The benefits of using patient-based methods of assessment

MEDIUM-TERM RESULTS OF AN OBSERVATIONAL STUDY OF SHOULDER SURGERY

J. Dawson, G. Hill, R. Fitzpatrick, A. Carr

From the Institute of Health Sciences, University of Oxford and the Nuffield Orthopaedic Centre, Oxford, England

We carried out a prospective study of 93 patients undergoing surgery for conditions of the rotator cuff during 1994 and 1995. They were assessed before operation and after six months, and four years, using the patient-based Oxford Shoulder Score (OSS), the SF-36 questionnaire and the Constant shoulder score. The response rates were higher for the OSS and SF-36. The correlation coefficients were high ($r > 0.5$) between all scores at each stage of the study. While all scores improved substantially at six months, the Constant score was reduced significantly at four years. This did not correlate with the patients' judgement of the change in symptoms or of the success of the operation. Our study suggests that patient-based measures of pain and function can reliably assess outcomes in the medium term after surgery to the shoulder.

J Bone Joint Surg [Br] 2001;83-B:877-82.

Received 8 May 2000; Accepted after revision 21 November 2000

There is little reliable information about the medium- and long-term results after many forms of surgery to the shoulder.¹ This is mainly because of the lack of appropriate methods of assessing outcome.

Patients can provide valid judgements about the benefits of treatment² which are independent of clinical assessment.³⁻⁶ There are two main types of patient-based measure: generic instruments which may be applied to a wide range of health problems, and condition-specific instruments, which are designed to address a single condition.

We have therefore compared a standard clinical assessment and two validated patient-based measures, one specific and one generic.

Patients and Methods

We carried out a prospective study of 93 consecutive patients who had been treated surgically between April 1994 and August 1995 for shoulder pain, with a provisional preoperative diagnosis of subacromial impingement, including 'calcified tendinitis', 'painful arc', partial or full-thickness tear of the rotator cuff, or any combination of these. All patients had received previous treatment consisting of intra-articular injections of corticosteroids, physiotherapy or both. The indication for surgery was that these measures had failed to alleviate symptoms. Patients with recurrent shoulder instability were excluded since they require different forms of assessment.⁷

Details of the surgical assessment were obtained from the surgeon, by questionnaire at the postoperative review after six months. The operations undertaken included arthroscopic or open subacromial decompression with or without partial or complete repair of tears of the rotator cuff.

None of the patients refused to participate. They completed a variety of assessments at the preadmission clinic, which were repeated after six months and four years later. One patient, who had not been assessed at six months, underwent an assessment at four years. Some patients who were unable (or refused) to attend for postoperative review completed their assessment by questionnaire.

Constant shoulder score.⁸ This contains subjective assessments of pain and activities of daily living which are allocated a maximum of 15 and 20 points, respectively, out of a total of 100. In addition, objective measurements of the active range of movement and power are allocated a maximum of 40 and 25 points, respectively. Low scores denote significant pain and poor function. The score was assessed by one surgeon (AC) preoperatively and by another (GH) at six months.

Oxford shoulder score (OSS).⁹ This is a condition-specific questionnaire which is completed by patients unaided. It contains 12 items, each of which has five categories of response. Scores are added to give a single score,

J. Dawson, DPhil, Research Fellow
R. Fitzpatrick, PhD, Professor
Division of Public Health and Primary Care, Department of Public Health,
Institute of Health Sciences, Old Road, Oxford OX3 7LF, UK.

G. Hill, FRCS Orth, Shoulder Fellow
A. Carr, FRCS, Professor of Orthopaedic Surgery
Upper Limb Unit, Nuffield Orthopaedic Centre, Windmill Road, Headington, Oxford OX3 7LD, UK.

Correspondence should be sent to Dr J. Dawson at the Oxford Centre for Health Care Research and Development, School of Health Care, Oxford Brookes University, 44 London Road, Headington, Oxford OX3 7PD, UK.

