The development and validation of a questionnaire for rotator cuff disorders: The Functional Shoulder Score

Anestis Iossifidis^{1,2}, Edward F Ibrahim^{1,2},
Charalambos Petrou^{1,2} and Antonis Galanos³

Abstract

Background: The purpose of the present study was to validate the Functional Shoulder Score (FSS), a new patient-reported outcome score specifically designed to evaluate patients with rotator cuff disorders.

Methods: One hundred and nineteen patients were assessed using two shoulder scoring systems [the FSS and the Constant–Murley Score (CMS)] at 3 weeks pre- and 6 months post-arthroscopic rotator cuff surgery. The reliability, validity, responsiveness and interpretability of the FSS were evaluated.

Results: Reliability analysis (test–retest) showed an intraclass correlation coefficient value of 0.96 [95% confidence interval (CI) = 0.92 to 0.98]. Internal consistency analysis revealed a Cronbach's alpha coefficient of 0.93. The Pearson correlation coefficient FSS–CMS was 0.782 pre-operatively and 0.737 postoperatively ($p < 0.0005$). There was a statistically significant increase in FSS scores postoperatively, an effect size of 3.06 and standardized response mean of 2.80. The value for minimal detectable change was ± 8.38 scale points (based on a 90% CI) and the minimal clinically important difference for improvement was 24.7 ± 5.4 points.

Conclusions: The FSS is a patient-reported outcome measure that can easily be incorporated into clinical practice, providing a quick, reliable, valid and practical measure for rotator cuff problems. The questionnaire is highly sensitive to clinical change.

Keywords

PROMs, reliability, rotator cuff disorders, shoulder questionnaires, validity

Date received: 23rd October 2014; accepted: 26th February 2015

Introduction

Shoulder pain is common and accounts for many working days lost per year in the general population.¹ Rotator cuff pathology is the most common diagnosis, accounting for several million presentations per year.²

Increasing pressures of the economic climate mean that clinicians must aim to prove that their interventions are clinically effective at minimum financial cost. During the last three decades, a large number of shoulder assessment questionnaires have been developed.^{3,4} Scoring tools provide evidence regarding patients' perceptions of treatment and provide a vehicle to collect data on the effectiveness of intervention over the long-term.⁵ Measurement of all the possible effects of a disease or intervention is recommended and more recently the need to report functional outcome has been accepted as being a key part of assessment.⁶

Selecting the appropriate outcome measure is a vital component of fully evaluating a procedure or therapy.⁷ Shoulder outcome measures may be general, disease-specific or condition specific.⁸ Questionnaires can also be classified into objective questionnaires (assessed by a clinician), subjective questionnaires (patient-reported) or mixed. Questionnaires containing an objective

¹Croydon University Hospital, London, UK

²Shoulder & Upper Limb Unit, Croydon, London, UK

³Statistician, KAT Hospital, University of Athens, Greece

Corresponding author:

Mr Anestis Iossifidis, Consultant in Trauma & Orthopaedic Surgery, Shoulder & Upper Limb Surgeon, Croydon University Hospital, 560 London Road, Croydon, London, UK.

Tel: +44 (0)2084013102

Email: anestis.iossifidis@croydonhealth.nhs.uk

component may have the advantage of recording patient's muscle strength and range of motion but require an experienced examiner and additional equipment to perform tests and are prone to inter- and intra-observer bias.⁹ The Constant–Murley Score (CMS) is the most commonly used outcome tool containing objective measurement¹⁰ and has been found to be valid and reliable for a range of shoulder pathologies, including rotator cuff disease.^{11,12} Patient-reported outcome measures (PROMs) allow comparison through quantification of subjective results and have become standard in the reported literature.⁵ They have the advantage of being free from examiner bias and are less time consuming to complete for both clinician and patient in a busy practice.

Any method used to evaluate treatment must have been developed appropriately and assessed for acceptability, reliability, validity and responsiveness.¹³ Furthermore, reporting a minimum clinically important difference (MCID) is desirable. MCID is the minimum change in a score over time that can be considered to represent a significant change in the patient's clinical outcome and unlikely to be a result of chance. Although several scoring systems are well described, many in use have not been properly tested for these essential attributes and comparison studies are limited by the modification of scales and weighting for statistical analysis,¹⁴ only a few report a MCID value.⁵ In addition, few of these outcome tools are validated as disease specific for rotator cuff pathology and as sensitive to surgical intervention.¹⁴

The purpose of the present study was to assess the use of a new patient-reported outcome questionnaire, called the Functional Shoulder Score (FSS), specifically designed to evaluate patients presenting with, and receiving surgical intervention for, rotator cuff disorders. Direct comparison is made with the CMS.

Materials and methods

All patients diagnosed with primary rotator cuff disease who presented to a single shoulder surgeon and subsequently underwent arthroscopic surgical intervention over a 1-year period (January 2009 to January 2010) were included in the present study. Participants were invited to complete the FSS questionnaire and be assessed for the CMS at presentation for anaesthetic pre-assessment (3 weeks pre-operatively) and at 6 months postoperatively in a prospective fashion. All operations consisted of arthroscopic subacromial decompression with or without arthroscopic rotator cuff repair. All procedures were performed either by or under direct supervision of the same surgeon. Patients were excluded if pathology other than rotator cuff disease was felt to be the

primary diagnosis (e.g. adhesive capsulitis, osteoarthritis or fracture).

The FSS (Figure 1) is an 11-item patient-reported subjective outcome questionnaire consisting of two major categories – pain (one question) and activities of daily living (ADLs) (10 questions). Subjects answer each question on a 10-point numeric rating scale, with a lower number always indicating worse pain or greater difficulty in function.

