The combined shoulder assessment: a convenient method for obtaining equivalent outcome scores

Emma Torrance¹, Linda Hallam¹, Michael J. Walton^{1,2,3},
Puneet Monga^{1,2,3}, Adam C. Watts^{1,2,4} and Lennard Funk^{1,2,3}

Abstract

Background: Outcome scores are useful tools with respect to quantifying how pain and limitations in strength and movement affect the patient's functionality and lifestyle. Three shoulder outcome methods [Oxford, Constant and QuickDASH (Disability of the Arm, Shoulder and Hand)] are commonly used in practice but collectively consist of 34 questions, taking approximately 7 min to complete. We have developed a concise shoulder assessment to derive three equivalent outcome scores in a 12-item patient-reporting questionnaire.

Methods: Outcome scores from 1285 outpatients of an upper limb clinic were collected. The patient cohort consisted of 462 females and 823 males with a mean (SD) age of 47.2 (16.79) years (range 13 years to 90 years). Using a correlation matrix, a 12-item questionnaire was drafted. The combined shoulder assessment was validated by 227 patients; consisting of 101 females and 126 males with a mean (SD) age of 47.91 (16.63) years (range 13 years to 88 years).

Results: Agreement was achieved between the two methods, with an equivalent Oxford Shoulder Score intraclass correlation (ICC) of 0.930, equivalent Constant Shoulder Score ICC of 0.942 and equivalent QuickDASH ICC of 0.869. Bland–Altman analyses showed no systematic differences. Large effect sizes highlighted the responsiveness to change.

Conclusions: The new combined shoulder assessment is a more convenient and patient-friendly method to obtain equivalent Oxford, Constant and QuickDASH shoulder outcome scores.

Keywords

equivalent Constant Shoulder Score, equivalent QuickDASH, equivalent Oxford Shoulder Score, outcome scores, shoulder

Date received: 3rd November 2016; revised: 8th June 2017; accepted: 21st September 2017

Introduction

The use of outcome scores in orthopaedics has evolved greatly in the past 20 years, moving from objective clinician-reported measurements to patient-reported subjective measures.¹ The purpose of functional outcome scores is to understand what effect the affected shoulder is having on the patient's lifestyle. Questions are posed with regards to pain, disturbances of sleep, occupational and recreational hindrances.^{2–4}

There are a multitude of scoring systems utilized daily in orthopaedic practice and research, many of which have been specifically developed for the shoulder. In Europe, three of the most commonly distributed shoulder scoring systems include the Oxford Shoulder

Score (OSS), Constant Shoulder (CS) Score and QuickDASH (Disability of the Arm, Shoulder and Hand) (QD).⁵ All three scoring instruments contain subjective questions, with the CS Score investigating an additional objective element regarding strength and range of motion (ROM). Multiple scoring methods

¹The Wilmslow Hospital, Wilmslow, UK

²Wrightington Hospital, Wigan, UK

³University of Salford, Manchester, UK

⁴University of Manchester, Manchester, UK

Corresponding author:

Lennard Funk, Wrightington Hospital Hall Lane Appley Bridge Wigan WN6 9EP, UK.

Email: lenfunk@shoulderdoc.co.uk

are often employed to obtain a complete picture of the affected shoulder because each scoring method targets slightly different aspects, with the CS Score weighted toward ROM, whereas the OSS and QD specifically target quality of life.

In our unit, all new patients complete OSS, CS Score and QD questionnaires prior to clinical intervention. Following surgery or conservative treatment, patients are regularly recalled to complete OSS, CS Score and QD to assess their progress. A disadvantage of this process lies in the number of questions patients are expected to answer: a total of 34 questions, taking approximately 7 min to complete.^{6,7} Furthermore, many of the questions posed in each scoring system are very similar, almost to the point of duplication. Previous studies have attempted to overcome this inconvenience by correlating endpoint scores, although this has met with varying success.^{8–10} Furthermore, this method does not allow for the extraction of numerical scores but, instead, merely the likelihood of a poor outcome in one scoring system associating with a poor outcome in another.