©2001 British Editorial Society of Bone and Joint Surgery
0301-620X/01/611316 \$2.00

with a range from 12 (best) to 60 (worse). It has been shown to be consistent, reproducible, valid and sensitive to clinical change at six months after surgery. Neither the Constant shoulder score nor the OSS is appropriate for assessing patients with instability.³

SF-36.¹⁰ This is a 36-item questionnaire, which is widely used to assess general health. It provides scores on eight dimensions: physical function, social function, limitations caused by physical symptoms, those caused by emotional problems, general mental health, vitality, pain, and perception of general health. A further item measures change by asking patients to compare their current health with that of 'one year ago'. Scores for each dimension range from 0 (poor health) to 100 (good health).

To summarise the different scoring systems, the OSS is scored in the reverse direction from that of the Constant and SF-36, and has a smaller range of possible scores. Thus, a low score for the OSS denotes fewer symptoms and good function, while a low score for the Constant and SF-36 denotes poor health and function. When a patient or surgeon occasionally missed out an item in the questionnaire, in all these methods of assessment the dimension including that item was not computed. This explains some of the slight variation in the numbers reported in the tables.

In addition to these measures, at both stages of follow-up, an item in the questionnaire asked patients to assess the success of their surgery, and offered four possible categories of response, ranging from 'very successful' to 'very disappointing'. Two 'transition' items invited patients to judge the degree of change in the symptoms arising from their shoulder and day-to-day life after surgery offering five possible responses: 'much better', 'slightly better', 'no change', 'slightly worse' or 'much worse'. The questionnaire at four years also included a question about the length of time during which the shoulder had been satisfactory since surgery.

The analysis included the calculation of the change score, which is the individual's preoperative score minus their score at six months or four years after surgery, for the Constant, OSS and SF-36 scores to assess the change in

clinical condition. The mean of the difference in these scores is presented when comparing the association of health status scores with categorical variables, such as grouped responses to questions about change.

We used the Spearman correlation coefficient to compare health status scores at each assessment and *t*-tests to test for statistical significance when comparing these scores for different groups of patients. While the distribution of absolute scores tends to be quite skewed, change scores approximate to normality. Paired *t*-tests have been used to compare differences in change scores with the passage of time. The 1% level ($p = 0.01$) was chosen to denote statistical significance in order to take account of multiple testing, i.e. the risk of obtaining a significant result on the basis of chance alone.

Results

At the preoperative assessment patients received a provisional diagnosis based on examination, and aided by radiography and/or ultrasonography whenever possible. Of the 93 patients, 52 (56%) were diagnosed as having subacromial impingement with or without a partial tear of the rotator cuff and 41 as having a full-thickness tear of the rotator cuff, many of whom also had signs of impingement. At the time of operation there were 61 men and 31 women with a mean age of 57.8 years (22.3 to 83.7); data were not available for one patient.

The most recent follow-up was at a mean of 3.9 years (2.2 to 5.6) after surgery, at which time one patient had died, six (7%) had been lost to follow-up, four (4%) had refused to participate, 82 (88%) had completed the patient-based questionnaire and 61 (66%) had attended for complete clinical assessment.

Pre- and postoperative health status scores. Table I shows the correlation coefficients for the preoperative and postoperative scores at six months and four years. Agreement was particularly high, and significant, between the OSS and Constant scores. The level of agreement in absolute scores increased with the passage of time. There was also agreement between the OSS and relevant dimensions

Table I. Correlation between the OSS and both the CSA and SF-36 scores, preoperatively and at six months and at four years after surgery

Method of assessment	Dimension	Preoperative		Six-month follow-up		Four-year follow-up	
		Spearman's correlation coefficient (r)	p value	Spearman's correlation coefficient (r)	p value	Spearman's correlation coefficient (r)	p value
CSA		-0.65	<0.001	-0.65	<0.001	-0.87	<0.001
SF-36	Physical functioning	-0.57	<0.001	-0.68	<0.001	-0.57	<0.001
	Pain	-0.64	<0.001	-0.75	<0.001	-0.71	<0.001
	Mental health	-0.45	<0.001	-0.52	<0.001	-0.51	<0.001
	Social functioning	-0.49	<0.001	-0.63	<0.001	-0.52	<0.001
	Role limitations – physical	-0.38	<0.001	-0.65	<0.001	-0.61	<0.001
	Role limitations – emotional	-0.44	<0.001	-0.54	<0.001	-0.37	0.001
	Vitality	-0.40	<0.001	-0.48	<0.001	-0.45	<0.001
	General health	-0.23	0.027	-0.33	0.002	-0.43	<0.001