The maximum score of 100 represents the total absence of pain and best possible function. The minimum score of 0 represents the worst possible result. Fifty points are allocated for pain and 50 for ADLs. The final score is calculated by multiplying the value indicated by the patient's answer to the question about pain by 5 and adding this to the half the sum of the total value recorded for ADLs. The FSS takes approximately 3 minutes for a patient to complete. The clinician can calculate the total FSS score in less than 1 minute.

The structure of the FSS was produced using an informal composite qualitative method combining patient opinion, experienced clinician opinion and observation of other validated shoulder scoring systems in use. Particular attention was paid to the questions used on the Oxford Shoulder Score¹⁵ and the Western Ontario Rotator Cuff index¹⁶ because the items on these particular scores were designed on the basis of rigorous qualitative methodology. The aim was to design a scoring system that was simple to use and understand, focused on symptoms that debilitated patient function the most and would be sensitive enough to detect significant clinical change. A score range of 0 (worst) to 100 (best) was chosen because it is intuitive, easy to remember and analogous to the percentage scoring system used in everyday life, such that an individual score would be instantly recognizable to a clinician as a good or poor result. Aligned with this, and to maximize the ability of the scale to detect small clinical differences without detracting from the speed of completion or calculation, scoring responses were graded on a scale of 0 to 10.

In the 8 years preceding the present study, the senior author (AI) had previously incorporated the use of the most popular shoulder specific scoring systems from the UK and North America at various times into a busy shoulder surgery practice as a means of auditing surgical results. The questions that appeared to be most discriminatory for the shoulder function of patients with rotator cuff disease were noted. Patient feedback was garnered on an often and anecdotal basis as to which questions they felt to be most relevant and important. Based on these factors, it was decided to incorporate 11 items into the questionnaire. Pain is widely acknowledged to be by far the most important assessment

PLEASE MARK ON THE SCALE FROM 0 TO 10

PAIN		0	1	2	3	4	5	6	7	8	9	10
How bad is your shoulder pain?		<div style="display: flex; justify-content: space-between;"> severe pain no pain </div>										
EVERYDAY ACTIVITIES												
Reach overhead	0	1	2	3	4	5	6	7	8	9	10	unable to do / can do normally
Comb / wash hair	0	1	2	3	4	5	6	7	8	9	10	unable to do / can do normally
Wash your opposite shoulder	0	1	2	3	4	5	6	7	8	9	10	unable to do / can do normally
Put jacket on	0	1	2	3	4	5	6	7	8	9	10	unable to do / can do normally
Reach your back pocket	0	1	2	3	4	5	6	7	8	9	10	unable to do / can do normally
Reach your midback	0	1	2	3	4	5	6	7	8	9	10	unable to do / can do normally
Sleep	0	1	2	3	4	5	6	7	8	9	10	unable to do / can do normally
Work activities	0	1	2	3	4	5	6	7	8	9	10	unable to do / can do normally
Leisure / sport activities	0	1	2	3	4	5	6	7	8	9	10	unable to do / can do normally
Lift objects at shoulder level	0	1	2	3	4	5	6	7	8	9	10	unable to do / can do normally

Figure 1. The Functional Shoulder Score (FSS). FSS Total Score (maximum 100) = (Pain Score × 5) + (Activities of Daily Living Score/2).

factor in several shoulder outcome measures.^{10,17} This was consistent with the experience from our clinic and hence the reported pain score on the FSS is multiplied by five to give a maximum score of 50. The remaining ten questions were those felt to be most relevant to activities of daily living specific to shoulder function in the specific subset of patients with rotator cuff disease.

The CMS is a mixed subjective and objective 100-point scoring system in which 35 points are derived from the patient's subjective report of pain (15 points) and questions pertaining to function of activities of daily living (20 points). The remaining 65 points are allocated by measurement of range of motion in forward flexion, abduction, external and internal rotation (40 points) as well as strength of resisted abduction in the plane of the scapula (25 points). In the present study, measurements of range of motion and strength were taken in accordance with the originally described method,¹⁰ which has been subsequently recommended by the European Society of Elbow Surgeons.¹⁸ To assess the strength component patients were asked to abduct their shoulder

in the scapular plane against a spring balance attached the forearm and anchored to the floor on five separate occasions. The mean recorded lift (kg) was calculated and compared with the mean recorded lift noted for the asymptomatic contralateral side. This modification of the CMS has been previously used and validated to allow a relative strength component to be calculated using the other shoulder as a comparator assumed to be normal for that individual.¹⁹

Separate to the items required for the FSS and CMS, participants were also required to complete an 'improvement scale' pertaining to overall change in the qualitative status of their shoulder after surgery (Figure 2). Patients could indicate whether their shoulder was 'much better', 'better', 'same', 'worse' or 'much worse'. This was included to allow patients to grade their impression of their own clinical status after surgery and, for purposes of statistical analysis, was considered to be the external criterion for a change to be clinically important.

Patients were consented to take part in a study comparing two shoulder outcome measures.

Ring the response that best indicates your shoulder's overall change after surgery:

Much worse	Worse	Same	Better	Much better
------------	-------	------	--------	-------------

Figure 2. Patient's improvement scale (IS).

A patient-completed questionnaire was administered to participants combining all elements of the FSS, the subjective aspect of the CMS and the satisfaction score (postoperatively only). Patients completed the questionnaire independently of clinician assistance. Pre-operatively the questionnaire was administered by a nurse in the pre-assessment clinic and this was followed by objective assessment of the CMS by an orthopaedic surgeon in training (CP). Postoperatively both subjective and objective components were administered by the trainee surgeon. This was following, and independent from, patient consultation with the senior operating surgeon at 6-month outpatient review. In accordance with our institutional policy, all patients attending clinic with poor grasp of the English language were accompanied by an interpreter who provided translation where required.

