

EXTENDED REPORT

A comparison of four shoulder-specific questionnaires in primary care

A Paul, M Lewis, M F Shadforth, P R Croft, D A W M van der Windt, E M Hay

Ann Rheum Dis 2004;**63**:1293–1299. doi: 10.1136/ard.2003.012088

See end of article for authors' affiliations

Correspondence to: Professor E M Hay, Primary Care Sciences Research Centre, Keele University, Keele ST5 5BG, UK; e.m.hay@keele.ac.uk

Accepted 21 October 2003

Objectives: To compare the validity, responsiveness to change, and user friendliness of four self completed, shoulder-specific questionnaires in primary care.

Methods: A cross sectional assessment of validity and a longitudinal assessment of responsiveness to change of four shoulder questionnaires was carried out: the Dutch Shoulder Disability Questionnaire (SDQ-NL); the United Kingdom Shoulder Disability Questionnaire (SDQ-UK); and two American instruments, the Shoulder Pain and Disability Index (SPADI) and the Shoulder Rating Questionnaire (SRQ). 180 primary care consultants with new shoulder region pain each completed two of the questionnaires, as well as EuroQoL and 10 cm visual analogue scales (VAS) for overall pain and difficulty due to the shoulder problem. Each participant was assessed by a standardised clinical schedule. Postal follow up at 6 weeks included baseline measures and self rated assessment of global change of the shoulder problem (seven point Likert scale).

Results: Strongest correlations were found for SDQ-UK with EuroQoL 5 score, and for SPADI and SRQ with shoulder pain and difficulty VAS. All shoulder questionnaires correlated poorly with active movement at the painful shoulder. SPADI and SRQ performed better on ROC analysis than SDQ-NL and SDQ-UK (areas under the curve of 0.87, 0.85, 0.77, and 0.77, respectively). However, SRQ scores changed significantly over time in stable subjects.

Conclusions: Cross sectional comparison of the four shoulder questionnaires showed they had similar overall validity and patient acceptability. SPADI and SRQ were most responsive to change. Additionally, SPADI was the quickest to complete and scores did not change significantly in stable subjects.

Most shoulder region pain is seen and managed in primary care.¹ Our understanding about the natural history and optimal treatment for shoulder complaints has been hampered, however, by the use of different outcome measures, including a variety of shoulder-specific questionnaires.^{2–4} It is therefore important to identify a preferred shoulder questionnaire to facilitate the consistent use of outcome measures and meaningful comparison of results in primary care based cross sectional and longitudinal studies.

Since 1990, a number of groups have developed self administered shoulder pain and disability questionnaires.^{5–13} Many were developed and tested in secondary orthopaedic care settings and were primarily intended for measuring outcomes after surgical procedures. To date, the validity and responsiveness of shoulder questionnaires has not been compared in primary care.

We describe the results from a prospective study comparing the validity, responsiveness, and acceptability of four shoulder pain and disability questionnaires when used to assess primary care consultants with new onset shoulder region pain.

METHODS

Selection of shoulder questionnaires

Self completed, non-disease-specific shoulder questionnaires available in English and published in peer reviewed journals since 1990 were identified by a Medline search in 1999, augmented by citation checking. These were the Shoulder Pain and Disability Index (SPADI),⁵ Simple Shoulder Test (SST),⁶ United Kingdom Shoulder Disability Questionnaire (SDQ-UK),⁷ American Shoulder and Elbow Surgeon's Shoulder Assessment Form (M-ASES),⁸ Oxford Shoulder Score (OSC),⁹ Subjective Shoulder Rating System (SSRS),¹⁰

Shoulder Rating Questionnaire (SRQ),¹¹ and Dutch Shoulder Disability Questionnaire (SDQ-NL).^{12–13}

Consensus among three authors (AP, DAWMvdW, and EMH) determined the selection of four shoulder questionnaires based on their suitability for use in primary care, as well as their face and content validity (table 1). The questionnaires were also selected to include a variety of time scales (for example, past day *v* past month), response options (for example, binomial *v* visual analogue scale), and scoring methods (for example, weighting *v* non-weighting). The questionnaires selected were the SDQ-NL, SDQ-UK, SPADI, and SRQ (table 2).

