show even very small change.

Brigham and Women's Hospital CTQ

This self-administered questionnaire was first described by Levine et al. in 1993 [9]. It was developed to assess the severity of symptoms and functional status and response to treatment in patients with CTS.

Development, item generation and reduction

Following consultation with hand surgeons and rheumatologists, Levine et al. identified six critical domains for the evaluation of CTS: pain, test–retest reliability of paraesthesia, numbness, weakness, nocturnal symptoms and overall functional status [9]. A symptom severity scale was developed comprising 11 questions incorporating these six domains. Twelve functional activities commonly affected in CTS, such as writing and holding a cup, were also identified. These were reduced to eight after pilot testing and included in the questionnaire as a functional status scale.

Scoring

The patients were asked to answer all 11 questions included in the symptom severity scale and the eight questions included in the functional status scale. The answer to each multiple-choice question ranged from mild (1 point) to most severe (5 points). The overall score was calculated as the sum of the mean of scores for all items on the symptom severity scale and functional status scale.

Validity

Content validity was tested by consulting a group of hand surgeons, rheumatologists and patients. A correlation between the scores on the scales and a variety of physical instruments, such as grip strength and pinch strength, used for measuring hand function was determined through a prospective study on 67 patients. The scores for severity of

symptoms had a moderate Spearman correlation with grip and pinch strength. The functional status scores had a high correlation with the severity of the symptoms and a moderate correlation with grip and pinch strength. A correlation was also calculated for patient satisfaction after the operation and the improvement in scores: a greater satisfaction was associated with a greater improvement in scores for both the severity of symptoms and functional status. All of these correlations were statistically significant. Overall, this indicated a good validity for the CTQ score.

Test–retest reliability

A group of 39 patients were asked to complete the questionnaire on two separate occasions on two consecutive days. The Pearson correlation coefficient was 0.91 for the symptom severity scale and 0.93 for functional status, indicating very good reproducibility.

Internal consistency

Cronbach's alpha indicating inter-item correlation within each scale was 0.89 for the symptom severity scale and 0.91 for the functional status scale. This implies an excellent internal consistency between the different items on the scale and also means that the scales could function well as a unidimensional index of severity of symptoms and functional status for patients with CTS.

Responsiveness

Responsiveness was tested on 38 patients who underwent carpal tunnel release surgery. The preoperative symptom severity score was 3.4 ± 0.67 (mean and SD); the mean postoperative score was 1.9 ± 1.0 . These scores indicate a substantial responsiveness to clinical change. The effect size was 1.4. The preoperative functional status score was 3.0 ± 0.93 compared with the postoperative functional score of 2.0 ± 1.1 , again a substantial improvement. The effect size in this case was 0.82. As an additional indicator of responsiveness, the correlation between patient satisfaction with the results of the operation and the reduction in score was calculated. This correlation was good, suggesting that CTQ is sensitive to change in the clinical picture in patients with CTS.

Patient-rated wrist evaluation score

The PRWE score was originally described by MacDermid et al. in 1998 [12]. The aim of the questionnaire is to provide a reliable and valid tool for quantifying patient-rated wrist pain and disability in order to assess outcome in patients with distal radius fractures.

Development, item generation and reduction

The questionnaire was developed by surveying wrist experts, reviewing the biomechanical literature and carrying out patient interviews. This resulted in the identification of the domains of pain and function as priorities for the evaluation of wrist function. The items in both these essential domains were further reduced by expert and patient review as well as pilot testing. Pain items were modified to incorporate the whole spectrum of severity, both in intensity and frequency. Functional items were modified to include items that were commonly performed by either hand, performed by most of the patients and easy to understand. The intention was that the questionnaire be simple and brief.

Scoring

It is self administered by the patient. The score consists of two domains – pain and function – both of which carry equal weight. There are five items in the pain domain and ten items in the function domain. The response to each item is scored on a scale of 0–10. The pain score is the sum of five items, a worse score of 50; the disability (function) score is the sum of ten items, divided by 2. Thus, the total function on the PRWE scale ranges from 0 (normal wrist) to 150 (worst possible score).

Construct validity

The change in the disability over time was evaluated in 101 patients with wrist fractures. A statistically significant improvement was found ($p < 0.0001$), with the amount of improvement being 74% as compared to the SF-36 score, which reported an improvement of 14% ($p < 0.0001$).