To assess test–retest reliability 40 patients were asked to complete the FSS at home immediately following initial assessment and return their completed questionnaire on the day of surgery 3 weeks later.

Statistical analysis

SPSS, version 13.0 (SPSS inc., Chicago, IL, USA) was used for statistical analysis. $p < 0.05$ was considered statistically significant.

Factor analysis. To confirm the usefulness of the structure of the 11-item FSS, a principal component factor analysis was conducted. Factor analysis is a statistical procedure that mathematically describes the correlations between a set of items arising from underlying common factors. Exploratory factor analysis (EFA) is useful to identify the most important variables when designing a collection of questions to measure a particular topic.²⁰ Here, it was used to identify the underlying relationships between measured items and their variance with respect to overall score. Factor loadings were calculated to indicate the strength of impact of each item on the overall score.

Reliability. Reliability is the overall consistency of a measure. A measure is said to have a high reliability if it produces similar results under consistent conditions. In the present study, we assessed reliability by determining internal consistency and test–retest reliability.

Internal consistency was determined by calculating Cronbach's alpha coefficient from all scores. A higher alpha value (range 0.0 to 1.0) indicates a consistent scale that measures a single underlying variable.²¹

Test–retest reliability was used to examine reproducibility over time and was assessed by comparing the answers of a subset of 40 patients at two distinct time points prior to surgery. These data were examined by calculating Pearson's correlation coefficient and Kendall Tau-b. Furthermore, relative reliability, the degree to which individuals maintain their opinion in a sample with repeated measurements, was evaluated using intraclass correlation coefficient (ICC) (i.e. the error in measurements as a proportion of the total variance). Because this coefficient does not correct for systematic differences and agreement by chance, the scores of the two assessments were also tested for systematic differences by using the Paired *t*-test.

Construct validity. The validity of an instrument concerns its ability to measure what it was intended to measure. The construct validity of the FSS was determined by establishing its correlation to the CMS using Pearson's correlation coefficient. We hypothesized that the patients' FSS would have moderate levels of correlation with the CMS and that this would be consistent with the levels of correlation seen in validation studies of other outcomes tools used in shoulder disorders.

Responsiveness to change. This refers to the ability of an instrument to reflect underlying change in patient status.⁷ Here, this was determined using the effect size (ES) and the standardized response mean (SRM) from FSS scores before and after surgery. An ES value greater than 1.0 and SRM greater than 0.8 indicates good responsiveness.^{22,23}

Cut-off point. To identify an FSS score that would represent an equivalent to the CMS 'cut-off' score of 70 between a 'good or excellent score' and a 'moderate score', the sensitivity and specificity of different cut-off points for the FSS were estimated from a receiver operating curve using only the postoperative CMS as a gold standard because all pre-operative CMS scores were < 70 . The areas under the receiver operating curve were calculated with standard error and 95% confidence interval using the maximum likelihood estimation method.

Interpretability. This refers to the degree to which one can assign qualitative meaning to quantitative scores.²⁴ This was represented using the standard error of measurement (SEM), the minimal detectable change (MDC) and the minimal clinically important difference (MCID). The SEM provides an estimate of how reliably a scale measures an individual's 'true score' at a single point in time and was calculated with a 90% confidence interval, using an equation based on the Cronbach's alpha coefficient. The MDC is defined as the minimum change of the FSS with which one can be confident that a 'true change' in the patient's clinical status has occurred²⁵ and is calculated by an equation based on the ICC of the test-retest sample.

The MCID is defined as the minimal change in the score that is considered to be worthwhile or important,²⁶ and is traditionally difficult to calculate.⁵ In this instance, the MCID was calculated by taking the mean change score of everyone who reported changing one increment on the qualitative improvement scale.^{27,28} Given that surgery is associated with a positive improvement after 6 months, the MCID can only be calculated by comparing the pre- and postoperative scores of those patients who indicated one positive incremental improvement on the scale (i.e. from 'same' to 'better').

Cook et al. have demonstrated that the error associated with the function subscale of self-reported shoulder scores could vary among different levels of functioning.²⁹ Therefore, we calculated SEM, MDC, and MCID for three different score ranges, 0 to 19, 20 to 39 and 40 to 59 on the overall FSS. We excluded patients with a pre-operative score of over 60 points as this accounted for only five patients and was thus not appropriate for statistical analysis.

Results

Of the patients invited to take part, FSS and CMS scores were collected for 119 patients. The study group consisted of 54 men and 65 women whose average age was 58 years at time of surgery (range 24 years

to 92 years; SD 11 years). The dominant arm was affected in 48 patients (40%). Pre-operative diagnosis was made after patient consultation and confirmed with shoulder imaging (either ultrasound or magnetic resonance imaging) where appropriate.

Table 1 presents the descriptive statistics obtained for the FSS and CMS pre- and postoperatively. Before operation mean FSS was 28.16 (range 0.00 to 83.50); after operation, this improved to 83.40 (range 46.00 to 100). Manuscript analysis revealed that no question remained unanswered.

Factor analysis

The Kaiser-Meyer-Olkin measure of sampling adequacy showed that the data were suitable for factor analysis. Exploratory factor analysis revealed that question 1 (regarding pain) accounted for 61.04% of the total variance (Table 2), identifying it as the most important distinguishing question. Factor loadings ranged from 0.59 to 0.89 and most items (80%) had loadings over 0.74.

Internal consistency

Cronbach's alpha for the FSS was 0.933, indicating that the items in the questionnaire measure the same construct with excellent internal consistency. All questions posted values higher than 0.9, suggesting that the items are interdependent and homogeneous in terms of the construct they measure (Table 3).

