Health-Related Quality of Life: Validity, Reliability, and Responsiveness of SF-36, EQ-15D, EQ-5D, RAQoL, and HAQ in Patients with Rheumatoid Arthritis

LOUISE LINDE, JAN SØRENSEN, MIKKEL ØSTERGAARD, KIM HØRSLEV-PETERSEN, and MERETE LUND HETLAND

ABSTRACT. ***Objective.*** To compare validity, reliability, and responsiveness of generic and disease specific health-related quality of life (HRQOL) instruments in rheumatoid arthritis (RA).

Methods. Two samples of patients completed the Medical Outcomes Study Short Form-36 Health Survey (SF-36), EuroQol (EQ)-5D, 15D, Rheumatoid Arthritis Quality of Life Scale (RAQoL), Health Assessment Questionnaire (HAQ), and visual analog scales (VAS) for pain, fatigue, and global RA. Validity (convergent, discriminant, and known-groups) was evaluated in a cross-section of 200 patients. Reliability was evaluated by agreement (intraclass correlation coefficient; baseline to 2 weeks) and internal consistency (Cronbach's alpha); and responsiveness by the standardized response mean stratified on improvement, status quo, or deterioration in health status after 6 months in 150 patients followed longitudinally. Followup questionnaires (at 2 weeks and 6 months) included questions about changes in health status since baseline.

Results. The cross-sectional sample included 77% women, median age 57 years (range 19–87), disease duration 6 years (0–58), with Disease Activity Score 28-joint count (DAS28) of 3.10 (1.21–6.47). The longitudinal sample included 80% women, median age 60 years (22–82). Validity: all instruments discriminated between low, moderate, and high DAS28. Reliability: RAQoL and HAQ displayed good repeatability (ICC > 0.95) and internal consistency (Cronbach's alpha > 0.90). Responsiveness: SF-36 bodily pain scale and VAS pain were responsive to both improvement and deterioration.

Conclusion. All instruments were valid measures for HRQOL in RA. The RAQoL and HAQ displayed the best reliability, while the SF-36 bodily pain scale and VAS pain were the most responsive. The choice of instrument should depend on the study objectives. (First Release May 15 2008; J Rheumatol 2008;35:1528–37)

Key Indexing Terms:
RHEUMATOID ARTHRITIS HEALTH-RELATED QUALITY OF LIFE
VALIDITY RELIABILITY RESPONSIVENESS

Rheumatoid arthritis (RA) is a chronic disabling disease affecting physical, mental, and social aspects of patients' lives. Traditional clinical disease markers of RA such as joint counts and serum C-reactive protein (CRP) quantify some of the physical aspects of the disease, but fail to fully identify the broad spectrum of disease effects. Consequently, patient-reported outcome measures of health-related quality of life (HRQOL) have been developed to complement the traditional disease markers, and the importance of such measures in determining improvement in clinical trials has been recognized by the American College of Rheumatology (ACR) and Outcome Measures in Rheumatoid Arthritis Clinical Trials (OMERACT)[1]. Some studies have even suggested patient-reported outcome measures to be potentially more sensitive to a clinical change in RA than traditional disease markers[2,3]. HRQOL measurement instruments can be either generic (i.e., general or disease-independent), such as the Medical Outcomes Study Short Form-36 Health Survey (SF-36), or disease-specific (designed for a specific disease), such as the Rheumatoid Arthritis Quality of Life Scale (RAQoL)[4]. HRQOL measures are now increasingly used in clinical trials of RA, but in order to draw conclusions from the results, the properties of the measurement instruments have to be evaluated in terms of validity, reliability, and responsiveness.

The psychometric properties of SF-36 have been demonstrated in various populations, and it is widely used as a

From the Department of Rheumatology, Copenhagen University Hospital, Hvidovre; CAST, University of Southern Denmark, Odense; and Gråsten Gigthospital, University of Southern Denmark, Gråsten, Denmark.

L. Linde, MD; M. Østergaard, MD, PhD, DMSc; M.L. Hetland, MD, PhD, Department of Rheumatology at Copenhagen University Hospital; J. Sørensen, MSc, CAST, University of Southern Denmark; K. Hørslev-Petersen, MD, DMSc, Gråsten Gigthospital.

Address reprint requests to Dr. L. Linde, Department of Rheumatology, 232, Kettegaard Allé 30, DK-2650 Hvidovre, Denmark.
E-mail: louiselinde@dadlnet.dk

Accepted for publication February 7, 2008.

measure of health status in clinical trials of RA. The EuroQol (EQ-5D) and 15D are preference-based generic tools that are well suited for cost-utility analyses. The EQ-5D has been validated in RA[5,6] and is often used in economic drug evaluation studies, while the 15D is less used, despite its broader coverage of health aspects.

