British Journal of Rheumatology 1997;**36:**463–469

USE OF THE SHORT FORM 36 (SF36) FOR HEALTH STATUS MEASUREMENT IN RHEUMATOID ARTHRITIS

J. TALAMO, A. FRATER,* S. GALLIVAN† and A. YOUNG

*Department of Rheumatology, City Hospital, St Albans, Herts. AL3 5PN, *Hertfordshire Health Agency, Charter House, Parkway, Welwyn Garden City, Herts. AL8 6JL and †Clinical Operational Research Unit, Department of Statistical Science, University College, London*

SUMMARY

The patient-administered Health Assessment Questionnaire (HAQ) is widely used in rheumatology studies. Another health quality assessment technique commonly used for other non-rheumatological conditions is the 'Short Form 36' (SF36). This has questions designed to assess eight aspects of health ranging from physical limitations to general perceptions of vitality and mental well-being. This study presents information on the health status of 137 patients with rheumatoid arthritis (RA) assessed by both the SF36 and HAQ. Summary statistics are given for the elements of the SF36 according to age, gender, disease measures of RA and the presence of co-morbidity. There were significant associations between the physical functioning score of the SF36 and the HAQ score, with other measures of disease activity and severity, and with co-morbidity, although there was considerable inter-patient variability. These findings suggest that future applications of health status questionnaires are possible.

KEY WORDS: Health Assessment Questionnaire, Outcome, Rheumatoid arthritis, SF36.

PATIENTS with rheumatoid arthritis (RA) often report many symptoms affecting quality of life. Yet there is no general agreement on the best measures to use in the management of this chronic disabling condition, and most are limited to joint involvement and acute-phase reaction such as the erythrocyte sedimentation rate (ESR). Many rheumatologists also undertake broader health assessments, including the use of the disability section of the Stanford Health Assessment Questionnaire (HAQ) [1]. This is used at present to assess physical disability and has been validated in several countries in Europe, including the UK [2]. It has been extensively used as an outcome measure in the USA and in both clinical trials and observational studies [3, 4]. A recent survey has shown that it is the most commonly used health status measure in rheumatology out-patients, although it is not known precisely to what extent the HAQ is used to influence management decisions [5]. There is growing evidence that it predicts outcome both in the long term [6, 7] and in the short term for patients with early disease [8]. Although a valuable screening instrument for problems in physical function in RA, it takes no account of emotional and psychological problems experienced by the patient, and which may affect patients' perceptions of their disease severity. In addition, the disability index of the HAQ has more emphasis on physical activities involving the upper limb than other aspects of physical functioning.

Using a health assessment instrument to audit the outcome of patient care in relation to quality of life might be especially useful for chronic conditions where clear outcomes are not expected and where success may best be measured in terms of maintaining an acceptable quality of life for the patient as the disease progresses. Ideally, health status measures should be valid (measure precisely what they set out to measure), reliable (produce repeatable results), acceptable to both patients and clinicians, providing the specific and general information needed in a form which is easily incorporated into a busy clinic setting. Sensitivity of the chosen measure to clinically important changes in health status over time is also important in rheumatology as therapeutic effects tend to be modest in the majority of patients [9, 10].

The Short Form 36 (SF36) [11], a general health status questionnaire designed in the USA, has been shown to be both valid and acceptable in a normal healthy population and reliable across diverse patient groups [12]. A version of the SF36 has been validated for use in the UK in a general population [13]. In common with the HAQ, it has the advantage of describing the impact of the disease in terms of patient-centred outcomes rather than the biological or disease-centred outcomes perceived by clinicians. It differs from the HAQ because it has questions designed to assess eight areas of health, ranging from those related to limitations in physical activities caused by disease to general perceptions of vitality and health.

This paper sets out to assess the value of the SF36 for the care of patients with RA. The specific objectives of our study have been to gather information about the SF36 in RA firstly to see how the physical functioning section of the SF36 compares with the HAQ and, secondly, to investigate how the assessments are influenced by factors such as age, gender and the presence of co-morbid conditions.