The present study aimed to develop a novel questionnaire, namely The combined shoulder assessment (CSA), aiming to obtain equivalent outcome scores to assess improvement in the patient's treatment cycle. The concise assessment was constructed and substantiated in one of the largest cohorts in validated shoulder outcome scores to date.

Materials and methods

Stage 1: CSA development

Patients. Data from 1285 outpatients of an upper limb unit were collected over a 10-month period. All patients had been treated by one of three senior shoulder surgeons. The patient cohort consisted of 462 females and 823 males with a mean (SD) age of 47.2 (16.79) years (range 13 years to 90 years).

Shoulder outcome scores. Patients routinely completed shoulder outcome scores throughout their time in the care of three senior shoulder surgeons, with the aim of assessing the functionality and impact of the affected shoulder on lifestyle. The outcomes measures included in the present study were OSS, CS Score and the QuickDASH. Responses from all patients had been systematically scored on a five-point scale and analyzed in a correlation matrix to assess the association of each response to the posed questions.

Data were assessed to be normal and fitting to Gaussian distribution. Correlations coefficients were determined using a Pearson's *r*-test. Statistically significant ($p < 0.05$) correlation coefficients were detailed in

accordance with Cohen.¹¹ Subjective questions from the OSS, CS Score and the QuickDASH with strong correlations were grouped appropriately to reduce the number of questions posed to patients (Figure 1). Questions were initially categorized into pain, quality of life, work and sport. Within each category, questions were grouped depending on their correlation value. Questions were only accepted into the 'group' if they had a strong correlation¹¹ between with questions within the same group. The posed question was selected as one within that group that had the highest mean *r*-value. When structuring the questions, some wording had changed slightly from the original questionnaires to accommodate the desired conciseness of the CSA (i.e. 'How would you describe tingling in your arm, shoulder or hand?' compared to 'Please rate the severity of the following: tingling in arm, shoulder or hand'). From these results, a 12-item patient reporting questionnaire was drafted, as depicted in the Supporting information. (supplementary material online).

OSS. The Oxford Shoulder Questionnaire was first described in 1996 for use in patients undergoing shoulder procedures other than stabilization.³ The questionnaire consists of 12 subjective questions to be answered by the patient independently. The responses are rated on a five-point scale and cumulated to obtain a numerical score. The numerical score ranges from 0 to 48 (best outcome). Previously, the lowest score was attributed to the best outcome, however an adjustment was enforced in 2009 to compare the results with the more conventional worse-is-lower methods.

CS Score. The CS Score was developed by Constant and Murley in 1987 and was one of the first outcome measures developed to assess shoulder function.² It comprises five clinician-reported physical examination findings and three subjective patient-reported assessments. Four objective measurements assess range of motion and another regarding force in abduction. The maximal score on the Constant assessment is 100 points, where a high score is achieved in optimal health state.

QuickDASH Score. The QuickDASH is a 19-item symptom scale concerning the patient's upper extremity.⁴ The questions regard the severity of pain, activity-related pain, tingling, weakness and degree of difficulty in performing various physical activities because of an arm, shoulder or hand problem. As with the OSS, the responses are rated on a five-point scale and cumulated to obtain a numerical score, ranging from zero to 100 (worst outcome). The QuickDASH contains two modules, concerning the ability to work, as well as to perform sports and/or play musical instruments.

Figure 1. Development of the combined shoulder assessment. Questions are categorized into pain, quality of life, work and sport/recreation. Pearson's coefficient and the p -value are noted.

Figure 1. Continued.

Figure 1. Continued.

Stage 2: CSA validation

The CSA was validated by 227 patients, consisting of 101 females and 126 males with a mean (SD) age of 47.91 (16.63) (range 13 to 88). Patients completed both the original outcome assessment and the CSA.