SDQ-NL is a 16 item questionnaire.^{12–14} Each question refers to the past 24 hours and has three responses (yes, no, and not applicable). Items refer only to pain related disability. Its properties have been tested in physiotherapy, primary and secondary care. SDQ-UK is a 23 item questionnaire.^{7–15} Each question refers to "today" and has three responses (yes, no, and not applicable). Non-applicable items are included in the final score, in contrast to SDQ-NL scoring. It includes items in the domains of pain, daily activities, sports/pastimes, and work/housework. Its cross sectional validity has been assessed for primary care and community subjects. SPADI is a 13 item questionnaire.^{5–16–20} The pain domain consists of five questions and the disability domain consists of eight. Each question refers to the past week. We used the version with 12 segment visual analogue scale (VAS) responses

Abbreviations: EQ, EuroQoL, ES, effect size; ROC, receiver operating characteristics; SDQ-NL, Dutch Shoulder Disability Questionnaire; SDQ-UK, United Kingdom Shoulder Disability Questionnaire; SPADI, Shoulder Pain and Disability Index; SRM, standardised responsiveness mean; SRQ, Shoulder Rating Questionnaire; TS, thermometer score; VAS, visual analogue scale

Table 1 Comparison of questionnaire contents

	SDQ-NL	SDQ-UK	SPADI	SRQ
Overall effect	–	–	–	1
Pain	16	4	5	4
Daily activities	–	12	8	6
Sports/pastimes	–	2	–	3
Work/housework	–	1	–	4
Other	–	4	–	–

Numbers refer to the number of questions in each category.

for this study.³ The validity and responsiveness to change of SPADI have been described in physiotherapy, walk in centres, and secondary care settings. SRQ is an 18 item questionnaire in five domains.¹¹ The first domain is a 10 cm VAS for current overall shoulder symptoms. Other domains (pain, daily activities, sports/recreation, and work) refer to the past month and are rated on five point rating scales. This questionnaire was developed and tested for validity and responsiveness to change in an orthopaedic setting.

Some differences between domains covered by the four shoulder questionnaires are worthy of note. SDQ-NL items refer solely to pain related complaints. Only SRQ and SDQ-UK contain items regarding sports and pastimes, but SRQ alone includes items regarding work. SPADI does not refer to sleep disturbance, although it does have an item regarding pain on lying on the affected side.

Study population

Two primary care groups were invited to refer consultants with a new episode of shoulder pain, aged 18 years and above, to a community based research clinic (catchment population approximately 200 000). Shoulder region pain was defined as “pain in the shoulder region brought on or exacerbated by movement at that shoulder”. A new episode was defined as “a consultation for a shoulder problem, where the subject had not consulted primary care for a similar problem in the same shoulder within the past 6 months”. The exclusion criteria listed on the primary care referral proforma included suspected or known inflammatory arthropathy; malignancy; polymyalgia rheumatica; fracture of neck or shoulder; subluxation or dislocation of the shoulder; pain of visceral origin. All subjects seen in the research clinic were invited to participate in the study, unless they were unable to complete the assessment packs or not available for follow up.

The study was approved by North Staffordshire research ethics committee and all participants gave written informed consent.

Baseline assessment

Baseline assessment consisted of self completed measures, observer rated shoulder measures, and an assessment of the acceptability of the shoulder questionnaires. In allocating shoulder questionnaires to individual patients, we wished to (a) compare scores of different shoulder questionnaires within individual patients; (b) compare shoulder questionnaire scores of groups of subjects with external constructs; (c) minimise participant burden in order to maximise effective follow up; (d) minimise any order effect; and (e) minimise differences between patient groups.

This was achieved as follows. The selection of four shoulder questionnaires provided six pairings (SDQ-UK+SDQ-NL, SPADI+SRQ, SDQ-NL+SRQ, SDQ-UK+SPADI, SDQ-UK+SRQ, and SDQ-NL+SPADI) to be allocated to every six consecutive subjects. Alternate groups of six participants completed their pair of shoulder questionnaires in reverse order. Each of the six pairs of questionnaires were, therefore, completed by 30 subjects. Thus, each individual questionnaire was completed by 90 participants and 180 participants were recruited in total.

Each participant was also asked to complete EuroQoL (EQ) and 10 cm VAS of current overall pain and difficulty due to the shoulder problem. EQ has two components: EQ5 score (poorest health state score is –0.59 and a score of 1 indicates full health) and the VAS thermometer score (TS) (0 is worst imaginable health and 100 is best imaginable health).²¹

After completing each shoulder questionnaire, participants used a four point Likert scale to rate ease of completion and a six point scale to rate how relevant they perceived the questionnaire to be to their shoulder problem. The study nurse timed completion of the shoulder questionnaires. Participants were then assessed using a standardised clinical schedule by a research fellow (AP), who did not know the results of the questionnaires. The clinical assessment included measurement of active range of movement at the painful shoulder. Shoulder abduction, flexion, and extension were measured with a plurimeter V inclinometer (a gravity referenced inclinometer designed and provided by Dr Rippstein, Zurich, Switzerland). Shoulder internal rotation was rated by visual estimation, and shoulder external rotation was measured using a universal goniometer.

Subsequent treatment of the shoulder region pain, including advice, analgesia, physiotherapy, steroids, and local anaesthetic injections, was based on clinical findings and was not part of the study protocol.