Submitted 20 May 1996; revised version accepted 19 September 1996.

Correspondence to: A. Young, Department of Rheumatology, St Albans City Hospital, Waverley Road, St Albans, Herts. AL3 5PN.

PATIENTS AND METHODS

Patients

Consecutive patients attending a rheumatology follow-up out-patient clinic with a clinical diagnosis of RA, as defined by the 1987 ACR criteria [14], were recruited to the study over a 4 month period. All patients approached agreed to take part. The sample included 137 patients, 40 males (29%) and 97 females (71%) with a mean age of 62 (26–91) yr and a mean duration of disease of 11 (0–43) yr. The demographic data of these patients (age, sex, the numbers and percentage of patients with radiological erosions, rheumatoid factor, any co-morbid, major and locomotor conditions) are summarized in Table I. Because of time constraints, only the first 65 of these patients attending for follow-up were reassessed 9–12 months later.

Sample and data collection

SF36 and HAQ questionnaires were given to consecutive patients with a diagnosis of RA during their routine out-patient examinations. Patients were asked to complete each questionnaire and the aims of the study were briefly explained to them. On completion of the questionnaires, patients were asked for their comments.

The disability section of the HAQ form asks a total of 20 questions in eight categories of function (dressing and grooming, rising, eating, walking, hygiene, reach, grip and activities) with two or three questions per category. For each question, the level of difficulty is scored from 0 to 3 (0 = patient is able to do without any difficulty, 1 = with some difficulty, 2 = with much difficulty or 3 = unable to do). The highest scores from each category are averaged to give a disability index, a continuous scale ranging from 0 to 3 with the highest scores representing the maximum impairment of function. Pain is recorded on a 10 cm visual analogue scale (VAS); patients are asked to mark pain over the last week. This is measured to the nearest millimetre and given a score of 0–100.

The SF36 assesses eight areas of general health as follows: limitations in physical activities caused by the disease; limitations in the social functioning of patients as a result of physical and/or emotional problems; limitations in the usual role functioning (work or other daily activities) as a result of emotional problems; limitations in the usual role functioning as a result of physical health problems; bodily pain; general mental health (feelings of well-being, depression etc.); vitality (energy and fatigue); general health perceptions. Scores in the range 0–100 are calculated for each of these different aspects of health, a low score indicating poor health status.

The clinical features of each subject were recorded as part of the information routinely gathered during follow-up examination of RA patients. These included: a joint score, recorded as the number of painful, swollen or tender joints; distribution of joint involvement including symmetry [15]; functional grade as assessed by Steinbrocker's method on a four-point scale [16]; hours of early morning joint stiffness; grip strength and ESR. Inactive RA was defined as less than half an hour of morning joint stiffness and no joint pain or swelling due to RA [15].

Details of co-existent medical conditions are also recorded routinely for all patients. Co-morbid conditions were grouped into three types, i.e. any co-existent condition, major co-existent disease (e.g.

TABLE I

Demography: co-morbidity and severity of rheumatoid arthritis in age and sex groups. Figures are numbers with percentages shown in parentheses

	Gender				Age group				Total	
	Male		Female		<55 yr		>55 yr			
	N	%	*N*	%	*N*	%	*N*	%	*N*	%
Number of co-morbid conditions										
None	15	(37)	37	(38)	15	(43)	37	(36)	52	(38)
One	9	(23)	23	(24)	13	(37)	19	(19)	32	(23)
Two or more	16	(40)	37	(38)	7	(20)	46	(45)	53	(39)
Locomotor co-morbidity										
None	26	(65)	59	(61)	24	(68)	61	(60)	85	(62)
One	8	(20)	21	(22)	9	(26)	20	(20)	29	(21)
Two or more	6	(15)	17	(18)	2	(6)	21	(20)	23	(17)
X-rays hands/feet										
Non-erosive	14	(35)	29	(30)	12	(34)	31	(30)	43	(31)
Erosive	26	(65)	68	(70)	23	(66)	71	(70)	94	(69)
Rheumatoid factor										
Positive	37	(93)	92	(95)	34	(97)	95	(93)	129	(94)
Negative	3	(7)	5	(5)	1	(3)	7	(70)	8	(6)
Disease activity										
Inactive RA	30	(75)	71	(73)	27	(77)	74	(73)	101	(74)
Active RA	10	(25)	26	(27)	8	(23)	28	(27)	8	(26)
Functional grade										
I	16	(40)	18	(18)	12	(34)	22	(22)	34	(25)
II	18	(45)	56	(58)	18	(52)	56	(55)	74	(54)
III	6	(15)	23	(24)	5	(14)	24	(23)	29	(21)