Test-retest reliability

A paired samples *t*-test between baseline and early follow-up administration of the FSS to a subsection of participants indicated no statistically significant differences (Table 4). Furthermore, Pearson's *r*, Kendall Tau-b and ICC coefficients between baseline and follow-up administration of the test ranged between 0.862 and 0.959 ($p < 0.0005$) (Table 4), indicating that FSS scores are appropriately consistent over a short

Table 1. Descriptive statistics pre- and postoperatively.

		Mean	SD	Minimum	Maximum
Pre-operatively	CMS	26.13	9.31	11.00	55.00
	FSS	28.16	18.07	0.00	83.50
Postoperatively	CMS	66.93	10.28	34.00	87.00
	FSS	83.40	11.41	46.00	100

CMS, Constant-Murley Score; FSS, Functional Shoulder Score.

time period during which it is assumed that there would be no clinical change.

Construct validity

The FSS demonstrated significant correlation with the CMS, providing evidence of its validation as a tool to measure shoulder pain and dysfunction. Table 5 summarizes this correlation as measured by Pearson's coefficient. The correlation coefficient between the FSS and CMS was 0.782 and 0.737 pre-operatively and post-operatively, respectively ($p < 0.0005$).

Responsiveness to change

The FSS demonstrated a high sensitivity to detect clinical change in patients after surgical intervention (Table 6). There was a statistically significant increase

in postoperative score ($p < 0.0005$). The measured ES and standardised response mean of the FSS (3.06 and 2.80, respectively) were far in excess of the minimum values taken to indicate good responsiveness (Table 7).

Cut-off point

If the postoperative CMS is taken as the gold standard (cut-off point = 70), a relatively high cut-off point of 86.25 was identified for the FSS, using a receiver operating curve (Figure 3). This value is taken as the boundary between a good score and a moderate score with calculated sensitivity of 74% and specificity of 72%. This means that patients with a score of more than 86.25 have a 74% probability of a true good or excellent result, while patients with a score less than 86.25

Table 2. Eigenvalues and explained variance.

Items	Eigenvalues	% of variance	Cumulative %
1	6.71	61.04	61.04
2	0.89	8.12	69.16
3	0.80	7.30	76.47
4	0.72	6.56	83.03
5	0.44	4.01	87.05
6	0.36	3.31	90.36
7	0.32	2.96	93.32
8	0.24	2.23	95.56
9	0.19	1.79	97.35
10	0.16	1.47	98.82
11	0.12	1.17	100.00

Table 3. Internal consistency of Functional Shoulder Score questionnaire (alpha = Cronbach's α value).

Items	Alpha if item deleted
1	0.940
2	0.920
3	0.924
4	0.922
5	0.921
6	0.926
7	0.924
8	0.936
9	0.929
10	0.926
11	0.926
Overall alpha = 0.933	

Table 4. Test-retest reliability for the Functional Shoulder Score (FSS).

Subscales (n = 40)	Pearson's correlation coefficients	Kendall Tau-b	ICC 95% CI	Paired samples t-test		
					Mean $\pm$ SD	P -value
FSS total	0.959*	0.871*	0.959* (0.92-0.98)	A	49.86 $\pm$ 19.5	0.781
				B	49.61 $\pm$ 19.9	

*All correlation coefficients are statistically significant ($p < 0.0005$).

A, baseline; B, follow-up. CI, confidence interval; ICC, intraclass correlation coefficient.

have a 72% probability of a true moderate or poor result (Table 8).

Interpretability

The SEM associated for the FSS was ± 7.85 points [based on a 90% confidence interval (CI)]. The MDC at the 90% confidence level was 8.38 points, indicating that if a patient's score falls outside of this range from visit to visit it is likely that a true change in score has occurred. In practical terms this means that if a patient scores 60 at first visit then only a score outside of the range 51 to 69 points would confidently indicate a real score change.

The overall MCID for the FSS was 24.7 ± 5.4 scale points. Given the overwhelming improvement in patient scores 6 months postoperatively, it was only

appropriate to calculate the MCID for those patients who indicated that they were 'better' after surgical intervention. Therefore the MCID value presented here can only be used to interpret improvement rather than deterioration. Some variation was noted for SEM, MDC, and MCID for the various pre-operative score ranges and the overall results regarding interpretability are presented in Table 7.

Discussion

The FSS is a short 11-item questionnaire that we recommend is used as a scoring system for patients with shoulder pain after diagnosis of rotator cuff dysfunction. It is intended to be of use to clinicians who wish to compare their patients' shoulder symptoms over time and after intervention. It is easy to use in the busy clinical environment and provides a quick numerical assessment of shoulder function.

The reliability of the FSS is one of its major strengths. The posted Cronbach's alpha value of 0.93 indicates that the items within the scale measure the same construct with an excellent level of homogeneity. For clinical application, Cronbach's alpha values between 0.90 and 0.95 are most desirable.³⁰ The FSS also demonstrates excellent test-retest reliability, with an ICC of 0.93. This compares well with other shoulder outcome measures with ICC values reported to range between 0.64 and 0.99.^{11,22,31,32} The intention was for our patient group to be retested within a 24-hour interval (although scores were not collected until 3 weeks later). We consider that our results would not be different if we had used longer intervals, as other studies have documented.^{33,34}

Table 5. Correlation between pre- and postoperative Functional Shoulder Score (FSS) and Constant-Murley Score.

		Pearson's (r)	p-value
FSS	Pre-operatively	0.782	< 0.0005
	Postoperatively	0.737	< 0.0005

Table 6. Comparison of Functional Shoulder Score (FSS) pre- and postoperatively.

	Pre-operative mean $\pm$ SD	Postoperative mean $\pm$ SD	p-value
FSS	28.13 $\pm$ 12.07	83.40 $\pm$ 11.42	<0.0005

Table 7. Standard error of measurement (SEM), minimal detectable change (MDC), minimal clinically important difference (MCID), effect size (ES) and standardized response mean (SRM) of the Functional Shoulder Score (FSS) for various score ranges and overall score.