Follow up

After 6 weeks, postal follow up included the same two shoulder questionnaires from each individual patient's baseline assessment, 10 cm shoulder pain and difficulty

Table 2 Comparison of questionnaire scoring used in the study

	SDQ-NL	SDQ-UK	SPADI	SRQ†
Total number of questions	16	23	13	18
Response options	Binary	Binary	VAS (0–11)	VAS × 1 (0–100)
“Not applicable” option	Yes	Yes	Yes	5 Point rating scale × 17 Yes (sports+recreation/work questions‡)
Use of domains	–	–	Yes	Yes
Weighting of domains	–	–	–	Yes
Total or domain score(s) calculated relative to total applicable score	Yes	–	Yes*	Yes
Range of subtotal scores if all questions applicable	0–16	0–23	0–100	17–100
Range of total scores	0–100	0–100	0–100	0–100

*If more than two items in a domain are “not applicable”, the domain score is zero; †scoring was reversed so that scores increased with increasing shoulder problems; ‡“not applicable” option for sports/recreation added for this study; work questions only to be completed by those in paid work, students, and home carers.

Table 3 Patient characteristics at baseline assessment

	Total (n = 180)	Questionnaire pairs (6 pairs×30 subjects)					SPADI and SRQ
		SDQ-NL and SDQ-UK	SDQ-NL and SPADI	SDQ-NL and SRQ	SDQ-UK and SPADI	SDQ-UK and SRQ	
Age‡	53.5 (19–85)	56.5	51.0	54.5	56.0	51.0	54.5
Male sex†	90 (50)	15	15	14	14	16	16
BMI (kg/m ²)‡	27.3 (19.9–41.2)	27.3	27.2	26.9	27.5	27.9	27.1
Employed†	99 (55)	18	17	16	15	16	17
Manual social class†	65 (36)	13	16	9	8	10	9
Use of painkillers†	108 (60)	25	18	15	20	15	15
Both sides affected†	16 (9)	1	5	4	3	3	0
Dominant side affected†	92 (51)	14	14	12	16	16	20
Overall pain VAS‡	60.0 (5–100)	51.5	54.5	66.0	62.0	47.5	63.0
Overall difficulty VAS‡	55.0 (0–100)	52.0	57.0	64.5	52.5	45.5	59.5
Abduction ‡	144.0 (22–184)	170.0	103.0	129.0	123.0	155.0	159.0
Flexion ‡	153.0 (24–182)	161.0	137.0	135.0	157.0	156.0	156.0
Extension ‡	48.0 (10–82)	50.0	40.0	44.0	47.0	54.0	48.0
Internal rotation ≥110*†	125 (69)	23	15	19	20	26	22
External rotation ‡	45.0 (0–80)	45.0	45.0	40.0	52.5	50.0	47.5
EQ5 score‡	0.66 (–0.18–1.00)	0.66	0.66	0.66	0.59	0.62	0.61
EQ thermometer score‡	70.0 (2–100)	80.0	55.0	60.0	60.0	70.0	79.5

* $p < 0.05$ by statistical test of between-paired groups comparisons (by χ^2 or Kruskal-Wallis tests); numbers are †counts (percentages for total sample); ‡median (range for total sample).

VAS, EQ, and a patient's global assessment of change of the shoulder problem (totally recovered = 1, moderately better = 2, slightly better = 3, same = 4, slightly worse = 5, moderately worse = 6, much worse = 7). Participants not returning their follow up packs received a postal reminder after 2 weeks and a phoned reminder after a further 2 weeks. Non-responders at this stage were replaced within the study.

Analysis

Shoulder questionnaire scores were standardised to a 0–100 scale across all four questionnaires, with scores increasing with increasing shoulder pain and disability. Cross sectional validity at baseline was investigated as follows: (a) questionnaire scores were correlated within pairs; (b) questionnaire scores were correlated with external standards. The external standards were observer rated active range of movement at the painful shoulder, a self completed generic measure (EQ), and self completed, single item, shoulder-specific measures (overall shoulder pain and difficulty VAS).

Internal and external responsiveness were assessed.²² Internal responsiveness characterises the ability of a questionnaire to change over time and external responsiveness

compares change in scores with the patients' global assessment of change (in this study, a seven point Likert scale). "Improved" subjects were defined as those who rated themselves as totally recovered to slightly better (groups 1–3), "stable" subjects as unchanged (group 4), and "worse" subjects as slightly to much worse (groups 5–7). Stable and worse subjects were also combined into one "not improved" category to enable receiver operator characteristic (ROC) analysis. Internal responsiveness was analysed in four ways:

- Baseline and follow up scores were compared using paired *t* tests. Improved and stable subjects were analysed separately.
- Effect size (ES), defined as the difference between means (baseline and follow up) in improved subjects divided by the standard deviation of baseline values in improved subjects, was determined.²³ The ES is large if >0.8 , moderate for 0.5–0.8, and small for 0.2–0.5.
- Standardised responsiveness mean (SRM), defined as the mean change score of improved subjects divided by the standard deviation of the change score in improved