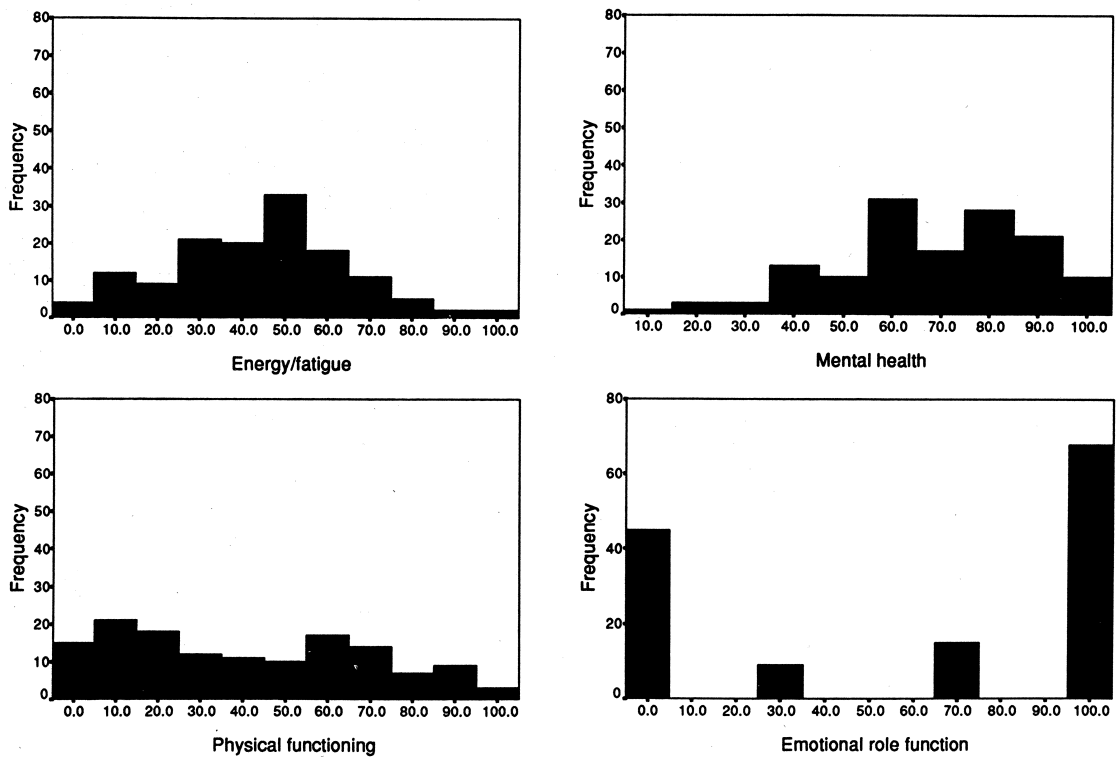

FIG. 1.—Distribution of scores for some elements of the SF36. Horizontal axis shows scores (range 0–100, with low scores indicating worse health status) in four of the possible eight dimensions covered by the SF36.

thyroid disease, diabetes, ischaemic heart disease) and conditions affecting limb function (e.g. cerebrovascular conditions, spinal disorders).

In order to investigate the responsiveness of the SF36 to change in functional status, 65 patients completed a second questionnaire ~9–12 months after the first when they returned for routine follow-up. The case notes of all these patients were reviewed by a clinician (AY) without knowledge of the patients' HAQ or SF36 responses. Patients were categorized according to whether or not major medical or life events had occurred between their first and second completion of HAQ and SF36 questionnaires. For example, hospitalization for joint surgery or co-morbid conditions, second-line drug changes or toxicity, etc.

