

The assessment of shoulder instability

THE DEVELOPMENT AND VALIDATION OF A QUESTIONNAIRE

J. Dawson, R. Fitzpatrick, A. Carr

From the University of Oxford and the Nuffield Orthopaedic Centre, Oxford, England

We have developed a 12-item questionnaire for completion by patients presenting with shoulder instability. A prospective study of 92 patients was undertaken involving two assessments, approximately six months apart, performed in an outpatient department. Each patient completed the new questionnaire and the SF36 form. An orthopaedic surgeon completed the Constant shoulder score and the Rowe assessment.

The new questionnaire and the Rowe clinical score each achieved a large standardised effect size (≥ 0.8) and compared favourably with relevant items on the SF36. By contrast, the Constant score barely registered any effect, confirming that it may be relatively insensitive to changes in clinical status for this particular condition.

The questionnaire provides a measurement of outcome for shoulder instability which is short, practical, reliable, valid and sensitive to changes of clinical importance.

J Bone Joint Surg [Br] 1999;81-B:420-6.

Received 20 April 1998; Accepted after revision 30 July 1998

The shoulder is designed for maximum mobility and is therefore the most common joint to develop instability.¹ This need not reach true dislocation, but subluxation, of which a person may be unaware, can cause symptoms.^{1,2} It may occur in all age groups but is more common under the age of 30 years.³

Frank dislocation may result from injury, and recurrence is very common after traumatic dislocation.^{1,2,4} Less often, the condition arises from congenital abnormality of the

ligaments and capsule of the joint and is then more likely to be multidirectional.

In some cases the condition responds well to physiotherapy, but a number of different operations have been recommended to stabilise the joint.^{1,5} There is, however, a dearth of evidence regarding the effectiveness of the surgical management of shoulder instability. Several methods of scoring have been developed to determine the outcome of the orthopaedic management of shoulder conditions but they derive from clinical and radiological data and depend upon the judgement of the surgeon.⁶ Patients and surgeons may differ in their concerns and priorities,⁷ and methods are required which elicit the patient's perception of the outcome.⁸ Patients can provide reliable and valid judgements of health status and of the benefits of treatment.⁹

There are advantages in using questionnaires designed to address the patient's perception of a single condition.¹⁰ They are intended to be most sensitive to the outcome in narrowly defined groups. Shorter questionnaires may be as sensitive to changes of importance to patients as longer versions,^{11,12} and have the added advantage of being more acceptable for routine use, particularly in the long term.¹³

We have previously described the development of an outcome measure for use in assessing treatment for the painful shoulder.¹⁴ During the early stages of that study it became clear that patients presenting with recurrent dislocation or subluxation of the shoulder were a distinct group whose condition was most usually characterised not so much by pain as by the anticipation of problems arising in relation to very specific activities. This group required a specific outcome measurement of its own.

We describe the development of a short questionnaire for use in assessing the outcome after treatment for instability of the shoulder, which is intended to be reliable, reproducible, valid and sensitive to change. The questionnaire was applied in a prospective study of patients, most of whom received either physiotherapy or surgery for instability of the shoulder.

Patients and Methods

Development of the questionnaire. Initially, we interviewed 20 patients attending an outpatient clinic to which they had been referred with instability of their shoulder, in

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

A. Carr, FRCS, Consultant Orthopaedic Surgeon
Upper Limb Unit, Nuffield Orthopaedic Centre, Windmill Road, Headington,
Oxford OX3 7LD, UK.

Correspondence should be sent to Dr J. Dawson.

©1999 British Editorial Society of Bone and Joint Surgery
0301-620X/99/39044 \$2.00

order to identify ways in which they had experienced and reported their problem. We then drafted an 18-item questionnaire and tested it on 20 new patients. They were also given a second copy of the questionnaire and asked to complete it at home on the following day, and to return it. They were invited to add their comments to this copy and to include any further shoulder problems which were not addressed by it.