Table II. Comparison between mean scores and mean change scores for the CSA, the OSS and the SF-36 health survey obtained preoperatively, at six months and at four years after surgery

Method of assessment	Preoperative scores (95% CI)	Six-month follow-up postoperative (95% CI)	Four-year follow-up postoperative (95% CI)	Mean change score (95% CI) Preoperative to six months	Mean change score (95% CI) Preoperative to four years
CSS	n = 93 51.7 (47.4 to 56.0)	n = 84 86.0 (81.9 to 90.2)	n = 60 74.0 (69.1 to 78.9)	n = 84 -33.8 (-39.0 to -28.6)	n = 60 -19.3 (-25.0 to -13.5)
OSS	n = 92 33.0 (31.3 to 34.8)	n = 88 22.1 (20.2 to 24.1)	n = 78 21.8 (19.4 to 24.2)	n = 87 10.8 (9.0 to 12.7)	n = 77 11.2 (9.0 to 13.4)
SF-36 dimensions	n = 90	n = 89	n = 75	n = 86	n = 78
Physical functioning	63.3 (58.7 to 67.8)	70.0 (64.5 to 73.1)	66.8 (60.5 to 73.1)	-7.5 (-11.3 to -3.7)	-1.6 (-7.9 to 4.8)
Pain	40.9 (36.9 to 44.8)	61.0 (55.5 to 66.6)	61.2 (55.2 to 67.2)	-20.6 (-26.2 to -15.0)	-19.6 (-25.9 to -13.4)
Mental health	71.8 (67.9 to 75.8)	74.0 (69.9 to 78.2)	75.8 (71.9 to 79.7)	-2.4 (-6.0 to 1.2)	-3.1 (-7.3 to 1.2)
Social functioning	75.3 (69.6 to 81.0)	76.7 (70.7 to 82.9)	80.8 (75.1 to 86.5)	-1.6 (-8.1 to 4.9)	-4.0 (-10.6 to 2.6)
Role limitations – physical	28.8 (21.6 to 36.0)	51.9 (43.2 to 60.7)	50.6 (40.6 to 60.7)	-22.8 (-32.7 to -12.8)	-21.8 (-34.9 to -8.6)
Role limitations – emotional	62.0 (52.3 to 71.6)	68.2 (59.3 to 77.1)	69.6 (61.0 to 78.3)	-6.5 (-16.6 to 3.6)	-5.1 (-16.6 to 6.4)
Vitality	54.8 (50.1 to 59.4)	57.5 (52.8 to 62.2)	57.8 (52.3 to 63.3)	-3.1 (-7.0 to 0.9)	-2.1 (-7.8 to 3.6)
General health	69.6 (65.2 to 74.0)	66.3 (61.7 to 70.9)	66.6 (61.6 to 71.6)	3.9 (0.5 to 7.2)	2.4 (-2.1 to 6.9)
Change in health status	46.5 (42.1 to 50.9)	63.2 (57.3 to 69.1)	50.6 (45.7 to 55.5)	-16.8 (-22.9 to -10.6)	-3.7 (-9.8 to 2.5)

on the SF-36 such as physical activity, pain and physical role limitations.

Table II presents the mean pre- and postoperative scores, the two independent health scores and the mean change scores (preoperative score minus postoperative score) for each follow-up assessment. Paired *t*-tests were used to compare the change scores at follow-up at six months and four years for each of the measures and dimensions of health status.

The mean change in Constant scores worsened between the assessments at six months and four years. The difference between these scores was significant (paired *t*-test, mean difference 13.38, *t* = 5.66, *df* 55, *p* < 0.001). The mean change, however, in scores at both follow-up assessments were very similar for the OSS. This was also the case for the dimensions of pain, role limitation due to physical and emotional factors, vitality and general health perception in the SF-36 (paired *t*-tests, OSS, mean difference 0.66, *t* = 0.76 *df* 72 *p* = 0.452; SF-36/physical function, mean difference 6.0, *t* = 2.2 *df* 71 *p* = 0.030; pain, mean difference 0.14, *t* = 0.14 *df* 76 *p* = 0.96; mental health, mean difference -1.45, *t* = -0.70 *df* 77 *p* = 0.487; social function, mean difference -4.6, *t* = -1.6 *df* 76 *p* = 0.108; role limitation, physical, mean difference 1.33, *t* = 0.25 *df* 74 *p* = 0.802; role limitation, emotional, mean difference 0.45, *t* = 0.09 *df* 73 *p* = 0.932; vitality, mean difference 1.49, *t* = 0.625 *df* 76 *p* = 0.534; general health perception, mean difference -0.96, *t* = -0.401 *df* 75 *p* = 0.69).