TABLE II
Median values and interquartile ranges (shown in parentheses) for different elements of the SF36 questionnaire and for the HAQ for 137 rheumatoid arthritis patients

	Male $N = 40$	Female $N = 97$	Age $\leqslant 55$ $N = 35$	Age > 55 $N = 102$	Group $N = 137$
Physical functioning	30 (52.5)	35 (50.0)	40 (40.0)	35 (51.2)	35 (50.0)
Social functioning	77.8 (41.6)	66.7 (44.5)	66.7 (33.3)	66.7 (44.5)	66.7 (44.5)
Role function (physical)	25.0 (75.0)	0.0 (50.0)	0.0 (50.0)	0.0 (50.0)	0.0 (50.0)
Role function (emotional)	100.0 (100.0)	66.7 (100.0)	100.0 (100.0)	66.7 (100.0)	66.7 (100.0)
Mental health	78.0 (31.0)	68.0 (24.0)	68.0 (36.0)	68.0 (28.0)	68.0 (28.0)
Energy/fatigue	40.0 (35.0)	45.0 (27.5)	40.0 (20.0)	45.0 (27.5)	45.0 (25.0)
Pain	44.4 (55.6)	33.3 (33.3)	33.3 (44.5)	33.3 (33.3)	33.3 (33.3)
General health perception	52.0 (27.0)	47.0 (39.5)	52.0 (37.0)	47.0 (33.2)	47.0 (37.0)
HAQ score	1.19 (2.00)	1.50 (1.43)	0.87 (1.50)	1.50 (1.40)	1.37 (1.50)

Statistical analysis

Other studies report population norms for scores in each of the eight areas of health status assessed by the SF36, quoting summary statistics in terms of mean values and standard deviation. It was hoped to carry out similar analysis in order to compare the scores of RA patients with those of the general population [17]. However, preliminary descriptive analysis revealed that, for the group of RA patients, constituent scores were not normally distributed, indeed there were gross departures from normality for some of the constituent scores of the SF36 (see Fig. 1). In such circumstances, although mean and standard deviation values can be calculated, they have little relevance. Instead, we report summary statistics for our sample of RA patients in terms of median and interquartile ranges. Graphical presentations of data are given in terms of scatter plots or box and whisker plots displaying the median value, the interquartile range and the range of data values. Where appropriate, Spearman's rank method and Mann–Whitney tests were used to analyse the strength of relationships between variables.

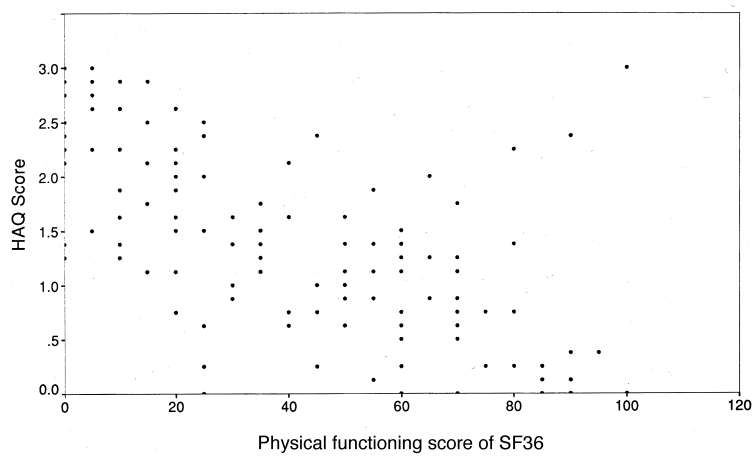

FIG. 2.—Scatter plot of HAQ (range 0–3) against physical functioning score of the SF36 (range 0–100). Worse function indicated by high scores for HAQ, low scores for SF36.