The original questionnaire was then modified and the revised version tested on two further groups of 20 patients until its final form was established. This contained 12 items, each of which had five response categories (Table I). Each item was scored from 1 to 5, from least to most difficulty or severity and combined to produce a single score with a range from 12 (least difficulties) to 60 (most difficulties).

We then examined whether the new questionnaire was internally consistent, reproducible, valid and sensitive to clinical change.¹⁵ Internal consistency determines whether items measure a single underlying concept. Reproducibility is concerned with whether the questionnaire yields the same results on repeated trials under the same conditions. Validity determines whether it does measure what it purports to. The validity of content shows whether items in a questionnaire cover the intended topics clearly; that of construct compares a set of relationships with other variables which may be expected. Sensitivity to change, or responsiveness, reflects the ability of the questionnaire to detect clinically significant changes over time.

Between March 1994 and August 1996 we recruited 92 patients in an outpatient clinic at the Nuffield Orthopaedic Centre, Oxford, who had been referred by their general practitioner with problems related to instability of the shoulder. They completed a variety of assessments on the first occasion and again at routine appointments in the same clinic approximately six months later after a course of physiotherapy, or six months after operation. Patients who did not attend the outpatient department at the appropriate time completed a proportion of the follow-up assessment by post, with up to two reminders sent.

The identity and details of four patients were inadequate leaving 53 men and 35 women with a median age of 25 years (12 to 54; mean 26.3; SD 9.1). One further individual had an imprecise diagnosis by his referring doctor. Diagnostic details were therefore obtained for 87 patients (95%), either through clinical assessment or, in a few cases, retrospectively from the medical notes, and two diagnostic subgroups were identified. Most patients (68, 78%) were assessed as having a unidirectional form of instability, while in 19 patients (22%) it was multidirectional. This latter group was younger with a mean age of 19.5 years (SD 6.1).

Of the 92 patients included in the study, 53 received an initial clinical assessment. Six recorded details which were inadequate rendering follow-up impossible, and a relatively high proportion of patients repeatedly failed to attend confirmed follow-up appointments. Of this latter group five did

not even respond to postal follow-up. Two patients were known to have moved abroad to unknown addresses; a further 15 had been discharged from the clinic and did not return postal questionnaires. Nevertheless, 64 patients (70%) had at least one follow-up in the form of a postal questionnaire or a questionnaire plus a clinical assessment. Unfortunately, of the 53 patients who received an initial standard clinical assessment, only 15 had a clinical assessment at follow-up. Change scores for the clinical assessments are presented for this latter subgroup.

Of the 64 patients who completed a follow-up assessment, 34 (53%) had undergone an operation to stabilise their shoulder followed by a course of physiotherapy. A similar number, although not necessarily the same people, received a course of physiotherapy; this included a small proportion of the 12 patients who either had surgery delayed beyond the period of study or had moved away before the appointed time.

Internal consistency. Internal consistency was tested by using Cronbach's alpha¹⁶ for the study questionnaire both preoperatively and at follow-up. This summarises the internal correlations of all items in a scale. The higher the alpha coefficient (0.0 to 1.0) the more consistent is the scale and the greater the likelihood that it is tapping an underlying single variable on the questionnaire. We examined correlations of all items with the overall score and assessed whether Cronbach's alpha was improved by removal of any item.

Reproducibility. Reproducibility (test-related reliability) was assessed by asking 43 patients at the first assessment to complete and return a second questionnaire 24 hours after the first. Correlations of the total scores were measured by Pearson correlation coefficients. Paired Student's *t*-tests were also carried out to examine whether there was a change in the distribution of scores between the two tests. The data were also examined by the coefficient of reliability according to the method described by Bland and Altman.¹⁷

Construct validity. This was examined by means of Pearson correlation coefficients between the total score of the questionnaire and other related measures obtained at the first assessment and the later follow-up. It was expected that scores should correlate moderately with those assessed by the methods of Constant and Rowe rated by an orthopaedic surgeon and with those from another health-status questionnaire (SF36) completed by the patient.