Table 4 Correlation matrix of associations between the four shoulder questionnaires with (a) each other; (b) other measures, including range of shoulder movement, generic, and VAS shoulder-specific measures

	SDQ-NL	SDQ-UK	SPADI	SRQ
SDQ-NL†	–	0.552**	0.330	0.432*
SDQ-UK†	0.552**	–	0.573**	0.716**
SPADI†	0.330	0.573**	–	0.829**
SRQ†	0.432*	0.716**	0.829**	–
Abduction ‡	–0.390**	–0.342**	–0.251*	–0.440**
Flexion ‡	–0.399**	–0.341**	–0.231*	–0.382**
Extension ‡	–0.439**	–0.225*	–0.090	–0.300**
Internal rotation ‡	–0.317**	–0.139	–0.209*	–0.264*
External rotation ‡	–0.213*	–0.020	–0.232*	–0.233*
EQ5 score‡	–0.432**	–0.680**	–0.473**	–0.498**
EQ TS‡	–0.449**	–0.448**	–0.288**	–0.427**
Pain VAS‡	0.480**	0.496**	0.656**	0.616**
Difficulty VAS‡	0.465**	0.413**	0.619**	0.598**

Values are Spearman's correlation coefficients (r_s).

* $p < 0.05$; ** $p < 0.01$ according to the value of r_s ; †n = 30; ‡n may be <90 owing to some missing data.

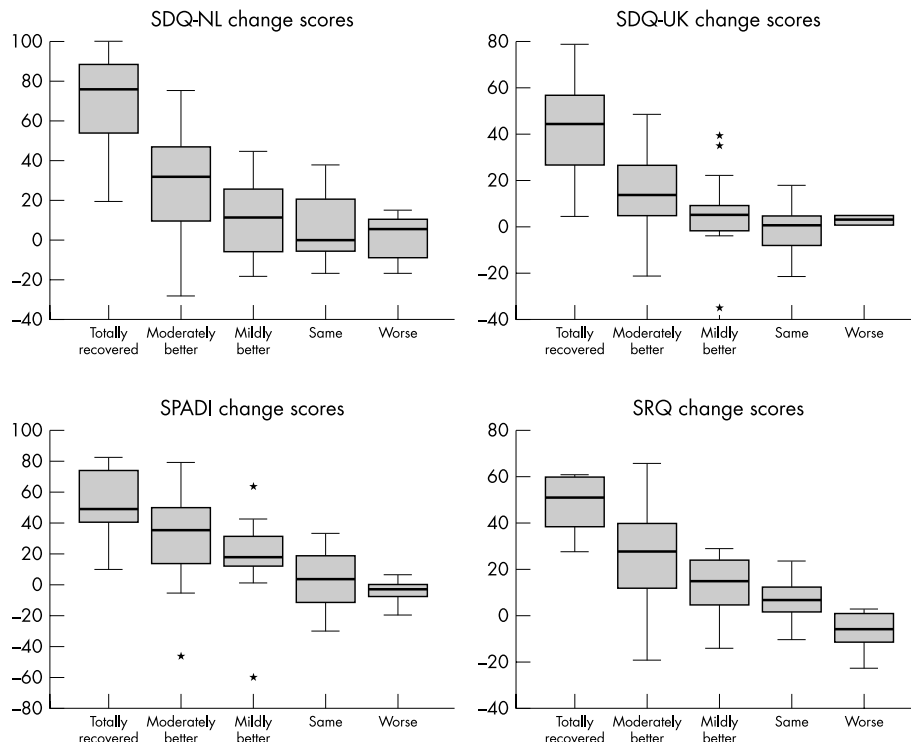


Figure 1 Box plots of the distribution of change scores for the shoulder questionnaires in relation to categories of self rated change.

subjects, was determined.²⁴ Interpretation of the SRM was as for the effect size.

- Responsiveness ratio (RR), defined as the difference between means (baseline and follow up) in improved subjects divided by the standard deviation of change in stable subjects, was determined.¹²⁻²⁵ A value >1 indicates a responsiveness, which is proportional to the magnitude of the responsiveness ratio.

External responsiveness was measured in two ways:

- Correlation of change in shoulder questionnaire scores with the self rated global assessment of change, using Spearman’s coefficient was determined.
- ROC analysis, performed by plotting sensitivity to change on the y axis and 1 – specificity on the x axis for all possible cut off values of the questionnaires against the patients’ global assessment of improved or not improved, was carried out.²⁶⁻²⁷ An area under the curve of 0.5 is expected by chance and a value of 1.0 indicates maximal

responsiveness. By examination of the intersections of the sensitivity and 1 – specificity plots nearest the upper left hand corner of the graph, the optimal cut off value for maximal average sensitivity and specificity for detecting change could be identified.