The shoulder-specific assessments showed a large change at both periods of follow-up. The size of this change differed between the Constant and OSS, however, and the mean difference between the change scores was statistically significant for the Constant (paired *t*-test, mean change preop to 6 months postop -32.27 *versus* mean change preop to four years postop -18.89, *t* = -5.7 *df* 55 *p* < 0.001). This was not the case for the OSS (paired *t*-test, mean change preop to six months postop 10.49 *versus*

mean change preop to four years postop 11.18, *t* = -0.7, *df* 71 *p* = 0.433).

When responding to questions about changes in health after shoulder surgery most patients (72/91, 79% at six months; 62/82, 76% at four years) said that their day-to-day life was either 'much better' or 'slightly better'. Most (78/91, 86% at six months; 69/82, 84% at four years) also said that symptoms from their shoulder were 'completely cured', 'much better', or 'slightly better', and most (67/82, 82%) said that their operation had been either 'very' or 'fairly successful'. The patients were divided into two groups for each item, on the basis of their responses (any improvement *versus* no change/worse). The mean changes of the Constant, OSS and SF-36 scores were analysed for the different grouped responses.

Table III shows comparisons between grouped responses to questions about retrospective assessment of change ('transition'), the success of the operation at six months and four years after surgery, and change scores of the Constant, OSS and SF-36 dimensions. There were marked differences in the mean change scores measured by the Constant assessment, OSS and the pain and physical role limitation dimensions of the SF-36 at six months after surgery, according to whether patients reported that their day-to-day life was much or slightly better rather than unchanged or worse. These differences diminished slightly for the Constant change score at four years, becoming barely statistically significant, while differences in the OSS and SF-36 pain dimension change scores were maintained.

Differences in OSS change scores were also seen in the patients' judgement of the success of the operation, with responses again grouped into two categories, at each period of follow-up. These differences were highly significant. The Constant shoulder score presented a different picture, with differences in change scores related to the patients' assessment of the change in symptoms only barely approaching statistical significance for the Constant assess-

Table III. Comparison between patients' grouped responses about postoperative changes in day-to-day life, improvements in shoulder-related problems and view of success of their surgery, compared with CSS, OSS and SF-36 scores at the six-month and four-year follow-ups