The Constant shoulder score¹⁸ contains both subjective and objective elements. The subjective assessments of pain and activities of daily living are allocated 15 and 20 points, respectively, out of a total of 100. Objective measurements of the active range of movement and power are allocated 40 and 25, respectively. We used a goniometer to assess the range of movement while power was measured in kilograms of resisted abduction using a spring balance, as described by Constant. It has been suggested that the Constant shoulder score is not appropriate for assessing instability.^{6,19}

Table I. Scores obtained on the 12-item shoulder instability questionnaire from 92 patients during the first assessment and from 64 patients at the six-month follow-up

Item	Scoring categories	First assessment		6-month follow-up	
		Number	Percentage	Number	Percentage
1. During the last six months, how many times has your shoulder slipped out of joint (or dislocated?)	1 Not at all in 6 months	16	17	37	58
	2 1 or 2 times in 6 months	33	36	13	20
	3 1 or 2 times per month	16	17	4	6
	4 1 or 2 times per week	6	7	0	0
	5 More often than 1 or 2 times/week	21	23	10	16
2. During the last three months, have you had any trouble (or worry) dressing because of your shoulder?	1 No trouble at all	25	27	29	45
	2 Slight trouble or worry	40	44	22	34
	3 Moderate trouble or worry	24	26	9	14
	4 Extreme difficulty	2	2	4	6
	5 Impossible to do	1	1	0	0
3. During the last three months, how would you describe the worst pain you have had from your shoulder?	1 None	5	5	5	8
	2 Mild ache	21	23	34	53
	3 Moderate	31	34	10	16
	4 Severe	23	25	12	19
	5 Unbearable	12	13	3	5
4. During the last three months, how much has the problem with your shoulder interfered with your usual work (including school or college work, or housework)?	1 Not at all	15	16	21	33
	2 A little bit	24	26	24	38
	3 Moderately	23	25	10	16
	4 Greatly	22	24	3	5
	5 Totally	8	9	6	9
5. During the last three months, have you avoided any activities due to worry about your shoulder - feared that it might slip out of joint?	1 Not at all	4	4	13	20
	2 Very occasionally	10	11	19	30
	3 Some days	29	32	10	16
	4 Most days or more than one activity	28	30	15	23
	5 Every day or many activities	21	23	7	11
6. During the last three months, has the problem with your shoulder prevented you from doing things that are important to you?	1 No, not at all	12	13	22	34
	2 Very occasionally	17	19	20	31
	3 Some days	26	28	11	17
	4 Most days or more than one activity	20	22	7	11
	5 Every day or many activities	17	19	4	6
7. During the last three months, how much has the problem with your shoulder interfered with your social life (including sexual activity - if applicable)?	1 Not at all	19	21	27	42
	2 Occasionally	29	32	21	33
	3 Some days	24	26	11	17
	4 Most days	16	17	5	8
	5 Every day	4	4	0	0
8. During the last four weeks, how much has the problem with your shoulder interfered with your sporting activities or hobbies?	1 Not at all	10	11	18	28
	2 A little/occasionally	21	23	20	31
	3 Some of the time	15	16	10	16
	4 Most of the time	15	16	10	16
	5 All of the time	31	34	6	9
9. During the last four weeks, how often has your shoulder been 'on your mind' - how often have you thought about it?	1 Never, or only if someone asks	3	3	8	13
	2 Occasionally	18	20	22	34
	3 Some days	9	10	8	13
	4 Most days	27	29	13	20
	5 Every day	35	38	13	20
10. During the last four weeks, how much has the problem with your shoulder interfered with your ability or willingness to lift heavy objects?	1 Not at all	19	21	26	41
	2 Occasionally	12	13	12	19
	3 Some days	15	16	9	13
	4 Most days	22	24	7	11
	5 Every day	24	26	10	16
11. During the last four weeks, how would you describe the pain which you usually had from your shoulder?	1 None	24	26	20	31
	2 Very mild	20	22	27	42
	3 Mild	20	22	5	8
	4 Moderate	17	19	9	14
	5 Severe	11	12	3	5
12. During the last four weeks, have you avoided lying in certain positions, in bed at night, because of your shoulder?	1 No nights	16	17	20	31
	2 Only 1 or 2 nights	9	10	9	14
	3 Some nights	17	19	11	17
	4 Most nights	24	26	12	19
	5 Every night	26	28	12	19