Patient acceptability of the shoulder questionnaires was assessed by considering levels of missing data, time taken to complete, ratings of ease of completion, and whether the shoulder questionnaires were relevant to their shoulder problem (Kruskal-Wallis test). Investigator acceptability was assessed subjectively by the first and second authors.

Statistical analysis was carried out using SPSS version 11.0 for Windows. Statistical significance was based on a two tailed significance level of $\alpha = 0.05$.

RESULTS

Study group

During 1999 and 2000, 237 patients with a new episode of shoulder pain were referred to the shoulder clinic.

Responsiveness measures	SDQ-NL	SDQ-UK	SPADI	SRQ
<i>Internal</i>				
t Test—“improved” (p value)	<0.001	<0.001	<0.001	<0.001
t Test—“stable” (p value)	0.549	0.633	0.553	0.013
ES	1.56	0.91	1.52	1.64
SRM	0.95	0.78	1.17	1.23
RR	1.73	1.39	1.67	2.76
<i>External</i>				
Correlation with self rated change	0.58	0.54	0.61	0.68
AUC (95% CI)	0.77 (0.68 to 0.87)	0.77 (0.67 to 0.88)	0.87 (0.79 to 0.95)	0.85 (0.77 to 0.93)

ES, effect size; SRM, standardised responsiveness mean; RR, Guyatt’s responsiveness ratio; correlation with self rated change, calculated using Spearman’s correlation; AUC, area under the receiver operator characteristic curve for “improved” versus “non-improved” subjects.

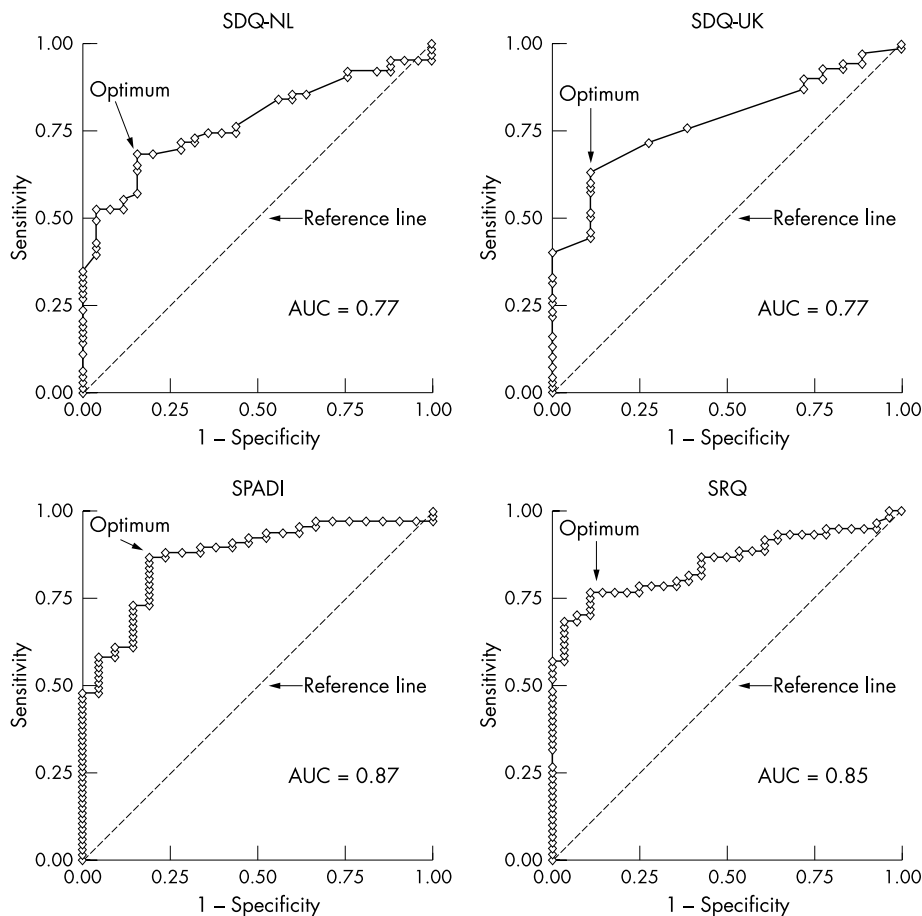


Figure 2 ROC curves of the shoulder questionnaires against self rated change (improved or not improved).

Non-participants included 19 subjects who failed to attend the research clinic, 10 who did not fulfil the study criteria, 12 who declined to participate, and 16 who initially agreed to participate but did not return their follow up packs. The median age of the 180 participants was 53.5 years (range 19–85) and 90 (50%) were female. Participants in the six groups of questionnaire pairs had similar demographic and clinical characteristics (table 3). Non-participants were slightly younger (median age 48.0 years) and a slightly higher percentage were male (55%).

Validity

Correlation between shoulder questionnaires

At baseline, the highest correlations were between SRQ and the other questionnaires, the lowest between SDQ-NL and the other questionnaires (table 4).