SF-36 health survey												
Patient responses	CSA		OSS		Physical functioning		Pain		Mental health		Social functioning	
	Mean change Mean difference (95% CI)	Mean change Mean difference (95% CI)	Mean change Mean difference (95% CI)	Mean change Mean difference (95% CI)	Mean change Mean difference (95% CI)	Mean change Mean difference (95% CI)	Mean change Mean difference (95% CI)	Mean change Mean difference (95% CI)	Mean change Mean difference (95% CI)	Mean change Mean difference (95% CI)	Mean change Mean difference (95% CI)	Mean change Mean difference (95% CI)
'How has your shoulder operation changed your day-to-day life so far?'												
Preoperatively to six months	n = 83	n = 87	n = 86	n = 89	n = 89	n = 89	n = 89	n = 89	n = 89	n = 87	n = 89	n = 88
Much/slightly improved	-39.8 p < 0.001	13.0 p < 0.001	-8.3 p = 0.42	-25.6 p < 0.001	-3.8 p = 0.12	-4.3 p = 0.12	-30.7 p < 0.001	-10.3 p = 0.16	-4.4 p = 0.30	3.9 p = 0.10	-19.4 p = 0.09	
Other responses	-14.9 -24.9 (-36.1 to -3.7)	2.0 11.0 (6.9 to 15.0)	-4.4 -3.9 (-13.3 to 5.6)	-2.3 -23.2 (32.4 to -14.0)	3.1 -6.9 (-15.8 to 1.9)	8.2 -12.5 (-28.2 to 3.2)	6.6 -37.3 (-54.4 to 0.1)	7.0 -37.3 (-54.4 to 20.1)	1.9 -6.31 (-16.1 to 3.5)	3.9 -0.01 (-8.4 to 8.3)	-6.6 -12.9 (-27.9 to 2.1)	
Preoperatively to four years	n = 60	n = 77	n = 73	n = 81	n = 80	n = 81	n = 78	n = 78	n = 80	n = 78	n = 82	
Much/slightly improved	-23.4 p = 0.013	14.4 p < 0.001	-3.4 p = 0.27	-24.2 p = 0.008	-4.1 p = 0.40	-7.7 p = 0.052	-27.5 p = 0.16	-6.1 p = 0.76	-3.2 p = 0.51	2.4 p = 0.97	-5.2 p = 0.37	
Other responses	-7.1 -16.3 (-29.0 to -3.6)	1.3 13.1 (8.91 to 17.32)	5.3 -8.7 (-24.3 to 6.96)	-4.7 19.5 (-33.7 to -5.3)	0.2 -4.3 (-14.3 to 5.8)	7.2 -14.9 (-29.9 to 0.2)	-1.5 -26.0 (-47.0 to -5.1)	-1.9 -4.3 (-31.7 to 23.2)	1.3 -4.5 (-17.9 to 8.90)	2.2 0.2 (-10.6 to 11.0)	1.2 -6.49 (-20.9 to 7.9)	
'How have your shoulder-related problems changed since your operation?'												
Preoperatively to six months	n = 83	n = 87	n = 86	n = 89	n = 89	n = 89	n = 89	n = 87	n = 89	n = 88	n = 91	
Much/slightly improved	-36.9 p = 0.02	12.0 p = 0.001	-8.44 p = 0.23	-23.2 p = 0.02	-3.4 p = 0.18	-4.7 p = 0.02	-28.6 p = 0.004	-10.8 p = 0.17	-4.4 p = 0.28	3.9 p = 0.28	-19.9 p = 0.01	
Other responses	-19.8 -17.2 (-31.3 to -3.0)	2.8 -9.2 (4.0 to 14.4)	-1.7 -6.8 (-17.8 to 4.2)	-5.1 -18.1 (-33.6 to -2.6)	3.7 -7.0 (-17.5 to 3.4)	16.2 -20.9 (-38.8 to -3.0)	11.5 -40.2 (67.1 to -13.2)	17.9 -28.8 (-71.1 to 13.6)	5.4 -9.8 (-28.7 to 9.0)	3.9 -0.04 (-9.9 to 9.8)	1.9 -21.8 (-38.9 to -4.7)	
Preoperatively to four years	n = 58	n = 74	n = 71	n = 79	n = 78	n = 78	n = 75	n = 76	n = 78	n = 76	n = 79	
Cured/much or slightly improved	-20.8 p = 0.61	13.2 p = 0.001	-3.7 p = 0.079	-22.5 p = 0.045	-4.5 p = 0.41	-7.5 p = 0.038	-25.4 p = 0.074	-7.1 p = 0.69	-5.1 p = 0.005	1.4 p = 0.12	-6.2 p = 0.24	
Other responses	-16.1 -4.6 (-22.5 to 13.3)	3.3 9.9 (3.9 to 15.8)	14.4 -18.1 (-38.4 to 2.2)	-3.3 -19.2 (-38.0 to -0.4)	0.8 -5.3 (-18.0 to 7.4)	13.3 -20.8 (-40.5 to -1.2)	0.0 -25.4 (-53.5 to 2.8)	0.0 -7.1 (-41.9 to 27.8)	19.0 -24.1 (-40.7 to -7.4)	11.9 -10.5 (-23.8 to 2.8)	5.0 -11.2 (-29.8 to 7.5)	
'How successful was your shoulder operation?'												
Preoperatively to six months	n = 81	n = 85	n = 84	n = 87	n = 87	n = 87	n = 87	n = 85	n = 87	n = 86	n = 89	
Very/fairly successful	-37.9 p = 0.003	12.0 p = 0.001	-8.3 p = 0.52	-23.1 p = 0.06	-2.9 p = 0.55	-3.2 p = 0.37	-25.0 p = 0.35	-8.8 p = 0.19	-3.4 p = 0.79	3.8 p = 0.92	-19.7 p = 0.007	
Other responses	-16.0 -22.0 (-36.4 to -7.5)	3.0 9.0 (3.7 to 14.3)	-4.6 -3.7 (-14.9 to 7.8)	-7.7 -15.4 (-31.2 to 0.33)	0.3 -3.2 (-13.8 to 7.4)	5.1 -8.3 (-26.6 to 10.1)	-11.5 -13.5 (-42.0 to 15.0)	10.3 -19.1 (-47.5 to 9.4)	0.8 -2.6 (-22.9 to 17.8)	4.3 -0.50 (-10.4 to 9.4)	3.8 -23.6 (-40.7 to -6.5)	
Preoperatively to four years	n = 60	n = 77	n = 73	n = 81	n = 80	n = 81	n = 77	n = 78	n = 70	n = 78	n = 82	
Very/fairly successful	-20.1 p = 0.53	12.8 p = 0.003	-2.5 p = 0.53	-22.6 p = 0.04	-4.2 p = 0.25	-6.9 p = 0.064	-23.0 p = 0.55	-6.8 p = 0.55	-3.6 p = 0.28	2.6 p = 0.81	-4.9 p = 0.42	
Other responses	-15.2 -4.9 (20.4 to 10.6)	4.5 8.3 (3.0 to 13.6)	2.9 -5.4 (-22.5 to 11.8)	-5.6 -17.0 (-33.2 to 0.82)	2.3 -6.5 (-17.7 to 4.7)	8.9 -15.8 (-23.5 to 0.9)	-15.4 -7.7 (-43.0 to 27.7)	2.4 -9.2 (-39.2 to 20.9)	4.6 -8.2 (-23.1 to 6.7)	1.2 1.4 (-10.4 to 13.3)	1.7 -6.5 (-22.5 to 9.5)	