The Rowe standard rating scale²⁰ was devised to assess outcomes after surgical repair for recurrent dislocation of the shoulder. The system requires a clinician to assess stability, range of movement and function. Each is accorded one of four possible values. With regard to stability 'no recurrence, subluxation or apprehension' has the maximum score (50 points) while 'recurrent dislocation' receives the minimum score (0). Assessment of movement gives the maximum score (20 points) for '100% of normal external and internal rotation and elevation', and the lowest score (0) for '50% of normal elevation and internal elevation; but no external rotation'. Function is graded as a maximum score of 30 points for 'no limitation in work or sports and little or no discomfort' with the lowest score (0) going to 'marked limitation and pain.' The totals for each are added to produce an overall range of scores between 0 and 100. A score of 90 to 100 is judged 'excellent', 75 to 89 as 'good', 51 to 74 'fair' and 50 or less 'poor'. Although it is widely used, this scoring system has not been assessed as regards its reliability, validity or sensitivity to change.

The SF36 is a 36-item questionnaire which is widely used to measure health status.²¹ It provides scores on eight dimensions namely physical function, social function, role limitation due to physical problems, role limitation due to emotional problems, general mental health, energy, bodily pain and general health perceptions. Scores for each item range from 0 (poor) to 100 (good).

All patients were asked to complete the new shoulder questionnaire and the SF36 at the first assessment and also at follow-up.

Sensitivity to change. The sensitivity to change of the questionnaire was examined by comparing scores before and at six months after surgery. The scores of 64 patients were available for these analyses.

We calculated the effect sizes for the new questionnaire and for the SF36. These were also calculated for the Constant and Rowe clinical assessments in 15 patients who had been evaluated by them in the outpatient clinic. This is a method of calculating the extent of change measured using a questionnaire in a standardised way which allows

comparison.²² In our study it was determined as the difference between the mean pretreatment and post-treatment scores for the samples, divided by the standard deviation of preoperative scores. An effect size of 1.0 is equivalent to a change of one standard deviation in the sample and is considered to be very large. An effect size of 0.2 is small.

Sensitivity was also examined by means of patient satisfaction and transition items²³ in which patients were asked to say how successful they thought that the treatment had been, and also to assess the change that had occurred in their shoulder problem. The first question offered four possible categories of response while the transition item offered five. Patients giving the most positive response in each case were compared with all others by way of their mean change in score, i.e., the preoperative minus the postoperative score, for each questionnaire. Patients were also asked to say whether they felt that there was still any 'room for improvement' in their shoulder. The change in score for those responding with 'little' or 'no improvement' was also compared with that of other patients.

Statistical significance was measured by the chi-squared and Student's *t*-tests.

Results

Table I gives the individual scores obtained for patients on the new shoulder questionnaire at first assessment and at follow-up. When first evaluated, the summed score for the questionnaire had a median value of 37.0 (15 to 57) and a mean of 36.6 (SD 10.6, 95% confidence limits (CL) 34.4 to 38.8) and at follow-up a median value of 26 (12 to 55) and a mean of 28.3 (SD 11.2, 95% CL 25.6 to 31.1).

Internal consistency. Cronbach's alpha for the study questionnaire was 0.91 at the pretreatment assessment (n = 92) and 0.92 at follow-up (n = 64). All items correlated with the total score of more than 0.5. Cronbach's alpha was also compared for the two diagnostic subgroups. In each case the alpha remained more than 0.9 (Table II).