	FSS pre-operatively			All patients
	0 to 19 (n = 39)	20 to 39 (n = 51)	40 to 59 (n = 24)*	
SEM	4.78	4.15	7.47	7.85
MDC	5.06	4.48	8.05	8.38
ES	10.85	9.31	3.14	3.06
SRM	5.31	3.67	3.25	2.80
MCID†	29.2 $\pm$ 4.4 (n = 14)	22.5 $\pm$ 4.4 (n = 10)	—	24.7 $\pm$ 5.4 (n = 24)

*Excluded five patients with pre-operative score > 60.

†Calculated only for those patients who reported becoming 'better' on the improvement scale.

SEM and MDC were calculated based on a 90% confidence interval.

Construct validity of the FSS was supported by good correlation with the CMS when administered in both pre-operative and postoperative phases (Pearson's coefficient 0.782 and 0.737, respectively). A high positive correlation coefficient is desired when the scales examined are, in theory, similar in nature.³⁵ The advantage of the FSS over the CMS is its ease and speed of use without the need for objective assessment from an experienced clinician.

The present study showed that the FSS is exceptionally sensitive to clinical change, expressed as a high ES of 3.06 and a high SRM of 2.80. Values greater than 1.0 indicate good responsiveness.^{22,23,36} The FSS appears to be more responsive than other shoulder questionnaires whose reported values range from 0.59 to 1.54.^{22,2,37,38} However, this cannot be inferred with complete confidence until a direct comparison is made on the same group of patients on a large scale.

Our results indicate that the SEM of the FSS was ± 7.85 scale points (based on a 90% CI). A clinician can therefore be at least 90% confident that the true value for a patient with FSS score reported at 60 points actually falls within a range of 52 to 68 points. The MDC at

the 90% confidence level was ± 8.38 scale points. This means that, if the same patient scores 70 points during the reassessment, the clinician can be reasonably confident that the patient has demonstrated true improvement because the change of 10 points is greater than the MDC value of 8.38. We calculated the SEM and MDC of the FSS at three different score ranges. Our results were consistent with those from previous studies in that the error was lower for mid-range scores.²⁸ Ceiling and floor effects do exist when patients score at the extremes of an outcome measure.^{28,39} A ceiling effect refers to a restricted range for improvement because patients begin at the highest level on the scale. Conversely, in the case of a floor effect, there is a restricted range for deterioration. In the present study, seven patients scored 2 points or below and one patient scored 83.5 pre-operatively. Because the MDC of the FSS is ± 8.38 , it appears that there is potential for both ceiling and floor effects to exist for a small number of individuals.

It is generally assumed that small differences in the scores of self-reported outcome tools may be statistically significant yet clinically unimportant.⁴⁰ Therefore, outcomes research is faced with the challenge of interpretability of the scores.^{27,41} The concept of the MCID has been proposed and refers to the smallest difference in a score that is considered to be worthwhile or important.⁴⁰ We were only able to calculate the MCID for those patients who had become at least 'better' after surgery. However, only four of the 119 patients (3%) reported that they were 'worse' or 'much worse'. The MCID for the FSS was 24.7 points and was higher than the MDC. This indicates that the amount of change a patient perceives as important is higher than the amount considered to be statistically significant. One possible explanation for this is that the majority of our patients did extremely well after surgery. They had very low pre-operative scores, post-operatively they were much more satisfied with only 24 patients reported being 'the same' after surgery. A limitation of these results is the relatively small sample size available for MCID calculation. A larger sample size of patients may yield a different MCID. Many studies have shown that retrospective report of change is associated with a larger prospective change for those with more room for change.^{40,42}

Figure 3. Receiver operating curve analysis used to calculate the best Functional Shoulder Score (FSS) cut-off point. Areas under the curve were calculated using the maximum likelihood estimation method.

Table 8. The cut-off point between good and moderate score for the Functional Shoulder Score (FSS).

	Area	SE	Significance	Cut-off Point	Sensitivity	Specificity	95% confidence interval	
FSS	0.805	0.039	0.0005	86.25	74%	72%	0.74	0.88

A smaller test result indicates a more positive test.

We used the patient's global rating of change (improvement scale) as the criterion for change in the present study. Many studies have used the same or a similar scale as the standard for change in their studies to assess functional status measures.^{22,32,37,43} However, the use of a global rating of change has been questioned.^{44,45} The reliability and validity of this global rating of change have not been established and a patient's recall of his previous condition may be inaccurate or biased.^{22,44–46} Because patient-reported questionnaires and the global rating of change are measures that involve a patient's judgment, the errors of the global rating of change and a self-assessment scale are most likely correlated potentially making the global rating a biased measure of change.^{22,45} However, several studies have provided evidence that the use of a global rating of change has the ability to differentiate changes in clinical status over time.^{22,32,37,43}

A myriad of instruments exist to measure shoulder symptoms and function. Schmidt et al. recently published the first standardized expert-based evaluation of shoulder PROMs with regard to development process, metric properties and administration issues using the systematic EMPRO (Evaluating Measures of Patient Reported Outcomes) tool.⁴⁷ Eleven instruments and 112 articles were included. They found the American Shoulder and Elbow Surgeons shoulder assessment – patient self-evaluation section (ASES-p), Simple Shoulder Test (SST) and the Oxford Shoulder Score (OSS) to be the best rated.

Table 9 compares the content and metric data pertaining to the FSS, ASES-p, SST and OSS. The FSS has comparable internal consistency, test–retest reliability and construct validity with the best rated PROMs.

According to our results, it has superior responsiveness compared to published results specific to rotator cuff disease. This can be explained by considering that the FSS has been particularly designed for use in patients rotator cuff disorders, whereas the other instruments are generic for most shoulder pathology including osteoarthritis. Further, the ASES includes two questions expressly pertaining to instability symptoms.