Criterion validity

The PRWE score was correlated with the SF-36 score and with an impairment score that was based on an assessment of physical functions, such as range of movement of wrist joint, grip strength and dexterity. The PRWE score showed a correlation with the SF-36 score of between 0.33 and 0.73. There was a low correlation with the SF-36 mental summary score and a high correlation with bodily pain score and physical function score. The PRWE score correlated poorly – 0.52 (weak to moderate correlation) – with an impairment score (score for the measurement of function impairment in patients, which raises questions over the validity of this score, as an impairment score is the aspect which corresponds to the function of the PRWE scale, an important aspect when evaluating outcome in patients with distal radius fractures).

Test–retest reliability

This was tested on three groups of patients. Groups 1 and 2 comprised patients with distal radius fractures currently undergoing physiotherapy and having completed physiotherapy, respectively, while Group 3 patients had scaphoid fracture non-union and were tested for long-term retest reliability. A short-term retest reliability testing was performed on the first two groups. An excellent intra-class correlation (ICC; > 0.90) was found for pain subscales for all three groups. The function subscales showed an excellent reliability in the distal radius fracture group (ICC > 0.85) but only moderate reliability over the long-term in Group 3 (ICC > 0.61). No appropriate testing for internal consistency and responsiveness was performed, which makes the PRWE score rather weak in terms of overall reliability.

Table 1 Outcome measures for analysing wrist and hand

Outcome measures ^a	Assesses	Anatomical region	Administrator	Format	Validity	Reliability	Responsiveness
DASH	Symptoms, function	Upper limb	Patient	30-item questionnaire	Good	Good	Good
CTQ	Symptoms, function	Carpal tunnel	Patient	19-item questionnaire	Good	Good	Good
PRWE	Symptoms, function	Wrist, number	Patient	15-item questionnaire	Fair	Good	Good
Gartland and Werley	Function	Wrist, hand	Clinician		None performed	None performed	None performed

^aDASH, Disability of shoulder, arm and hand questionnaire; CTQ, the Brigham and Women's Hospital carpal tunnel questionnaire; PRWE, patient-rated wrist evaluation questionnaire

Gartland and Werley score

This is one of the most commonly used outcome measures for evaluating wrist and hand function. This was initially described in 1951 by Gartland and Werley [4]. This score is completed by the administrator after the patient has been examined.

This system is based on a demerit point system which involves an objective evaluation of wrist function. It relies on the concept that a minimum of 45° dorsiflexion, 30° palmar flexion, 15° ulnar and radial deviation and 50° pronation and supination is normal. Demerit points are given based on the presence of a specific arbitrarily determined degree of loss of range of movement. For example, five points are given for a 45° loss of dorsiflexion, and only one point is given for loss of palmar flexion of more than 30°. Depending on the number of points scored, the outcome is classified as excellent, good or poor. Sarmiento et al. [13] later modified the system to include a loss of pronation and grip strength.

Lucas and Sachtjen [10] further modified it by adding such non-objective variables of hand as median nerve impairment, reflex sympathetic dystrophy and the stiffness of digits. They removed grip strength from the criteria of functional outcome. These changes were implemented to incorporate all of the possible outcomes and complications that can occur following wrist injuries, particularly distal radius fractures.

Validity, reliability and responsiveness

Despite the extensive use of this outcome measure, there have been no validity studies carried out to date. This is the one of the very few outcome measures which we found to provide an objective evaluation of outcome and may well be the reason that makes this measure very popular among orthopaedic surgeons. However, no appropriate methodology seems to have been applied for identifying the domains which makes this less reliable for use.

Discussion

The relatively large number of outcome measures available for evaluating wrist and hand function provides clinicians with a wide range of choice, thereby enabling them to use that outcome instrument which is the most appropriate and suitable. The choice of an outcome measure is determined by the clinical condition one wishes to assess; the resources available and the psychometric properties are often additional determining factors [1].

We analysed the DASH score because this is the only score which considers the whole upper limb as a single

unit; as such, it may be useful for assessing outcome in any upper limb pathology irrespective of the site [6]. PRWE was analysed as this score is specific for the outcome from one joint. The Brigham CTQ was chosen for analysis as it is disease-specific. The Gartland and Werley score is the most commonly used outcome measure in the literature and the only one dependent on the administrator's objective assessment.