Reproducibility. In the test-retest sample (n = 34), the correlation between the total scores for the questionnaire

Table II. Internal consistency of the shoulder instability questionnaire for 92 patients before treatment and 64 after treatment

Question	Content	Pretreatment			Follow-up		
		Mean score (SD)	Item-total correlation	Alpha if item removed	Mean score (SD)	Item-total correlation	Alpha if item removed
1	Dislocations in six months	2.8 (1.4)	0.53	0.90	2.0 (1.4)	0.55	0.92
2	Trouble with dressing	2.1 (0.9)	0.69	0.90	1.8 (0.9)	0.78	0.91
3	Worst pain	3.2 (1.1)	0.59	0.90	2.6 (1.0)	0.70	0.92
4	Interference with work	2.8 (1.2)	0.71	0.89	2.2 (1.2)	0.72	0.91
5	Avoidance of activities	3.6 (1.1)	0.66	0.90	2.8 (1.3)	0.69	0.92
6	Prevented activities of importance	3.1 (1.3)	0.73	0.89	2.2 (1.2)	0.80	0.91
7	Interference with social life	2.5 (1.1)	0.67	0.90	1.9 (1.0)	0.74	0.92
8	Interference with sport/hobbies	3.4 (1.4)	0.68	0.90	2.5 (1.3)	0.62	0.92
9	Shoulder 'on your mind'	3.8 (1.2)	0.74	0.89	3.0 (1.4)	0.78	0.91
10	Interference with lifting	3.2 (1.5)	0.67	0.90	2.4 (1.5)	0.75	0.91
11	Usual level of pain	2.7 (1.4)	0.67	0.90	2.2 (1.2)	0.68	0.92
12	Avoided positions in bed at night	3.4 (1.4)	0.38	0.91	2.8 (1.5)	0.51	0.93
				First assessment Cronbach Alpha: 0.91	Post-treatment Cronbach Alpha: 0.92		

was high ($r = 0.97$, $p < 0.0001$). No significant change occurred in the distribution of scores between the two assessments for reliability (first assessment, mean 35.5 and SD 11.0; retest mean 35.2 and SD 11.6; paired Student's t -test > 0.05). The differences in scores between the first and second which was obtained 24 hours later, were plotted against their means. The scatter appeared normal and showed the same variability across the range of scores at all levels of disability or severity. The estimated mean (0.26) of score differences was not significantly different from 0. The coefficient of reliability was calculated as 5.7 using the Bland and Altman method¹⁷ and 95% of score differences fell between 0 ± 5.7 . Overall, 88% of score differences lay between 0 ± 4 points.

Construct validity. The new questionnaire correlated well with the Constant and Rowe clinical scores both before operation and at the six-month follow-up (Table III). It also agreed significantly with the related parts of the SF36, particularly in physical function and pain.

Sensitivity to change. Patients reported substantial improvement in health status at the six-month follow-up. Overall, 44 patients (73%) described their treatment as either very or fairly successful and 47 (78%) said that problems relating to their shoulder were either much or slightly better. This agreed with the effect size obtained for the new questionnaire, which was large (0.8). The two clinical assessments performed very differently from one another. The Rowe method measured an even larger effect than the new shoulder questionnaire (1.2), while the Constant score was only able to detect a small effect (0.2). The SF36 was also only able to detect a small effect overall. The 'physical role limitation' measured the largest effect size of 0.5, denoting a moderate degree of effect only (Table IV).