Completed assessments were entered into a patient management database. Two independent researchers extracted data of either the original or CSA, blinded to the status of the corresponding score.

Statistical analysis

Statistical analysis was carried out using Prism, version 5 (GraphPad Software Inc., San Diego, CA, USA) and SPSS (IBM Corp., Armonk, NY, USA). Data used to develop the CSA fit to Gaussian distribution by the inclusion of pre-intervention and postintervention scores and subsequent parametric statistics were performed.

Correlation. Pearson's correlation was used to assess the association between the responses to all posed questions of OSS, CS Score and QuickDASH outcome. When $r \geq 0.5$, an association is considered to be 'strong', in accordance with Cohen's guidelines (1988).

Reliability. Reliability of a measurement refers to the extent to which measurements give consistent or accurate results. Two frequently used measures of reliability are internal consistency and test-retest reliability.

Cronbach's α was calculated in the present study. It comprises a reliability coefficient that measures the internal consistency of a measurement instrument, ranging from 0 to 1, with a Cronbach's α of 0.7 and more indicating sufficient internal consistency.^{12,13}

Test-retest reliability is defined as how well the scores of the participants can be distinguished from each other on the questionnaire, despite existing measurement errors.¹⁴ For all variables, test-retest reliability was calculated by the intraclass correlation coefficient (ICC) using a two-way random effects model with an absolute agreement definition. For the present study, an ICC exceeding 0.70 was defined as good reliability, an ICC of 0.40 to 0.70 as moderate reliability and an ICC of less than 0.40 as poor reliability.¹⁵

Agreement. Agreement concerns the measurement error and assesses how close the original scoring system is with the novel CSA method. For this purpose, the Bland-Altman method was used by plotting the mean difference between the two consecutive measurements against the SD of this difference.¹⁶ The limits of agreement were calculated as the mean difference $\pm$ 1.96 times the SD of the differences. The Bland-Altman plot provides a visual interpretation of possible systematic variation in differences over the range of measurement and outliers that are not revealed by regular correlation analyses.

Responsiveness to change. Responsiveness to change was assessed 20 months following the initial CSA scores. Sixty-two patients were included in the responsiveness

to change analysis. This is a method of calculating the extent of change measured by a questionnaire in a standardized way to allow comparison. To measure this responsiveness, the effect size was calculated as (mean CSA follow-up score – mean CSA first score)/SD of CSA first score.¹⁷ Effect sizes were categorized as small (0.20), moderate (0.50) or large (0.80), in accordance with the literature.¹⁸ Paired *t*-tests were also conducted to support the effect sizes.

Results

All scoring systems in questions produced an acceptable test–retest reliability of > 0.70; equivalent OSS (eOS) re-test ICC of 0.92, equivalent CS Score (eCS) re-test ICC of 0.78 and equivalent QuickDASH (eQD) re-test ICC of 0.98.

The mean outcome scores obtained by both original individual questionnaires and equivalent values from The CSA were investigated. As depicted in Figure 2, comparable outcomes were achieved between the two methods: Oxford 30.32 (7.806) versus eOS CSA 29.35 (8.316); Constant 43.48 (20.54) versus eCS CSA 40.01 (17.70); and QuickDASH 46.79 (17.68) versus eQD CSA 47.43 (17.43). Furthermore, narrow and comparable 95% confidence intervals were also highlighted between the two methods: Oxford 29.30 to 31.34 versus eOS CSA 28.26 to 30.44; Constant 40.79 to 46.17 versus eCS CSA 37.70 to 42.33; and QuickDASH 44.46 to 49.11 versus eQD CSA 45.15 to 49.71. No significant difference was noted between the means of each scoring mechanism; eOS ($p = 0.204$), eCS ($p = 0.055$) and eQD ($p = 0.696$).