RESULTS

Both the HAQ and the SF36 were acceptable to patients. None of those asked refused to complete the forms. Three patients gave contradictory answers to questions concerning the physical functioning section of the SF36. Two patients stated for walking 100 yards that they were 'limited a lot'; however, for walking over a mile they stated that they were only 'limited a little'. Another patient claimed to be 'limited a lot' climbing one flight of stairs, but only 'a little' in climbing several flights of stairs.

Table II summarizes the median and interquartile ranges for each of the aspects of health status covered by the SF36 and HAQ according to gender and age band. Median and interquartile ranges for the different elements of the SF36 according to whether or not patients had active RA at the same time as completing the questionnaire, whether or not they had a co-morbid condition and according to their functional grade are shown in Table III.

Figure 2 shows a scatter plot of patients' HAQ against the physical functioning score of the SF36. There is clearly a strong association between the two, confirmed by Spearman's rank correlation analysis ($rs = 0.72$, $P < 0.001$); however, there is also considerable inter-patient variability. The three outlying points in the upper right of the scatter diagram may well correspond to patients who gave inconsistent replies to the SF36 questionnaire.

The HAQ score and the physical functioning score of the SF36 are both associated with the number of co-morbid conditions, as shown in Figs 3 and 4. There

TABLE III
Median values and interquartile ranges (shown in parentheses) for different elements of the SF36 questionnaire and for the HAQ for 137 rheumatoid arthritis patients

	Active disease $N = 36$	Inactive disease $N = 101$	Co-morbidity present $N = 85$	No co-morbidity $N = 52$	Functional grade I $N = 34$	Functional grade II/III $N = 103$
Physical functioning	20.0 (35.0)	45.0 (55.0)	20.0 (50.0)	55.0 (43.8)	70.0 (35.0)	20.0 (45.0)
Social functioning	55.6 (52.8)	66.7 (33.3)	66.7 (50.0)	77.8 (33.3)	88.9 (33.3)	66.7 (33.3)
Role function (physical)	0.0 (25.0)	0.0 (75.0)	0.0 (25.0)	25.0 (75.0)	62.5 (100.0)	0.0 (25.0)
Role function (emotional)	66.7 (100.0)	100.0 (100.0)	66.7 (100.0)	100.0 (66.7)	100.0 (66.7)	66.7 (100.0)
Mental health	60.0 (34.0)	72.0 (26.0)	68.0 (28.0)	74.0 (28.0)	76.0 (29.0)	64.0 (28.0)
Energy/fatigue	30.0 (35.0)	45.0 (25.0)	40.0 (25.0)	47.5 (30.0)	50.0 (35.0)	40.0 (25.0)
Pain	22.2 (82.0)	44.5 (44.5)	22.2 (33.3)	44.5 (44.5)	66.7 (36.1)	27.8 (25.0)
General health perception	31.0 (37.0)	52.0 (34.5)	42.0 (31.0)	57.0 (32.0)	58.5 (26.25)	42.0 (32.0)
HAQ score	1.62 (1.22)	1.25 (1.50)	1.87 (1.37)	0.87 (1.12)	0.37 (0.75)	1.75 (1.25)

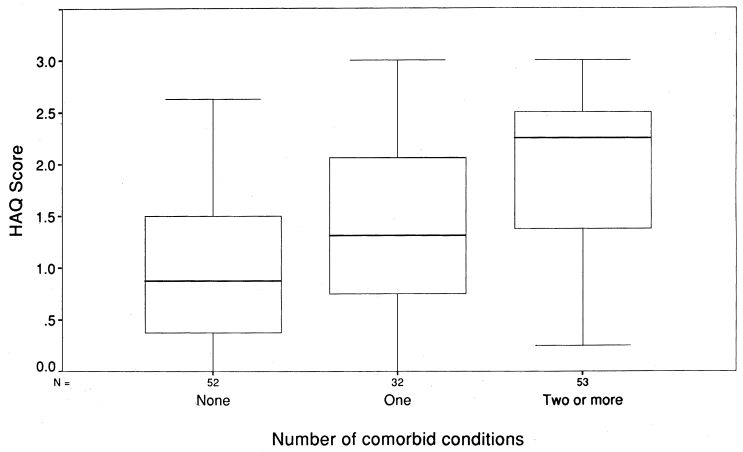

FIG. 3.—HAQ score and co-morbidity. Box and whisker plot to show median, interquartile range, and full range of HAQ in patients with no co-morbidity, with one, or with two or more co-existent conditions.

is a statistically significant difference between the scores recorded for patients having no co-morbidity and those having at least one co-morbid condition ($P < 0.001$; Mann–Whitney test).