Table V gives further evidence for the sensitivity of the new questionnaire to change in clinical condition. In judging patient satisfaction, transition and scope for improvement, the new questionnaire distinguished clearly between those patients who rated the most positive change in their shoulder after treatment, with little scope for further

Table III. Correlation between the 12-item shoulder questionnaire, the Constant and Rowe shoulder assessment, and SF36 health-status measure

	Correlation coefficients	
	First assessment	Follow-up
Constant	-0.56*	-0.76*
Rowe	-0.51*	-0.84*
SF36		
Physical activity	-0.71*	-0.67*
Pain	-0.70*	-0.73*
Mental health	-0.35†	-0.27‡
Social function	-0.58*	-0.43*
Role limitations - physical reasons	-0.69*	-0.61*
Role limitations - emotional reasons	-0.30†	-0.49*
Vitality	-0.20	-0.33†
Health perceptions	-0.08	-0.42†

* $p < 0.0001$

† $p < 0.01$

‡ $p < 0.05$

improvement, and those who said that improvement had been only slight or that any change was for the worse. Neither clinical assessment was able to detect any difference in any of the items, although this was most likely to be due to the small numbers compared. In each case the level of significance achieved was much higher for the new questionnaire than for any of the relevant parts of the SF36; indeed the latter was unable to distinguish any difference in the success of treatment.

Discussion

Instability of the shoulder poses particular problems for assessment as symptoms are often intermittent, and characterised less by the everyday presence of pain than by the anticipation of problems arising in relation to specific activities.

We have developed and tested a short 12-item questionnaire which patients have found easy to complete and which provides reliable, valid and responsive information as to their perception of shoulder instability. It is intended for use as an outcome measure, and poses few difficulties

Table IV. Comparison by effect size of the 12-item shoulder questionnaire with other health-status measures regarding their ability to measure clinical change in the condition of the patient between first assessment (Time 1) and follow-up (Time 2)

Test	Mean score				Effect size
	Time 1	95% CL	Time 2	95% CL	
12-item questionnaire	36.6	34.4 to 38.8	28.3	25.6 to 31.1	0.8
Constant	83.8	78.2 to 89.4	80.5	71.4 to 89.6	0.2
Rowe	42.6	36.6 to 48.7	68.4	54.6 to 82.1	1.2
SF36					
Health perceptions	75.6	72.0 to 79.3	77.8	74.0 to 81.5	-0.1
Health change	50.8	47.5 to 54.2	56.0	51.3 to 60.6	-0.3
Physical activity	84.0	81.2 to 86.7	87.3	83.6 to 90.9	-0.3
Pain	64.1	58.5 to 69.7	75.7	69.3 to 82.0	-0.4
Role limitation - physical	54.6	46.0 to 63.2	73.8	64.8 to 82.9	-0.5
Role limitation - mental	81.0	74.2 to 87.7	81.5	73.3 to 89.7	0.0
Mental health	75.3	71.9 to 78.7	76.0	72.0 to 79.9	0.0
Vitality	65.3	61.3 to 69.3	62.6	50.4 to 74.9	0.1
Social function	80.0	74.9 to 85.0	86.1	80.8 to 91.4	-0.3

Table V. Comparison of change scores for different health-status assessments according to appraisal of change by the patient, and remaining scope for change as measured by patient satisfaction and one transition item