The present study indicated that the FSS has the best value for MDC but the lowest value for MCID. As explained earlier, we consider this to be related to the small number of patients reporting a one category up-shift on the improvement scale post-surgery. Furthermore, it is difficult to compare the MCID across studies because the 'improvement scale' used tends to vary greatly with regard to wording and number of items. It is worth noting that the SST has an extremely high MDC value most likely related to its binary yes/no question format and narrow range of scores possible. The FSS has the least items and uses an easily recognizable positively-scored 100-point scale, which we consider to be advantageous for the busy clinician especially when compared to the OSS.

We acknowledge that assessment of construct validity could have been strengthened by comparing the FSS with other validated shoulder-specific scores, as well as scores pertaining to more general disability (e.g. Short Form-36 and Euroqol-5D). However the similarity of Pearson correlation coefficients obtained in the present study compared to values obtained in previous studies comparing the SST,⁴⁸ ASES,⁴ and OSS^{49–51} with the CMS provides surrogate reassurance on this point.

Table 9. Comparison of item number, scale and published metric data for the Functional Shoulder Score (FSS), shoulder assessment – self-evaluation section (ASES), Simple Shoulder Test (SST) and Oxford Shoulder Score (OSS).

PROM	Items	Score range	Reliability		Construct validity	Interpretability (100-point scale)		Responsiveness for cuff disease	
			Cronbach's α	ICC	Pearson's (versus CMS)	MDC	MCID	ES	SRM
FSS	11	0 to 100	0.93	0.96	0.74 to 0.78	8.38	24.7	3.06	2.80
ASES	28	0 to 100	0.61 to 0.96 ^{22,53–56}	0.84 to 0.96 ^{22,53–55}	0.71 ⁴⁸	11.2 ²²	16.9 ⁵⁷	–	1.42 ⁵⁸
SST	12	0 to 12	0.85 ⁵⁹	0.97 ³²	0.70 ⁴	32.3 ⁵⁹	19.4 ⁵⁷	–	1.09 ⁶⁰
OSS	12	60 to 12	0.94 ⁵⁰	–	0.65 to 0.87 ^{49–51}	12.5 ⁶¹	12.5 ⁶¹	1.10 to 1.88 ^{15,49,62}	1.10 to 1.14 ^{15,49,62}

PROM, Patient-reported outcome measures; CMS, Constant–Murley Score; MDC, minimal detectable change; MCID, minimal clinically important difference; ES, effect size; SRM, standardized response mean.

No formal qualitative methodology was used in the design phase of the FSS, and this is a common feature with the design of the majority of shoulder scores.⁵² However, principal component factor analysis (Table 2) indicated that all 11 items accounted for significant variance in answer and lends justification to the designated score-weighting of 50 points for the pain item. Although the wording of the questionnaire is deliberately simplistic, and no question remained unanswered, it may have been useful to garner documented formal patient feedback on questionnaire design to refine this before the study started.

At present, the FSS is potentially limited by several factors. We have made no attempt to validate its use in shoulder pathology other than cuff disease and therefore it should be considered as a disease-specific score. We do not recommend its use as a blanket score for all shoulder disease. Accurate use will depend on accurate diagnosis and this may not always be possible on first consultation. Multiple conditions may mimic rotator cuff disease and there is currently no 'gold standard' outpatient test for diagnosis. There may have been patients included in the present study in whom the primary pathology was not rotator cuff disease; however, the overwhelmingly positive improvement following surgery directed towards the rotator cuff on both FSS and CMS would likely suggest accurate diagnosis in the majority. In the clinical environment, a lack of improvement in FSS following surgical intervention should alert a clinician to question the diagnosis. Furthermore, the FSS is only validated to detect change following rotator cuff surgery, as opposed to other interventions such as physiotherapy and steroid injection. However, it should be noted that previous studies have tended to find that SRM and ES values for PROMs tend to be similar across a range of treatments for cuff disease.⁵² Further studies to assess the responsiveness and interpretability of the FSS for other interventions are encouraged. Finally, although our institution is situated in a large metropolitan area with a multicultural population, the FSS requires further validation for cross-cultural use and multilingual adaptation.

Conclusions

The present study has provided evidence for the use of FSS as a shoulder specific outcome measure for patients with rotator cuff disorders. The FSS is practical and easy to administer. It is reliable, valid and highly responsive to clinical change, providing surgeons with comparable data for clinical audit and research studies.

Acknowledgements

The data included in this manuscript were previously presented at the European Society of Shoulder and Surgeons (ESSE-SECEC) annual meeting in Lyon 2011

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.

Author contributions

AI was responsible for concept design, data collection, approval and critical revision. EFI was responsible for drafting the article and critical revision. CP was responsible for data collection. AG was responsible for data analysis and interpretation.