ment at follow-up at six months; the differences at four years were not statistically significant. The differences in Constant change scores, however, which related to the patients' assessment of the success of their operation, were highly significant at six months but not at four years.

The mean differences in change scores on the SF-36 domains relating to responses about specific questions concerning symptoms and the patients' judgement of the success of their operation were occasionally large at both assessments, particularly with regard to pain and physical role limitations. The variation associated with these differences, however, was also large and the level of statistical significance correspondingly low, particularly at the assessment at four years.

Discussion

It is difficult to assess the long-term outcome after surgical treatment for conditions of the shoulder. It is costly to provide appropriate extra clinic appointments, and patients are often unable or unwilling to attend for assessment some years after treatment. The most significant problem, however, is the lack of appropriate methods of assessing outcome.

In our study only 66% of the original sample underwent a clinical assessment at a mean period of 3.9 years (2.2 to 5.6) after surgery. We acknowledge that this variation in the period of follow-up may have affected the clinical validity of the investigation.

The correlation coefficients between absolute scores on the OSS, the Constant shoulder assessment and relevant dimensions of the SF-36 were generally high ($r > 0.5$) and highly significant. The mean change scores for the Constant, OSS and relevant dimensions of the SF-36 showed that substantial improvement had occurred at six months after surgery. These were maintained after four years, as measured by the OSS and the pain and physical role limitation dimensions of the SF-36. This was surprising since the SF-36 is a generic measure and will therefore reflect any changes in general health.¹¹ Such interference may have been expected after a period of four years. More surprising, however, was the deterioration in the mean change score for the clinically assessed Constant shoulder score, which had almost halved at four years.

Further support for patient-based assessments, in particular the shoulder-specific OSS, came from comparisons between mean change scores, grouped responses and a question regarding the success of the operation. Patients reported considerable differences in mean change scores at six months; thus those who indicated that their life was much or slightly better because of their operation had greater mean differences in change scores at six months after surgery on the Constant, OSS and relevant domains of the SF-36 than did patients who indicated no improvement. These differences were highly significant. Equivalent analyses regarding the four-year assessment showed similar

results for the OSS and pain dimension of the SF-36 while differences obtained on the Constant score barely approached statistical significance. Regarding comparisons between similarly grouped responses to more specific questions about change in symptoms and mean differences in change scores, such differences were found to be similar at six months and at four years for the OSS and for relevant domains of the SF-36, whereas for the Constant score the mean difference in change score was significantly reduced between the two postoperative assessments. Thus while the mean difference in Constant change score at the six-month assessment approached statistical significance, by the four-year assessment it was reduced and not statistically significant.