Figure 2. Mean (SEM) for outcome and equivalent outcomes scores retrieved by the original and combined shoulder assessment (CSA) methods. OS, Oxford Shoulder; CS, Constant Shoulder; QD, QuickDASH (Disability of the Arm, Shoulder and Hand); eOS, equivalent OS Score; eCS, equivalent CS Score; eQD, equivalent QuickDASH.

Table 1 reports the internal consistency and the reliability of the original scoring mechanism with the CSA. Internal consistency tested by the Cronbach's α exceeded 0.90 for eOS, eCS and eQD elements, showing high internal consistency. The ICCs for the scoring systems ranged between 0.869 and 0.942 with narrow confidence intervals, thus indicating good reliability.

The Bland–Altman plots were utilized to visually represent the agreement of scores obtained from the original method and novel CSA. It can be deduced that 97% of all outcome scores obtained by the CSA fall within the limits of agreement between the two methods. Figure 3 indicates a low bias with little variability between the methods, not related to the magnitude of the score. There were no floor or ceiling effects observed in this analysis.

Patients reported a significant improvement in eOS, eCS and eQD scores at 20 months following clinical intervention ($p < 0.0001$). Large effect sizes were recorded for eOS of 1.6, eCS of 1.4 and eQD of –1.9, highlighting the responsiveness to change.

Discussion

The present study introduces the CSA, comprising a short 12-item questionnaire to obtain three commonly used outcome scores. It has been validated as a novel method during specialist treatment and imposes very little burden on patients. The CSA reduces the number of patient reporting questions from 34 to 12. Subsequently, the time to taken to complete the questionnaire is reduced from approximately 7 min to 2 min and 30 s.^{6,7} In this era of technological advancement, we demand a more efficient service in all aspects of our lives and therefore reducing the time taken to complete the CSA is ultimately advantageous for patient experience.

To our knowledge, no other orthopaedic group has developed an amalgamated outcome scoring system. Previous studies have explored how endpoint numerical scores correlate between each other, with varied success.^{8–10,19} Razmjou et al.⁹ investigated the association between CS Scores and other outcome scores: Western

Table 1. Reliability of combined shoulder assessment with original scoring systems.

	Cronbach's α	ICC	ICC – 95% CI
Oxford	0.960	0.930	(0.909 to 0.949)
Constant	0.970	0.942	(0.925 to 0.955)
QuickDASH	0.930	0.869	(0.833 to 0.898)

Figure 3. Bland–Altman plots to represent agreement between the two scoring mechanisms. OS, Oxford Shoulder; CS, Constant Shoulder; QD, QuickDASH (Disability of the Arm, Shoulder and Hand); eOS, equivalent OS Score; eCS, equivalent CS Score; eQD, equivalent QuickDASH.

Ontario Rotator Cuff Index and the Rotator Cuff Quality of Life index. Correlations were notable between 0.56 and 0.75; however, the study focused solely on patients with rotator cuff pathology. Conversely, Placzek et al.¹⁰ performed a correlation study of numerous functional shoulder tests, including the CS Score, American Shoulder and Elbow Surgeons Shoulder Score (ASES), University of California at Los Angeles Shoulder Score (UCLA), Shoulder Pain and Disability Index (SPADI), Simple Shoulder Test (SST) in a group of 70 patients with shoulder pain and dysfunction. They reported generally poor

correlations among scoring instruments.¹⁰ Thus, it could be deduced that the correlation between endpoint of outcome scores can often be unreliable and therefore this highlights the opportunity to construct and review our novel method of obtaining shoulder outcome scores.