The change in patients' scores for the HAQ and the physical functioning element of the SF36, from their first assessment to a subsequent assessment 9–12 months later, is illustrated in Figs 5 and 6. In these scatter plots, we have distinguished between patients whose condition was judged to have deteriorated and those who have remained the same or improved. For descriptive purposes, linear regression lines of 'best fit' have been included in these figures for the two groups.

For patients whose condition did not deteriorate, both health status scores give a scatter roughly equally divided between the upper and lower quadrants. As might be expected, although there is not a perfect match between initial and subsequent scores, roughly as many increase as decrease and the mean change in score is close to zero.

For the group of patients whose condition deteriorated, however, the scatter of points indicates a corresponding change in health status score. For the SF36 physical functioning scores, most data points lie in the lower quadrant and, for the HAQ, most lie in the upper quadrant. The SF36 showed a change in the opposite direction to the HAQ score. Further work is needed to assess fully the value of the instrument in monitoring change over time.

DISCUSSION

RA is a debilitating condition having its greatest impact on the pain and mobility of patients. This study shows that both the SF36 and HAQ are potentially

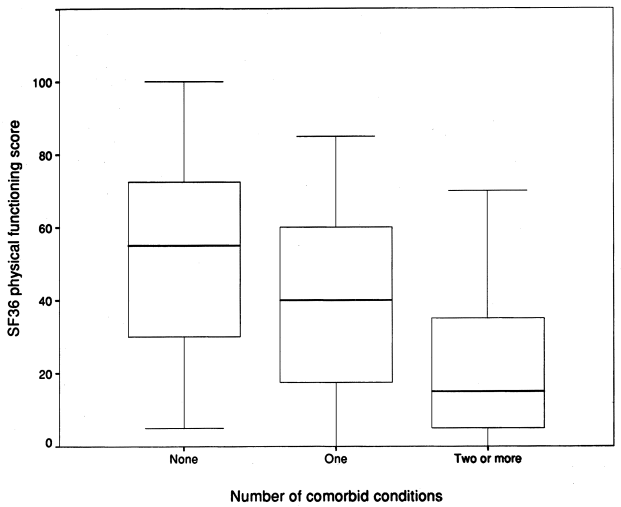

FIG. 4.—SF36 score and co-morbidity. Box and whisker plot to show median, interquartile range, and full range of HAQ in patients with no, with one and with two or more co-existent conditions.

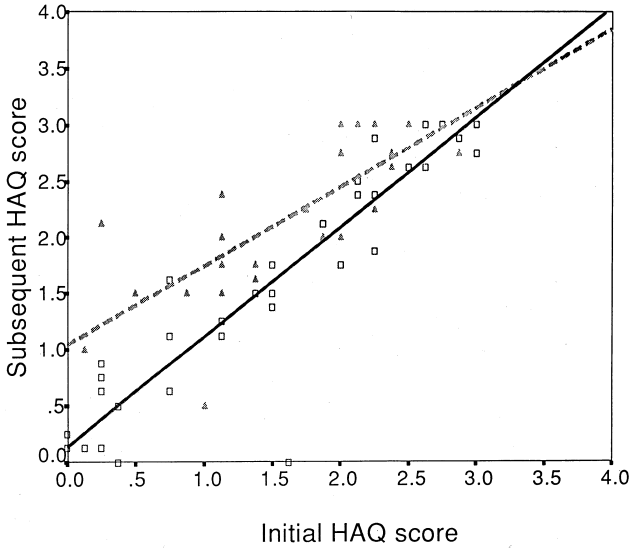

FIG. 5.—Change in HAQ. Scatter plot to illustrate change in HAQ over a 9–12-month period in patients whose overall condition deteriorated and those who remained the same. A linear regression line of 'best fit' has been superimposed. – – – ▲, worse; —— □, no worse. See text.