With regard to grouped responses to an item about their judgement of the success of the operation, differences in change scores for OSS mirrored differences in the patients' judgements at both preoperative assessments, while differences in the Constant score were not significantly maintained for longer than six months. Differences in change scores for the relevant dimensions of the SF-36 were not generally of statistical significance. These findings suggest reduced reliability and sensitivity of the Constant score relative to the OSS over the longer term.

This could have arisen for a number of reasons. It may have been due to a smaller sample having been assessed by the former, or that the clinical assessments were carried out by two clinicians, one at each phase of follow-up. There may either have been differences in the way in which the clinicians carried out the assessment, or that patients moderated their responses according to the assessor. A previous study has also questioned the interobserver reliability of the Constant shoulder assessment.¹²

Our study supports the use of patient-based measures of outcome and particularly the shoulder-specific OSS to supplement clinical assessment in the follow-up of patients after shoulder surgery.

Financial support was provided by grants from Oxford Regional Health Authority (Audit) and The Medical Research Fund Committee, Faculty of Clinical Medicine, Oxford University.

We thank Vicky Bloomfield and Kim Clipsham for their assistance with conducting this study, together with all nursing staff in the preadmission and other outpatient clinics at the Nuffield Orthopaedic Centre. We also thank individual patients for giving their valuable time.

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

References

1. Green S, Buchbinder R, Glazier R, Forbes A. Interventions for shoulder pain (Cochrane review). *The Cochrane Library, Issue 3*. Oxford: Update Software, 2000.
2. Fitzpatrick R, Fletcher A, Gore S, et al. Quality of life measures in health care. I: Applications and issues in assessment. *BMJ* 1992;305:1074-7.
3. MacDonald DA. The shoulder and elbow. In: Pynsent PB, Fairbank JC, Carr A, eds. *Outcome measures in orthopaedics*. Butterworth Heinemann, Oxford, 1993:144-73.
4. Van der Windt DA, Van der Heijden GJ, Scholten RJ, Koes BW, Bouter LM. The efficacy of non-steroidal anti-inflammatory drugs (NSAIDs) for shoulder complaints: a systematic review. *J Clin Epidemiol* 1995;48:691-704.

5. **Van Der Heijden GJ, van der Windt DA, Kleijnen J, Koes BW, Bouter LM.** Steroid injections for shoulder disorders: a systematic review of randomised clinical trials. *Br J Gen Pract* 1996;46:309-16.
6. **Dawson J, Fitzpatrick R, Murray D, Carr A.** Comparison of measures to assess outcomes in total hip replacement surgery. *Qual Health Care* 1996;5:81-8.
7. **Dawson J, Fitzpatrick R, Carr A.** The assessment of shoulder instability: the development and validation of a questionnaire. *J Bone Joint Surg [Br]* 1999;81-B:420-6.
8. **Constant CR, Murley AHG.** A clinical method of functional assessment of the shoulder. *Clin Orthop* 1987;214:160-4.
9. **Dawson J, Fitzpatrick R, Carr A.** Questionnaire on the perceptions of patients about shoulder surgery. *J Bone Joint Surg [Br]* 1996;78-B:593-600.
10. **Ware JE, Sherbourne CD.** The MOS 36-item short-form health survey (SF36). I. conceptual framework and item selection. *Med Care* 1992;30:473-83.
11. **Dawson J, Fitzpatrick R, Murray D, Carr A.** The problem of 'noise' in monitoring patient-based outcomes: generic, disease-specific and site-specific instruments for total hip replacement. *J Health Serv Res Policy* 1996;1:224-31.
12. **Conboy VB, Morris RW, Kiss J, Carr AJ.** An evaluation of the Constant-Murley shoulder assessment. *J Bone Joint Surg [Br]* 1996;78-B:229-32.