Question 1 'How successful do you feel your shoulder treatment has been?'			
Test	Very successful Mean change score (n = 23)	All other responses Mean change score (n = 37)	t-test p value
12-item questionnaire (n = 60)	14.5	4.6	3.2*
Constant (n = 15)	-0.3	-2.2	NS
Rowe (n = 15)	-45.0	-27.5	NS
SF36 (n = 60)			
Health perceptions	-5.8	2.0	NS
Health change	-8.7	-2.7	NS
Physical activity	-4.8	-1.1	NS
Pain	-9.7	-9.6	NS
Role limitation - physical	-30.4	-14.9	NS
Role limitation - mental	-1.4	-0.9	NS
Mental health	-0.5	-0.2	NS
Vitality	-5.7	-2.6	NS
Social function	-7.2	-5.1	NS
Question 2 'Do you feel that there is still 'room for improvement' in your shoulder?'			
	Little or none Mean change score (n = 26)	Moderate or great Mean change score (n = 36)	t-test p value
12-item shoulder score (n = 62)	14.8	3.5	-4.0‡
Constant (n = 15)	-9.4	3.9	NS
Rowe (n = 15)	-49.0	-24.5	NS
SF36 (n = 62)			
Health perceptions	-5.9	4.4	2.6‡
Health change	-11.5	-1.4	2.2‡
Physical activity	-8.3	-1.8	2.9*
Pain	-11.1	-9.3	NS
Role limitation - physical	-36.5	-8.3	2.6‡
Role limitation - mental	-7.7	6.7	NS
Mental health	-2.6	1.7	NS
Vitality	-0.6	8.1	NS
Social function	-7.3	-3.7	NS
Transition item: 'How have the problems related to your shoulder changed following treatment?'			
	Much better Mean change score (n = 33)	All other responses Mean change score (n = 27)	t-test
12-item shoulder score (n = 60)	14.2	1.3	4.6†
Constant (n = 15)	-6.4	10.3	NS
Rowe (n = 15)	-38.0	-18.8	NS
SF36 (n = 60)			
Health perceptions	-5.0	3.9	-2.4‡
Health change	-8.3	-0.9	NS
Physical activity	-5.5	0.2	NS
Pain	-13.8	-4.5	NS
Role limitation - physical	-33.3	-5.6	-2.6‡
Role limitation - mental	-9.1	9.0	NS
Mental health	-2.1	1.8	NS
Vitality	-0.5	8.9	NS
Social function	-6.1	-5.8	NS

* p < 0.01

† p < 0.001

‡ p < 0.05

for the patients. The items are internally consistent and reproducible, and therefore the questionnaire may be considered to be at least as reliable as clinical scores used to assess outcomes.²⁴

Every effort was made to derive the items to be included in the questionnaire from exploratory interviews with patients, rather than by imposing clinical assumptions. Draft versions of the questionnaire were tested on

patients and the final content only agreed when patients understood it and felt that no important items had been omitted. Construct validity was tested by examining the level of agreement with the clinical data and the SF36 assessment. All correlations were in the expected direction with poor scores on our shoulder questionnaire corresponding to poorer scores for the Constant, Rowe and the SF36 assessments. In the case of the last, correlations

were highest in the measurement of pain and physical function.

Responsiveness, or the sensitivity to clinically important change, is least likely to be examined, despite its being of the greatest importance in any form of prospective outcome study.²⁵ This point was clearly demonstrated when the sensitivity to change of our questionnaire was compared with that from the Constant and Rowe scores. The Constant score, which appeared to agree well with other assessments at the pretreatment stage, performed poorly when measuring change in the condition of the patients. This represents additional evidence in support of a previous report suggesting that Constant assessment is not appropriate to measure outcome in patients with shoulder instability,¹⁹ probably because it contains no elements which relate directly to this. Instead, it places undue weight on concerns such as pain and power which are frequently not relevant to patients with instability. In comparison, the Rowe assessment outperformed all other measures in detecting change although the numbers were rather small.

The standardised effect size was much higher for our questionnaire than with the SF36. The new questionnaire was much better at distinguishing between subgroups of patients, based on their own assessment of the change in their shoulder after either physiotherapy or surgery. This may be because the SF36 was not developed to be used specifically in relation to outcomes of surgical treatment. Both clinical measurements lacked adequate numbers at follow-up to allow useful conclusions to be drawn from comparisons of subgroups.

Measurements such as we have carried out assess the outcome from the point of view of the patient. Such methods are intended to supplement and not replace conventional measures of outcome. It is becoming increasingly clear that systematic studies are required to examine effects on outcome of alternative methods of treatment and operations in the management of orthopaedic conditions.^{26,27} Because variations in outcomes may be quite modest, such studies will need large sample sizes to detect differences and will therefore have to be multicentred. Standardised patient-based measurements of outcome have obvious advantages, particularly if they are highly sensitive to clinical change. A particularly long period of follow-up is required to detect differences in outcome for many orthopaedic conditions. Short uncomplicated questionnaires which can be delivered by post may make such long-term studies more feasible. The shoulder questionnaire reported here is intended for such use.