References

1. Hegmann KT and Moore JS. Common neuromusculoskeletal disorders. *Sourcebook of Occupational Rehabilitation*. Philadelphia, USA: Springer, 1998, pp.19–41.
2. Oh LS, Wolf BR, Hall MP, Levy BA and Marx RG. Indications for rotator cuff repair: a systematic review. *Clin Orthop Relat Res* 2007; 455: 52–63.
3. Bot S, Terwee C, Van der Windt D, Bouter L, Dekker J and De Vet H. Clinimetric evaluation of shoulder disability questionnaires: a systematic review of the literature. *Ann Rheum Dis* 2004; 63: 335–41.
4. Romeo AA, Mazzocca A, Hang DW, Shott S and Bach Jr BR. Shoulder scoring scales for the evaluation of rotator cuff repair. *Clin Orthop Relat Res* 2004; 427: 107–14.
5. Roller AS, Mounts RA, DeLong JM and Hanypsiak BT. Outcome instruments for the shoulder. *Arthroscopy* 2013; 29: 955–64.
6. Beaton DE and Richards RR. Measuring function of the shoulder. A cross-sectional comparison of five questionnaires. *J Bone Joint Surg* 1996; 78: 882–90.
7. Jackowski D and Guyatt G. A guide to health measurement. *Clin Orthop Relat Res* 2003; 413: 80–9.
8. Wright RW and Baumgarten KM. Shoulder outcomes measures. *J Am Acad Orthop Surgeons* 2010; 18: 436–44.
9. Pynsent PB, Fairbank JCT and Carr A. *Outcome Measures in Orthopaedics*. Oxford, UK: Butterworth-Heinemann, 1993.
10. Constant C and Murley A. A clinical method of functional assessment of the shoulder. *Clin Orthop Relat Res* 1987; 214: 160–4.
11. Conboy VB, Morris RW, Kiss J and Carr AJ. An evaluation of the Constant–Murley shoulder assessment. *J Bone Joint Surg* 1996; 78: 229–32.
12. Johansson KM and Adolfsson LE. Intraobserver and interobserver reliability for the strength test in the

- Constant-Murley shoulder assessment. *J Shoulder Elbow Surg* 2005; 14: 273–8.
13. Rees J, Dawson J, Hand G, et al. The use of patient-reported outcome measures and patient satisfaction ratings to assess outcome in hemiarthroplasty of the shoulder. *J Bone Joint Surg* 2010; 92: 1107–11.
 14. Longo UG, Vasta S, Maffulli N and Denaro V. Scoring systems for the functional assessment of patients with rotator cuff pathology. *Sports Med Arthrosc Rev* 2011; 19: 310–20.
 15. 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.
 16. Kirkley A, Alvarez C and Griffin S. The development and evaluation of a disease-specific quality-of-life questionnaire for disorders of the rotator cuff: The Western Ontario Rotator Cuff Index. *Clin J Sport Med* 2003; 13: 84–92.
 17. Richards RR, An K-N, Bigliani LU, et al. A standardized method for the assessment of shoulder function. *J Shoulder Elbow Surg* 1994; 3: 347–52.
 18. Constant C. *Constant Scoring Technique for Shoulder Function*. SECEC-ESSE, Lyon, France: SECEC Information number 3.
 19. Fialka C, Oberleitner G, Stampfl P, Brannath W, Hexel M and Vecsei V. Modification of the Constant–Murley shoulder score-introduction of the individual relative Constant score Individual shoulder assessment. *Injury* 2005; 36: 1159–65.
 20. Harman HH. *Modern Factor Analysis*. Chicago, IL: University of Chicago Press, 1976.
 21. Cronbach LJ. Coefficient alpha and the internal structure of tests. *Psychometrika* 1951; 16: 297–334.
 22. Michener LA, McClure PW and 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: 587–94.
 23. Portney L and Watkins M. *Foundations of Clinical Research: Applications to Practice*. East Norwalk, CT: Appleton & Lange, 1993, p.148.
 24. 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: 979–92.
 25. Leggin BG, Michener LA, Shaffer MA, Brenneman SK, Iannotti JP and Williams Jr GR. The Penn shoulder score: reliability and validity. *J Orthop Sports Phys Ther* 2006; 36: 138–51.
 26. Kovacs FM, Abaira V, Royuela A, et al. Minimal clinically important change for pain intensity and disability in patients with nonspecific low back pain. *Spine* 2007; 32: 2915–20.
 27. Beaton DE, Boers M and Wells GA. Many faces of the minimal clinically important difference (MCID): a literature review and directions for future research. *Curr Opin Rheumatol* 2002; 14: 109–14.
 28. Jaeschke R, Singer J and Guyatt GH. Measurement of health status: ascertaining the minimal clinically important difference. *Control Clin Trials* 1989; 10: 407–15.
 29. Cook KF, Gartsman GM, Roddey TS and Olson SL. The measurement level and trait-specific reliability of 4 scales of shoulder functioning: an empiric investigation. *Arch Phys Med Rehabil* 2001; 82: 1558–65.
 30. Bland JM and Altman DG. Statistics notes: Cronbach's alpha. *BMJ* 1997; 314: 572.
 31. Roach KE, Budiman-Mak E, Songsiridej N and Lertratanakul Y. Development of a shoulder pain and disability index. *Arthritis Rheum* 1991; 4: 143–9.
 32. Beaton D and Richards RR. Assessing the reliability and responsiveness of 5 shoulder questionnaires. *J Shoulder Elbow Surg* 1998; 7: 565–72.
 33. Marx RG, Menezes A, Horovitz L, Jones EC and Warren RF. A comparison of two time intervals for test-retest reliability of health status instruments. *J Clin Epidemiol* 2003; 56: 730–5.
 34. Wiesinger GF, Nuhr M, Quittan M, Ebenbichler G, Wölfl G and Fialka-Moser V. Cross-cultural adaptation of the Roland-Morris questionnaire for German-speaking patients with low back pain. *Spine* 1999; 24: 1099–103.
 35. Michener LA and Leggin BG. A review of self-report scales for the assessment of functional limitation and disability of the shoulder. *J Hand Ther* 2001; 14: 68–76.
 36. Stucki G, Liang MH, Fossel AH and Katz JN. Relative responsiveness of condition-specific and generic health status measures in degenerative lumbar spinal stenosis. *J Clin Epidemiol* 1995; 48: 1369–78.
 37. Heald SL, Riddle DL and Lamb RL. The shoulder pain and disability index: the construct validity and responsiveness of a region-specific disability measure. *Phys Ther* 1997; 77: 1079–89.
 38. L'Insalata JC, Warren RF, Cohen SB, Altchek DW and Peterson MG. A Self-administered questionnaire for assessment of symptoms and function of the shoulder. *J Bone Joint Surg* 1997; 79: 738–48.
 39. Binkley JM, Stratford PW, Lott SA and Riddle DL. The Lower Extremity Functional Scale (LEFS): scale development, measurement properties, and clinical application. *Phys Ther* 1999; 79: 371–83.
 40. Hays RD and Woolley JM. The concept of clinically meaningful difference in health-related quality-of-life research. *Pharmacoeconomics* 2000; 18: 419–23.
 41. Testa MA. Interpretation of quality-of-life outcomes: issues that affect magnitude and meaning. *Medical Care* 2000; 38: II-166–74.
 42. Baker DW, Hays RD and Brook RH. Understanding changes in health status: is the floor phenomenon merely the last step of the staircase? *Medical Care* 1997; 35: 1–15.
 43. Gummesson C, Atroshi I and 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 Musculoskelet Disord* 2003; 4: 11.
 44. Norman GR, Stratford P and Regehr G. Methodological problems in the retrospective computation of responsiveness to change: the lesson of Cronbach. *J Clin Epidemiol* 1997; 50: 869–79.