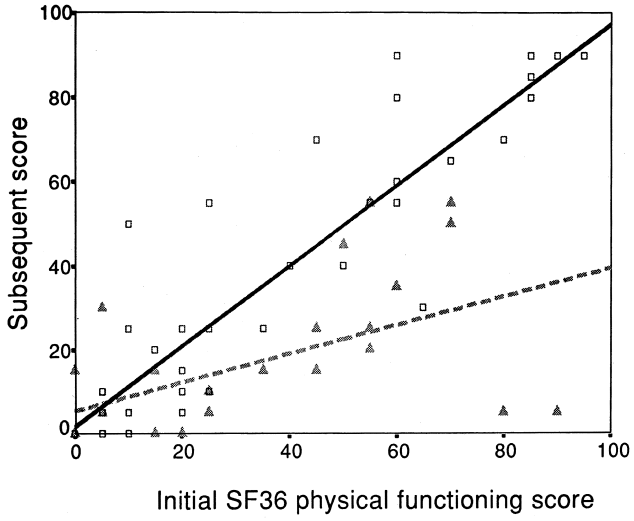

FIG. 6.—Change in SF36 physical functioning score. Scatter plot to illustrate change in SF36 over a 9–12 month period in patients whose overall condition deteriorated and those who remained the same. A linear regression line of 'best fit' has been superimposed. – – – ▲, worse; —— □, no worse. See text.

useful measures to help patient and clinician better understand the influence of disease on wider aspects of quality of life. Some patients with RA clearly suffer a decline in all aspects of health, both physical and emotional.

It could be argued that the identification and treatment of non-physical aspects of chronic disease should be considered as part of therapeutic management for this condition, e.g. anxiety and depression, and problems concerning social and role functioning. Routine use of the SF36 might be useful to help identify patients for whom intervention is needed. The value of the information for influencing clinical management or more closely involving patients in their care was not fully developed by this study.

As might be expected, the results showed that poor health status, as assessed by the SF36 and the HAQ, was consistent with current disease activity, poor functional grade and the presence of co-morbidity. Patients with co-morbid conditions had worse scores for both life quality measures, similar to other reports [18, 19], but not all [20]. Also, worse scores (lower for SF36, higher for HAQ) were seen for female patients aged 55 or above. These findings suggest that future applications of health status questionnaires are possible.

Collecting information on health status using HAQ and SF36 was acceptable to patients, though unfamiliar to them. Informally, patients reported a preference for the HAQ because they found it easier to complete. However, the HAQ is limited to providing information on functional health and pain, whereas the SF36 provides a broader picture of health, including emotional status, role functioning, and anxiety and depression. These dimensions have not been fully explored in this study and warrant further attention. The introduction of these measures for assisting with diagnosis, assessing prognosis and auditing outcome of health care will rely on their acceptability to patients and their potential for monitoring the progress of the disease in relation to health status. Further work is needed to explore this question, including a comparison of other possible measures and with particular emphasis on sensitivity to change.

More work is needed to clarify how different aspects of health quality for disease-specific groups compare with the healthy population. Although this was one of the original aims of our study, the non-normal distributions of scores from the SF36 made formal comparison with other work impossible. Technically, so-called parametric statistical methods for comparing a sample mean with the mean of a reference population can be carried out using the reported mean and standard deviation for the reference population. However, such statistical testing relies on the underlying distribution being normal. If this is not the case, non-parametric statistical methods are required. Unfortunately, these cannot be carried out if the only information available is the reference population mean and standard deviation. This is discussed at length by Treasure *et al*. [21]. In the extreme case where health status measures are dichotomous, which is the case for some dimensions of the SF36, it is unclear what information the mean and standard deviation are meant to convey. They certainly do not convey any useful summary of the 'central tendency' of the distribution concerned.