We acknowledge financial support by grant from Oxford Regional Health Authority. We would also wish to thank the nursing staff in the outpatient clinics, as well as 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. Carr AJ. Biomechanics of shoulder stability. *Current Orthop* 1996;10: 146-50.
2. Walch G. Chronic anterior glenohumeral instability. *J Bone Joint Surg [Br]* 1996;78-B:670-7.
3. Emery RJ, Mullaji AB. Glenohumeral joint instability in normal adolescents: incidence and significance. *J Bone Joint Surg [Br]* 1991; 73-B:406-8.
4. Vermeiren J, Handelberg F, Casteleyn PP, Opdecam P. The rate of recurrence of traumatic anterior dislocation of the shoulder: a study of 154 cases and a review of the literature. *Int Orthop* 1993;17:337-41.
5. Marcacci M, Zaffagnini S, Petitto A, et al. Arthroscopic management of recurrent anterior dislocation of the shoulder: analysis of technical modifications on the Caspari procedure. *Arthroscopy* 1996; 12:144-9.
6. MacDonald DA. The shoulder and elbow. In: Pynsent PB, Fairbank JC, Carr A, eds. *Outcome measures in orthopaedics*. Butterworth Heinemann, Oxford, 1993:144-73.
7. Wright JG, Rudicel S, Feinstein AR. Ask patients what they want: evaluation of individual complaints before total hip replacement. *J Bone Joint Surg [Br]* 1994;76-B:229-34.
8. Amadio PC. Editorial. Outcomes measurement: more questions; some answers. *J Bone Joint Surg [Am]* 1993;75-A:1583-4.
9. 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.
10. Patrick DL, Deyo RA. Generic and disease-specific measures in assessing health status and quality of life. *Med Care* 1989;27:(3 Suppl):217-32.
11. Katz JN, Larson MG, Phillips CB, Fossel AH, Liang MH. Comparative measurement sensitivity of short and longer health status instruments. *Med Care* 1992;30:917-25.
12. Fitzpatrick R, Ziebland S, Jenkinson C, Mowat A, Mowat A. A comparison of the sensitivity to change of several health status instruments in rheumatoid arthritis. *J Rheumatol* 1993;20:429-36.
13. 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;4:224-31.
14. 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.
15. Cox D, Fitzpatrick R, Fletcher A, et al. Quality of life assessment: can we keep it simple? *J Roy Stat Soc Series A* 1992;155:353-93.
16. Cronbach L. Coefficient alpha and the internal structure of tests. *Psychometrika* 1951;16:297-334.
17. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1986;i: 307-10.
18. Constant CR, Murley AHG. A clinical method of functional assessment of the shoulder. *Clin Orthop* 1987;214:160-4.
19. Conboy V, Morris R, Kiss J, Carr A. An evaluation of the Constant-Murley shoulder assessment. *J Bone Joint Surg [Br]* 1996;78-B: 229-32.
20. Rowe CR, Patel D, Southmayd WW. The Bankart procedure: a long-term end-result study. *J Bone Joint Surg [Am]* 1978;60-A:1-16.
21. 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.
22. Kazis LE, Anderson JJ, Meenan RF. Effect sizes for interpreting changes in health status. *Med Care* 1989;27:178-89.
23. Ziebland S, Fitzpatrick R, Jenkinson C, Mowat A, Mowat A. Comparison of two approaches to measuring change in health status in rheumatoid arthritis: the health assessment questionnaire (HAQ) and modified HAQ. *Ann Rheum Dis* 1992;51:1202-5.
24. Koran LM. The reliability of clinical methods, data and judgements. *N Engl J Med* 1975;293:642-6.
25. Guyatt G, Walter GS, Norman G. Measuring change over time: assessing the usefulness of evaluative instruments. *J Chronic Dis* 1987;40:171-8.
26. Murray DW, Carr AJ, Bulstrode C. Survival analysis of joint replacements. *J Bone Joint Surg [Br]* 1993;75-B:697-704.
27. 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.