The present study presents high ICC values (0.869 to 0.930) with narrow confidence intervals (0.833 to 0.955), indicative of good reliability. The ICC values reported in the present study are comparable with the results previously described in translated versions of shoulder outcome scores with the original English version.^{20–22}

In clinical research, the impact of statistical significance often overshadows the clinical importance of the data. In the present study, we note a small change in the mean values between the original scoring systems and equivalent outcome scores from the CSA, although we must question what impact this has in clinical practice. A useful indicator is minimally clinically important change. The three scoring systems investigated in the present study have the benefit of an established minimal clinically important difference. Recently, Christiansen et al.²³ reported the minimally clinically important change for the OSS as 6 points and that for the CS Score as 11 points.²³ This supports the use of the CSA, which reports a change in the mean of 0.97 and 3.47 points for the equivalent OSS and CS Score, respectively. With regards to the QuickDASH, Franchignoni et al.,²⁴ report the minimally clinically important difference as 12.85 points. With a small difference in mean of 0.64 between the original and CSA, we can be confident that the novel questionnaire is in keeping with clinical confidence limits.

One notable difference between the original scoring mechanisms and the CSA was the timescale of the questions posed. Both the Oxford Shoulder and Constant scoring systems question the impact of the affected shoulder on the patient's lifestyle in the past 4 weeks, whereas the QuickDASH queries the impact of the affected shoulder within 1 week. The CSA Score omits any timescale when posing patient reporting questions. This decision was based on both analysis in the present study and previous work. Questions regarding 'normal' pain in all three scoring instruments were shown to have no significant differences despite varying timescale ($p=0.158$); this was also consistent for questions relating to sleep disturbance ($p=0.246$). The inability of patients to retrospectively recall pain is not a new phenomenon in clinical practice, or specifically patients with shoulder pathology.^{25,26}

Limitations

All patients were included in the design of the CSA; however, the OSS has not been validated in patients

with instability. The equivalent Oxford Instability Shoulder Score is currently in the process of validation as an additional module to the CSA.

All patients involved in the validation process were provided with both sets of questionnaires via post or e-mail prior to first consultation with a senior shoulder surgeon. The interval between patients completing the original assessments and the CSA were not measurable. Furthermore, the validation process included solely pre-intervention scores. Nevertheless, all shoulder conditions were included in the present study and therefore provided a full distribution of outcome scores. It should be noted that outcome scores obtained by the CSA are defined as 'equivalent' scores. In future studies, the method by which equivalent outcomes scores were obtained by the CSA should be clearly stated. Additionally, the CSA is not designed to facilitate retrospective calculation of equivalent outcomes scores and can only be utilized prospectively.

Conclusions

The CSA is a more convenient method to obtain equivalent Oxford, Constant and QuickDASH outcome scores in one patient-friendly questionnaire. The validation process highlights comparable means obtained by the novel method, with good reliability and agreement. The process undertaken to construct and substantiate the CSA has the potential to be expanded to include other shoulder scores and to be applied in other orthopaedic fields.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

Ethical review and patient consent

All patients signed consent forms to agree to the use of their anonymized data for scientific and research purposes. The return of the questionnaire was further confirmation of implied consent.