The value of subjective information on the health status of patient populations would be enhanced by the possibility of making comparisons with norms for attributes such as physical and emotional functioning in the general population. Researchers must ensure that norms are published using non-parametric summary statistics for non-normally distributed data or provide basic information from which they could be derived.

REFERENCES

1. Fries JF, Spitz P, Kraines RG, Holman HR. Measurement of patient outcome in arthritis. Arthritis Rheum 1980;23:137–45.
2. Kirwan JR, Reeback JS. Stanford Health Assessment Questionnaire modified to assess disability in British patients with rheumatoid arthritis. Br J Rheumatol 1986;25:206–9.
3. Leigh JP, Fries JF, Parikh N. Severity of disability and duration of disease in rheumatoid arthritis. J Rheumatol 1992;19:1906–11.
4. Buchbinder R, Bombardier C, Yeung M, Tugwell P. Which outcome measure should be used in rheumatoid arthritis clinical trials? Clinical and quality-of-life measures' responsiveness to treatment in a randomised controlled trial. Arthritis Rheum 1995;38:1568–80.
5. Carr A, Thompson P, Young A. Do health status measures (HSM) have a role in rheumatology? A survey of the use of and attitudes towards health status measures in the UK. Arthritis Rheum 1996; ACR Abstracts S261.
6. Wolfe F, Cathey MA. The assessment and prediction of functional disability in rheumatoid arthritis. J Rheumatol 1991;18:1298–306.

7. Leigh JP, Fries JF. Predictors of disability in a longitudinal sample of patients with rheumatoid arthritis. Ann Rheum Dis 1992;51:581–7.
8. Young A. Short-term outcomes in recent-onset rheumatoid arthritis. Br J Rheumatol 1995;34(suppl. 2):79–86.
9. Fitzpatrick R, Ziebland S, Jenkinson C, Mowat A, Mowat A. Importance of sensitivity to change as a criterion for selecting health status measures. Quality Health Care 1992;1:89–93.
10. Liang MH, Jette AM. Measuring functional ability in chronic rheumatoid arthritis. A critical review. Arthritis Rheum 1981;24:80–6.
11. Ware JE, Sherbourne CD. The MOS 36-item short-form health status survey (SF-36). 1. Conceptual framework and item selection. Med Care 1992;30:473–83.
12. McHorney CA, Ware JE, Lu JFR. The MOS 36-item short-form health survey (SF36). III. Tests of data quality, scaling assumptions and reliability across diverse patient groups. Med Care 1994;32:40–66.
13. Brazier JE, Harper R, Jones NMB *et al*. Validating the SF-36 health survey questionnaire: new outcome measure for primary care. Br Med J 1992;305:160–4.
14. Arnett FC, Edworthy SM, Bloch DA *et al*. The ACR 1987 revised criteria for classification of rheumatoid arthritis. Arthritis Rheum 1988;31:315–24.
15. Young A, Corbett M, Brook A. The clinical assessment of joint inflammatory activity in rheumatoid arthritis related to radiological progression. Rheumatol Rehabil 1980;19:14–9.
16. Steinbrocker O, Traeger CH, Batterman RC. Therapeutic criteria in rheumatoid arthritis. J Am Med Assoc 1949;140:659–62.
17. Jenkinson C, Coulter A, Wright L. Short form 36 (SF36) health survey questionnaire: normative data for adults of working age. Br Med J 1993;306: 1437–40.
18. Young A, Talamo J, Frater A, Gallivan S. Comorbidity and health perception in rheumatoid arthritis. Br J Rheumatol 1994;33(suppl. 2):22.
19. Rasker JJ. Duration of rheumatoid arthritis has greater influence than comorbidity on functional outcome in RA. Br J Rheumatol 1994;33(suppl. 2): 22.
20. Deighton CM, Surtees D, Walker DJ. Influence of the severity of rheumatoid arthritis on sex differences in Health Assessment Questionnaire scores. Ann Rheum Dis 1992;51:473–5.
21. Treasure T, Gallivan S. Quality of life data: how can we get the best quality from them? J R Soc Med 1986;88:25–6.