Correlation between shoulder questionnaires and observer rated measures

Overall, there were weak correlations between shoulder questionnaire scores and active shoulder movement ($r_s = -0.02$ to -0.44) (table 4). Correlations were highest with abduction and flexion and lowest with shoulder rotation.

Correlation between shoulder questionnaires and other self rated measures

At baseline, all shoulder questionnaires correlated significantly with EQ scores and overall shoulder pain and difficulty VAS (table 4). The strongest correlations were observed for SDQ-UK with EQ5 score, and for SPADI and SRQ with overall shoulder pain and difficulty VAS.

Responsiveness to change

At 6 weeks’ follow up, 19 subjects (11%) reported total recovery, 79 (44%) were moderately and 32 (18%) were mildly better, 29 (16%) were the same, 4 (2%) were mildly worse, 11 (6%) moderately worse, 2 (1%) much worse, and 4 (2%) results were missing. Hence, 130 (72%) were classified as improved and 46 (26%) as not improved. Figure 1 plots the distribution of questionnaire change scores for categories of self rated change. Mean change scores for improved and stable subjects were 29.2 and 2.8 for SDQ-NL, 16.5 and -1.6 for SDQ-UK, 31.3 and 3.3 for SPADI, and 25.1 and 5.9 for SRQ. Table 5 shows the results of tests of responsiveness.

Internal responsiveness

For the improved group, paired *t* tests for differences between baseline and follow up scores were highly significant for all shoulder questionnaires. Paired *t* tests for stable subjects showed no significant difference ($p > 0.05$) for SDQ-NL, SDQ-UK, and SPADI. However, a significant difference was found for SRQ ($p = 0.013$).

All four shoulder questionnaires showed moderate or large ES, SRM, and responsiveness ratios, although SDQ-UK was consistently worst across all three tests.

External responsiveness

SDQ-UK had the lowest correlation value with self rated change of the shoulder problem ($r_s = 0.54$) and SRQ had the highest ($r_s = 0.68$).

Using ROC curves (fig 2), the highest areas under the curve were recorded for SPADI and SRQ. Optimal cut off points, above which any improvement could be identified, were an

Table 6 Comparison of shoulder questionnaire attributes at baseline assessment

	SDQ-NL	SDQ-UK	SPADI	SRQ
Time to complete (min)*‡	2.0 (1.0–8.0)	3.0 (1.0–8.0)	2.0 (1.0–4.0)	4.0 (1.0–16.0)
Reflects shoulder problem (Likert scale 0–5)*‡§	5.0 (2.0–5.0)	4.0 (1.0–5.0)	5.0 (1.0–5.0)	4.0 (0.0–5.0)
Missing data†				
0	83 (92)	80 (90)	84 (93)	80 (89)
1	3 (4)	7 (8)	4 (5)	3 (3)
2	2 (2)	0 (0)	1 (1)	1 (1)
≥3	2 (2)	2 (2)	1 (1)	6 (7)
Ease of completion†				
Very easy	71 (79)	62 (69)	61 (69)	57 (64)
Moderately easy	19 (21)	26 (29)	27 (30)	32 (36)
Moderately difficult	0 (0)	2 (2)	1 (1)	0 (0)
Very difficult	0 (0)	0 (0)	0 (0)	0 (0)

* $p < 0.05$ by statistical test of between-groups comparisons (by the Kruskal-Wallis test); n may be < 90 owing to some missing data; numbers are †counts (percentages), ‡median (range); §Likert scale: 0 = definitely not, 5 = definitely did.

improvement of 14 out of 100 (2–3 out of 16 items) for SDQ-NL, 4–8 (1–2 out of 23 items) for SDQ-UK, 8 for SPADI, and 13 for SRQ.

Acceptability

On average, each shoulder questionnaire took less than 5 minutes to complete (table 6). SPADI was the quickest shoulder questionnaire to complete and SRQ took the longest ($p < 0.001$). Participants rated SDQ-NL as best, on a 0 to 5 scale, for relevance to their shoulder problem ($p = 0.047$). Levels of missing data were low for all shoulder questionnaires and participants generally found them easy to complete. SRQ was the most difficult and time consuming to score. The SDQ-UK was the easiest and quickest to score, followed by SDQ-NL and SPADI.

Table 7 summarises the relative properties of the shoulder questionnaires.

DISCUSSION

We report a study specifically designed to compare the validity, responsiveness to change, and user friendliness of four self completed shoulder questionnaires in subjects presenting in primary care. We selected primary care consultants with a new episode of shoulder region pain, who required further management, for the study. Such patients are likely to be representative of those included in outcome and intervention studies of shoulder problems in primary care. Referral to our research clinic was maximised by research team contact with the primary care physicians involved, reminders to participating practices, ease of referral, and prompt assessment of patients in local hospitals.