References

1. Wylie JD, Beckmann JT, Granger E and Tashjian RZ. Functional outcomes assessment in shoulder surgery. *World J Orthop* 2014; 5: 623–633.
2. Constant CR and Murley AH. A clinical method of functional assessment of the shoulder. *Clin Orthop Relat Res* 1987; 214: 160–164.
3. Dawson J, Fitzpatrick R and Carr A. Questionnaire on the perceptions of patients about shoulder surgery. *J Bone Joint Surg Br* 1996; 78: 593–600.
4. Hudak PL, Amadio PC and 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: 602–608.
5. Booker S, Alfahad N, Scott M, Gooding B and Wallace WA. Use of scoring systems for assessing and reporting the outcome results from shoulder surgery and arthroplasty. *World J Orthop* 2015; 6: 244–251.
6. Noorani AM, Roberts DJ, Malone AA, et al. Validation of the Stanmore percentage of normal shoulder assessment. *Int J Shoulder Surg* 2012; 6: 9–14.
7. Gabel CP, Michener LA, Burkett B and Neller A. The Upper Limb Functional Index: development and determination of reliability, validity, and responsiveness. *J Hand Ther* 2006; 19: 328–348.
8. Provencher MT, Frank RM, Macian D, et al. An analysis of shoulder outcomes scores in 275 consecutive patients: disease-specific correlation across multiple shoulder conditions. *Mil Med* 2012; 177: 975–982.
9. Razmjou H, Bean A, MacDermid JC, van Osnabrugge V, Travers N and Holtby R. Convergent validity of the Constant-Murley outcome measure in patients with rotator cuff disease. *Physiotherapy Canada* 2008; 60: 72–79.
10. Placzek J, Lukens S, Badalanmenti S, Rouba PJ, Freeman CD and Wiater MJ. Shoulder outcome measures. *Am J Sports Med* 2004; 32: 1270–1277.
11. Cohen J. *Statistical power analysis for the behavioral sciences*, 2nd edn. Mahwah, NJ: Lawrence Erlbaum, 1988.
12. Cronbach LJ. Coefficient alpha and the internal structure of tests. *Psychometrika* 1951; 16: 297–334.
13. Nunnally JC and Bernstein IR. *Psychometric theory*. New York, NY: McGraw-Hill, 1994.
14. de Vet HC, Terwee CB, Knol DL and Bouter LM. When to use agreement versus reliability measures. *J Clin Epidemiol* 2006; 59: 1033–1039.
15. Rankin G and Stokes M. Reliability of assessment tools in rehabilitation: an illustration of appropriate statistical analyses. *Clin Rehabil* 1998; 12: 187–199.
16. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1986; 1:307–110.
17. Lohr KN, Aaronson NK, Alonso J, et al. Evaluating quality-of-life and health status instruments: development of scientific review criteria. *Clin Ther* 1996; 18: 979–992.
18. Kane R. Outcome measures. In: Kane R (ed.) *Understanding health care outcomes research*. Gaithersburg: Aspen Publishers, 1997, pp.17–18.
19. Williams GN, Gangel TJ, Arciero RA, Uhorchak JM and Taylor DC. Comparison of the Single Assessment Numeric Evaluation method and two shoulder rating scales. Outcomes measures after shoulder surgery. *Am J Sports Med* 1999; 27: 214–221.
20. Berendes T, Pilot P, Willems J, et al. Validation of the Dutch version of the Oxford Shoulder Score. *J Shoulder Elbow Surg* 2010; 19: 829–836.

21. Murena L, Vulcano E, D'Angelo F, et al. Italian cross-cultural adaption and validation of the Oxford shoulder score. *J Shoulder Elbow Surg* 2010; 19: 335–341.
22. Wiertsema SH, Rietberg MB, Hekman KM, Schothorst M, Steultjens MP and Dekker J. Reproducibility of the Dutch version of the Western Ontario rotator cuff Index. *J Shoulder Elbow Surg* 2013; 22: 165–170.
23. Christiansen DH, Frost P, Falla D, Haahr JP, Frich LH and Svendsen SW. Responsiveness and minimal clinically important change: a comparison between 2 shoulder outcome measures. *J Orthop Sports Phys Ther* 2015; 45: 620–625.
24. Franchignoni F, Vercelli S, Giordano A, Sartorio F, Bravini E and Ferriero G. Minimal clinically important difference of the disabilities of the arm, shoulder and hand outcome measure (DASH) and its shortened version (QuickDASH). *J Orthop Sports Phys Ther* 2014; 44: 30–39.
25. Miranda H, Gold JE, Gore R and Punnett L. Recall of prior musculoskeletal pain. *Scand J Work Environ Health* 2006; 32: 294–300.
26. Tyler J and Singh B. 2 weeks after an arthroscopic shoulder procedure patients have forgotten their pre-operative symptoms. *Shoulder Elbow* 2015; 7: 309–332.