The Stanford Health Assessment Questionnaire (HAQ) was developed in 1980 as a disease-specific measure of functional disability in RA[7]. The HAQ is extensively used in both clinical trials and clinical practice in RA. The RAQoL is a more recently developed RA quality of life instrument, well suited for use in clinical trials of RA[8-10].

Some studies have suggested that disease-specific instruments are more sensitive to treatment-induced changes in RA patients than generic instruments[11-13], and others moreover report greater sensitivity to self-reported changes in RA[14,15]. Comparative work across the categories is sparse, and there is thus a need for further studies to elucidate which of the available generic and disease-specific HRQOL instruments are best suited for assessment and monitoring in RA.

The purpose of our study was to investigate and compare the validity, reliability, and responsiveness of 3 well established (SF-36, EQ-5D, HAQ) and 2 less employed (15D, RAQoL) generic and disease-specific HRQOL measurement instruments in routine care of patients with RA.

MATERIALS AND METHODS

Patients. We collected data from 2 samples of patients with RA. The patients were recruited from the outpatient clinics at Gråsten and Hvidovre Hospitals in Denmark; the inclusion criterion was a diagnosis of RA according to the ACR 1987 criteria.

Sample one (P1) was a cross-sectional review of 200 patients with RA, which was part of a larger study (QUEST-RA)[16]. Patients with dementia were excluded, as their contribution to our study is complicated by a diminished recall capacity. Data were collected between January and April 2005 and comprised a clinical evaluation performed by a physician and a questionnaire completed by the patient. The clinical evaluations were done by experienced rheumatologists and included: ACR criteria, 42-joint count, visual analog scale (VAS) score for disease activity, serum-CRP (s-CRP), radiographic joint examination of hands and feet, extraarticular disease, medical history, comorbidities, and history of joint surgery. The patient questionnaire included items regarding work-status, physical function (HAQ), VAS scores (fatigue, pain, global RA, arthritis activity), lifestyle, health status, and quality of life (SF-36, EQ-5D, 15D, RAQoL). A health professional aided those in need of assistance with the questionnaire.

The other sample (P2) included 150 patients with RA followed longitudinally. The patients were enrolled by their physician or nurse at the Hvidovre outpatient clinic during a 2-week period in December 2005. Patients with dementia, blindness, or deafness were excluded, as were patients with a language barrier. Data were collected by means of a patient questionnaire administered at baseline, 2 weeks, and 6 months and included age and sex, physical function (HAQ), VAS scores (fatigue, pain, global RA), health status, and quality of life (SF-36, EQ-5D, 15D, RAQoL). Moreover, the 2-week and 6-month questionnaires included a question on self-reported change (better, no change, worse) in RA and general health since baseline. The baseline questionnaire was administered in the clinic, while the 2-week and 6-month questionnaires were sent by mail to those who returned the baseline questionnaire. The patients were instructed to complete all the questionnaires at home and to return them in a prepaid envelope. To increase the response rate the nonresponders were contacted by telephone and encouraged to return the questionnaire.

Questionnaires. The SF-36 is a generic measure of health status covering both physical and mental aspects of health. The 36 multiple-choice questions produce 8 dimensions of health: physical functioning (PF), physical role limitations (RP), bodily pain (BP), general health perceptions (GH), vitality (VT), social functioning (SF), emotional role limitations (RE), and mental health (MH). The scores range from 0 (poor health) to 100 (perfect health). We used the Danish standard (4-week recall) version 1[17].

The EQ-5D is a generic preference-based health status instrument including 5 dimensions of health (mobility, self-care, usual activities, pain/discomfort, anxiety/depression), each divided into 3 levels of severity. Patients are asked to give their health state as of today. The 243 possible health states have been weighted, yielding an index score between 0 (death) and 1 (perfect health)[18,19].

The 15D instrument is a generic preference-based health status measure including 15 dimensions of health [mobility, vision, hearing, breathing, sleeping, eating, speech (communication), elimination, usual activities, mental function, discomfort and symptoms, depression, distress, vitality, and sexual activity], each divided into 5 levels of severity[20]. Patients are asked to give their health state as of today. Due to the large number of possible health states, the weighting is modeled by an algorithm providing an index score between 0 (death) and 1 (perfect health)[19].