Karnezis et al. [8] compared the association between objective clinical variables, such as grip strength and wrist movements, and PRWE score by means of regression analysis, which revealed the limitations of objective assessment in reflecting the level of disability of the wrist. These researchers were unable to establish an association between PRWE and the Gartland and Werley score, which proved that movements of the wrist joint and grip strength alone are not a reliable way of measuring outcome. Gay and et al. [5] analysed the comparative responsiveness of the DASH score, the Brigham wrist score and the SF-36 to clinical change after carpal tunnel release. The instrument most sensitive to clinical change, assessed at 12 weeks post-carpal tunnel release, was the Brigham score (effect size/standardised response means; 1.71/1.66), followed by the DASH score (1.01/1.13) and the SF-36 score (0.57/0.52). There was a good correlation between the DASH and the Brigham score (Spearman correlation coefficient: 0.87), which makes the Brigham score a reliable, valid and sensitive tool for assessing outcome in patients with CTS.

Beaton et al. [2] compared the validity, reliability and responsiveness of the DASH score with those obtained from joint-specific measures and found that the former correlated well with other joint-specific measures such as the Brigham and Women's CTQ score for the wrist joint and the SPADI score for the shoulder joint. They also found that the responsiveness of the DASH score to self-rated or expected change was comparable to or better than other joint-specific measures, both in the whole group and in each region. This confirmed the usefulness of the DASH score across the whole upper limb, particularly in patients with multiple upper limb joint involvement.

Macdermid et al. [11] compared the responsiveness of the DASH, PRWE and SF-36 scores in evaluating recovery after distal radius fractures. The PRWE score was the most responsive of the three in this particular group of patients (SRM: 2.27), followed by the DASH (SRM: 2.01) and the SF-36 (SRM: 0.92). This makes the PRWE score a reasonably reliable, valid and sensitive tool for assessing outcome in patients with distal radius fractures.

Conclusion

Table 1 presents a summary of the four outcome instruments discussed in detail in this article. The DASH score is the best

instrument for evaluating patients with disorders involving multiple upper limb joints. The Brigham score is a disease-specific validated outcome instrument for carpal tunnel syndrome. The PRWE score is a validated tool for assessing outcome in patients with distal radius fractures, and the Gartland and Werley score provides an objective assessment of outcome, but its use has not yet been validated.

References

- Bland JM, Altman DG (1986) Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1:307–310
- Beaton DE, Katz JN, Fossel AH et al (2001) Measuring the whole or the parts? Validity, reliability, and responsiveness of the DASH outcome measure in different regions of the upper extremity. *J Hand Ther* 14:128–142
- Bergner M, Bobbitt RA, Carter WB et al (1981) The sickness impact profile: development and final revision of a health status measure. *Med Care* 19:787–805
- Gartland JJ Jr, Werley CW (1951) Evaluation of healed Colles' fracture. *J Bone Joint Surg Am* 33:895–907
- Gay RE, Amadio PC et al (2003) 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 [Am]* 28:251–254
- Gummeson C, Atroshi I, Ekdahl C (2003) The disabilities of the arm, shoulder and hand (DASH) outcome questionnaire: longitudinal construct validity and measuring self-rated health change after surgery. *BMC Musculoskeletal Disorders* 12:349–362
- Hudak PL, Amadio PC, Bombardier C (1996) Development of an upper extremity outcome measure: the DASH. *Am J Int Med* 29:602–608
- Karnezis IA, Fragkiadakis EG (2002) Association between objective clinical variables and patient rated disability of the wrist. *J Bone Joint Surg Br* 84:967–970
- Levine DW, Simmons PB, Koris MJ et al (1993) A self administered questionnaire for the assessment of severity of symptoms and functional status in carpal tunnel syndrome. *J Bone Joint Surg Am* 75:1585–1991, Nov
- Lucas GL, Sachtjen KM (1981) An analysis of hand function in patients with Colles' fracture treated by rush rod fixation. *Clin Orthop* 155:172–179
- MacDermid JC, Richards RS, Donner A, Bellamy N, Roth JH (2000) Responsiveness of SF-36, DASH and PRWE and physical impairments in evaluating recovery after distal radius fracture. *J Hand Surg [Am]* 25:330–340
- MacDermid JC, Turgeon T, Richards RS, Beadle M, Roth JH (1998) Patient rating of wrist pain and disability: a reliable and valid measurement tool. *J Orthop Trauma* 12:77–86
- Sarmiento A, Pratt GW, Berry NC, Sinclair WF (1975) Colles' fracture: functional bracing in supination. *J Bone Joint Surg Am* 57:311–317
- Ware JE Jr, Kosinski M, Bayliss MS et al. (1995) Comparison of methods for the scoring and statistical analysis of SF-36 health profile and summary measures: summary of results from the Medical Outcomes Study. *Med Care* 33 [Suppl 4]:264–279