The four shoulder questionnaires selected have face and content validity for the assessment of shoulder pain and disability. The correlation of shoulder questionnaire scores with each other can be interpreted as a confirmation of these

properties. The lower correlations of SDQ-NL with the others is perhaps a reflection of the single domain of pain measured by this instrument.

In the absence of a true “gold standard” against which to assess criterion validity, we compared the shoulder questionnaires with external constructs likely to reflect the impact of shoulder problems. The significant correlation of shoulder questionnaire scores with overall shoulder pain and difficulty VAS reassures us that the shoulder questionnaires are indeed reflecting pain and disability due to the affected shoulder. Of note, SPADI and SRQ correlation coefficients were the strongest.

The further testing of construct validity was based on the hypotheses that (a) EQ, a generic instrument, would measure the same domains of pain, wellbeing, and ability to perform tasks as the shoulder questionnaires and (b) objective measures of shoulder movement would be influenced by pain in that region and would influence the ability to perform tasks. Shoulder questionnaire scores were significantly correlated with general health measured by the EQ5 score and TS. In particular, SDQ-UK correlated well with the EQ5 score. In contrast, range of shoulder movement did not correlate well with shoulder questionnaire scores. This may be due to several factors. Firstly, the shoulder can be painful in the presence of a full range of movement. Secondly, many day to day arm activities can be performed with hands below shoulder level, therefore, shoulder restriction may need to be severe before notable disability is present. Alternatively, as highlighted by repeatability studies, we may not be accurately measuring active range of movement at the painful shoulder.²⁸ Finally, the shoulder questionnaires tested may simply not be sensitive to disability resulting from restricted shoulder movement.

The testing of responsiveness to change is similarly hampered by the lack of an external “gold standard” of relevant change. For the purposes of this study, we selected patients’ global assessment of change in shoulder symptoms

Table 7 Relative properties of the shoulder questionnaires

	SDQ-NL	SDQ-UK	SPADI	SRQ
Content validity	+	++	++	+++
Construct validity	+	++	+++	+++
Responsiveness to change over time	++	+	+++	+++
Stable to no change over time	+	++	+	–
Participant acceptability	+++	++	++	+
Ease of scoring	+++	++++	++	+

– not satisfactory; + satisfactory; ++ good; +++ very good; ++++ excellent.

as our external comparator. All tests of responsiveness indicated that the shoulder questionnaires were at least moderately able to detect true change over time. SDQ-UK was the most stable in subjects who rated themselves as unchanged, SRQ was the most responsive overall to true change, whereas SPADI performed best in relation to ROC analysis. SRQ was unstable in subjects reporting no change in their shoulder problem, detected by comparing baseline with follow up scores of stable subjects. Additionally, the SRM of the SRQ in stable subjects (mean SRQ change score in stable subjects/standard deviation of that change) was of moderate size. SRQ should therefore probably be considered inappropriate for use in longitudinal observational studies and trials.

The differences in responsiveness between the shoulder questionnaires have important implications for the calculation of sample size and power estimates for clinical studies of shoulder problems. Power calculations using the responsiveness estimates from this study would result in substantial differences in study sample sizes, depending on the questionnaire chosen and the predicted difference in improvement between study groups.

Most participants found the shoulder-specific questionnaires easy to complete and relevant to their shoulder problem. A small proportion of subjects were unable to complete the questionnaires because of problems with comprehension, reading, feeling generally too unwell, or the shoulder symptoms themselves. SPADI and SDQ-NL were the quickest to complete and all SPADI questionnaires were completed within 5 minutes.

We recommend that future studies of shoulder pain and disability in primary care should use a core of health measures to enable comparison between studies and data pooling, including a self completed shoulder-specific questionnaire, 10 cm VAS scores of pain and difficulty due to the affected shoulder, and a generic health measure. The choice of which shoulder questionnaire to use will depend on the purpose for which it is required (for example, cross sectional v longitudinal study) and practical considerations (for example, time to complete and ease of scoring). Owing to its combined validity, responsiveness to true change, and acceptability, the SPADI appears to be the preferred shoulder-specific questionnaire for assessing shoulder problems presenting in primary care.

ACKNOWLEDGEMENTS

We thank participating patients and practices of South and North Stoke Primary Care Groups, the study nurses, and Kathryn Jones for her help with preparation of the manuscript.

AP was funded by the Haywood Rheumatism Research and Development Foundation. The study was funded by the North Staffordshire Primary Care Consortium using NHS(E) R&D Budget 1 funds.