45. Stratford PW, Binkley JM and Riddle DL. Health status measures: strategies and analytic methods for assessing change scores. *Phys Ther* 1996; 76: 1109–23.
46. Fritz JM and Irrgang JJ. A comparison of a modified Oswestry low back pain disability questionnaire and the Quebec back pain disability scale. *Phys Ther* 2001; 81: 776–88.
47. Schmidt S, Ferrer M, Gonzalez M, et al. Evaluation of shoulder-specific patient-reported outcome measures: a systematic and standardized comparison of available evidence. *J Shoulder Elbow Surg* 2014; 23: 434–44.
48. Angst F, Pap G, Mannion AF, et al. Comprehensive assessment of clinical outcome and quality of life after total shoulder arthroplasty: usefulness and validity of subjective outcome measures. *Arthritis Rheum* 2004; 51: 819–28.
49. Dawson J, Hill G, Fitzpatrick R and Carr A. The benefits of using patient-based methods of assessment. Medium-term results of an observational study of shoulder surgery. *J Bone Joint Surg* 2001; 83: 877–82.
50. Huber W, Hofstaetter JG, Hanslik-Schnabel B, Posch M and Wurnig C. The German version of the Oxford Shoulder Score – cross-cultural adaptation and validation. *Arch Orthop Trauma Surg* 2004; 124: 531–6.
51. Christie A, Hagen KB, Mowinckel P and Dagfinrud H. Methodological properties of six shoulder disability measures in patients with rheumatic diseases referred for shoulder surgery. *J Shoulder Elbow Surg* 2009; 18: 89–95.
52. Angst F, Schwyzer HK, Aeschlimann A, Simmen BR and Goldhahn J. Measures of adult shoulder function: Disabilities of the Arm, Shoulder, and Hand Questionnaire (DASH) and its short version (QuickDASH), Shoulder Pain and Disability Index (SPADI), American Shoulder and Elbow Surgeons (ASES) Society standardized shoulder assessment form, Constant (Murley) Score (CS), Simple Shoulder Test (SST), Oxford Shoulder Score (OSS), Shoulder Disability Questionnaire (SDQ), and Western Ontario Shoulder Instability Index (WOSI). *Arthritis Care Res* 2011; 63: S174–88.
53. Padua R, Padua L, Ceccarelli E, Bondi R, Alvi F and Castagna A. Italian version of ASES questionnaire for shoulder assessment: cross-cultural adaptation and validation. *Musculoskelet Surg* 2010; 94: S85–90.
54. Kocher MS, Horan MP, Briggs KK, Richardson TR, O'Holleran J and Hawkins RJ. 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* 2005; 87: 2006–11.
55. Goldhahn J, Angst F, Drerup S, Pap G, Simmen BR and Mannion AF. Lessons learned during the cross-cultural adaptation of the American Shoulder and Elbow Surgeons shoulder form into German. *J Shoulder Elbow Surg* 2008; 17: 248–54.
56. Cook KF, Roddey TS, Olson SL, Gartsman GM, Valenzuela FF and Hanten WP. Reliability by surgical status of self-reported outcomes in patients who have shoulder pathologies. *J Orthop Sports Phys Ther* 2002; 32: 336–46.
57. Tashjian RZ, Deloach J, Green A, Porucznik CA and Powell AP. Minimal clinically important differences in ASES and simple shoulder test scores after nonoperative treatment of rotator cuff disease. *J Bone Joint Surg* 2010; 92: 296–303.
58. Razmjou H, Bean A, van Osnabrugge V, MacDermid JC and Holtby R. Cross-sectional and longitudinal construct validity of two rotator cuff disease-specific outcome measures. *BMC Musculoskelet Disord* 2006; 7: 26.
59. Roddey TS, Olson SL, Cook KF, Gartsman GM and Hanten W. 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: 759–68.
60. MacDermid JC, Drosdowech D and Faber K. Responsiveness of self-report scales in patients recovering from rotator cuff surgery. *J Shoulder Elbow Surg* 2006; 15: 407–14.
61. van Kampen DA, Willems WJ, van Beers LW, Castelein RM, Scholtes VA and Terwee CB. Determination and comparison of the smallest detectable change (SDC) and the minimal important change (MIC) of four-shoulder patient-reported outcome measures (PROMs). *J Orthop Surg Res* 2013; 8: 40.
62. Dawson J, Hill G, Fitzpatrick R and Carr A. Comparison of clinical and patient-based measures to assess medium-term outcomes following shoulder surgery for disorders of the rotator cuff. *Arthritis Rheum* 2002; 47: 513–19.