Authors' affiliations

A Paul, M F Shadforth, E M Hay, Staffordshire Rheumatology Centre, The Haywood, Stoke on Trent, ST6 7AG, UK

M Lewis, P R Croft, E M Hay, Primary Care Sciences Research Centre, Keele University, Keele, ST5 5BG, UK

D A W M van der Windt, Institute for Research in Extramural Medicine, Department of General Practice, VU University Medical Centre, Amsterdam, The Netherlands

REFERENCES

- Felson DT**, Meenan RF, Dayno SJ, Gertman P. Referral of musculoskeletal disease patients by family and general practitioners. *Arthritis Rheum* 1985;**28**:1156-62.
- Croft P**. Measuring up to shoulder pain. *Ann Rheum Dis* 1998;**57**:65-6.
- Green S**, Buchbinder R, Glazier R, Forbes A. Systematic review of randomised controlled trials of interventions for painful shoulder: selection criteria, outcome assessment, and efficacy. *BMJ* 1998;**316**:354-60.
- Michener LA**, 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.
- Roach KE**, Budiman-Mak E, Songsiridej N, Lertratanakul Y. Development of a shoulder pain and disability index. *Arthritis Care Res* 1991;**4**:143-9.
- Lippitt SB**, Harryman DT, Matsen III FA. A practical tool for evaluating function: the simple shoulder test. In: Matsen III FA, Fu FH, Hawkins RJ, eds. *The shoulder: a balance of mobility and stability*. Rosemont, Illinois: American Academy of Orthopaedic Surgeons, 1993:501-18.
- Croft P**, Pope D, Zonca M, O'Neill T, Silman A. Measurement of shoulder related disability: results of a validation study. *Ann Rheum Dis* 1994;**53**:525-8.
- Richards RR**, An K-N, Bigliani LU, Friedman RJ, Gartsman GM, Gristina AG, et al. A standardized method for the assessment of shoulder function. *J Shoulder Elbow Surg* 1994;**3**:347-52.
- Dawson J**, Fitzpatrick R, Carr A. Questionnaire on the perceptions of patients about shoulder surgery. *J Bone Joint Surg Br* 1996;**78**:593-600.
- Kohn D**, Geyer M. The subjective shoulder rating system. *Arch Orthop Trauma Surg* 1997;**116**:324-8.
- L'Insalata JC**, Warren RF, Cohen SB, Altchek DW, Peterson MG. A self-administered questionnaire for assessment of symptoms and function of the shoulder. *J Bone Joint Surg Am* 1997;**79**:738-48.
- van der Windt DAWM**, van der Heijden GJ, de Winter AF, Koes BW, Deville W, Bouter LM. The responsiveness of the Shoulder Disability Questionnaire. *Ann Rheum Dis* 1998;**57**:82-7.
- van der Heijden GJ**, Leffers P, Bouter LM. Shoulder disability questionnaire design and responsiveness of a functional status measure. *J Clin Epidemiol* 2000;**53**:29-38.
- de Winter AF**. *Diagnosis and classification of shoulder complaints*. Amsterdam: EMGO Institute Press, 1999 (PhD thesis).
- Pope DP**, Croft PR, Pritchard CM, Silman AJ. Prevalence of shoulder pain in the community: the influence of case definition. *Ann Rheum Dis* 1997;**56**:308-12.
- Williams JW Jr**, Holleman DR Jr, Simel DL. Measuring shoulder function with the Shoulder Pain and Disability Index. *J Rheumatol* 1995;**22**:727-32.
- Beaton DE**, Richards RR. Measuring function of the shoulder. A cross-sectional comparison of five questionnaires. *J Bone Joint Surg Am* 1996;**78**:882-90.
- Beaton D**, Richards RR. Assessing the reliability and responsiveness of 5 shoulder questionnaires. *J Shoulder Elbow Surg* 1998;**7**:565-72.
- Heald SL**, Riddle DL, 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.
- Roddey TS**, Olson SL, Cook KF, Gartsman GM, 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.
- Kind P**, Dolan P, Gudex C, Williams A. Variations in population health status: results from a United Kingdom national questionnaire survey. *BMJ* 1998;**316**:736-41.
- Husted JA**, Cook RJ, Farewell VT, Gladman DD. Methods for assessing responsiveness: a critical review and recommendations. *J Clin Epidemiol* 2000;**53**:459-68.
- Kazis LE**, Anderson JJ, Meenan RF. Effect sizes for interpreting changes in health status. *Med Care* 1989;**27**(suppl 3):S178-89.
- Liang MH**, Fossel AH, Larson MG. Comparisons of five health status instruments for orthopedic evaluation. *Med Care* 1990;**28**:632-42.
- Guyatt G**, Walter S, Norman G. Measuring change over time: assessing the usefulness of evaluative instruments. *J Chronic Dis* 1987;**40**:171-8.
- Hanley JA**, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* 1982;**143**:29-36.
- van der Schouw YT**, Verbeek AL, Ruijs JH. ROC curves for the initial assessment of new diagnostic tests. *Fam Pract* 1992;**9**:506-11.
- Jordan K**. *Statistical assessment of cervical spine and shoulder range of motion*. Keele: Department of Health Planning and Management and Department of Mathematics, 2000 (PhD thesis).