The RAQoL is a disease-specific measure of quality of life in RA[10]. The patients are asked to answer 30 dichotomous questions (answered by yes/no) regarding physical, emotional, and social limitations caused by the disease at the moment. For each affirmative answer 1 point is assigned; the scores thus range from 0 (best score) to 30 (worst score).

The HAQ is a measure of functional disability in RA. It includes 20 questions (8 dimensions) on the ability to perform activities of daily living (ADL) at the moment. The response options range from 0 (with no difficulty) to 3 (unable to do). The highest scores of each dimension are summed and divided by 8, resulting in a possible range of total scores (HAQ-score) from 0 (no difficulty) to 3 (unable to do)[21].

Questionnaire analysis. The basic structure of the instruments, including the relevance of their content, has been thoroughly documented[5,10,20,22-25]. We applied the Danish translations of SF-36[17], EQ-5D (C. Gudex, unpublished data), 15D (D. Gyrd-Hansen, unpublished data), RAQoL, and HAQ[8], and computed the summary scores (SF-36, VAS, RAQoL, HAQ) and utility indexes (EQ-5D, 15D) according to the guidelines proposed by the developers. The HAQ was scored without including aids or help from other people. The score distributions were examined graphically.

Missing data: the proportion of missing items (i.e., unanswered questions) was assessed for each item individually in SF-36, EQ-5D, 15D, VAS, RAQoL, and HAQ. Missing data were imputed at the item level, and values were therefore replaced using median imputation.

Validity: a valid instrument measures what it purports to measure, hence it is without systematic measurement error. Construct validity is an important element, and 3 aspects are often considered: (1) known-groups, (2) convergent, and (3) discriminant validity. Known-groups validity is based on the assumption that different groups of patients are expected to yield different HRQOL scores and is tested by examining the sensitivity of the instruments to these differences. Convergent and discriminant validity are tested by exploring hypotheses on the strength of association between instruments or scales measuring related and unrelated concepts[26]. In our study, construct validity was illustrated by employing tests for known-groups, convergent, and discriminant validity in P1. Known-groups validity was investigated by examining the sensitivity of the instruments to differences between predefined groups of differing disease severity. The measurement instruments were hypothesized to show poorer scores in the groups thought to have more severe RA. Four variables were chosen to illustrate disease severity: (1) Disease Activity Score based on CRP and 28-joint count (DAS28)[27], (2) VAS for self-reported current arthritis activity, (3) presence of bone erosions on conventional radiographs of hands and feet,

and (4) disability pension status. DAS28 and VAS were grouped into low, moderate, and high disease activity, while presence of bone erosions and disability pension status were dichotomous variables. The sensitivity to the known difference in the groups was evaluated by effect sizes (ES) for the dichotomous variables and score comparisons for the disease activity variables. Guided by the score distributions, we used either nonparametric methods (Kruskal-Wallis for group comparisons and Mann-Whitney for 2-sample comparisons) or 1-way analysis of variance (ANOVA) with adjustment for multiple comparisons by Tukey's method. Known-groups validity was evaluated mainly in terms of the magnitude of the mean or median differences (and to a lesser extent the statistical significance) and the ES values. Cohen's graduation of ES into small < 0.5, medium 0.5–0.8, and large (> 0.8) was used[28]. Convergent and discriminant validity was investigated by examining the association between instruments or scales hypothesized to be related (convergent validity) and unrelated (discriminant validity). We used the multitrait-multimethod (MTMM) correlation matrix, in which associations of different methods to assess a specific trait are studied. In this case, we studied associations of different HRQOL instruments or scales to assess certain aspects of health. We created 6 health aspect groups (physical function, pain, fatigue, global RA, overall health, and mental/social function) and all instruments and scales were subsequently assigned to their appropriate group. Instruments or scales grouped together were hypothesized to be more associated with each other (convergent validity) than with those in the other groups (discriminant validity). As the distributions of some of the data were assumed to be non-normal, we used Spearman's correlations. Correlations above 0.70 (strong) and below 0.30 (weak) indicate good convergent and discriminant validity, respectively[29].

Reliability: a reliable test is precise, hence it has no or minimal random measurement error.

The 2 most common ways of evaluating reliability are: (1) repeatability, which is the ability of a test to produce similar results in repeated measurements; and (2) internal consistency, which estimates the interrelatedness of items in multi-item scales. Thus repeatability can be tested for any instrument or scale, while internal consistency can be investigated only for multi-item scales[26].

Repeatability was investigated in P2 by test-retest of SF-36, EQ-5D, 15D, RAQoL, HAQ, and VAS (fatigue, pain, global RA) with a 2-week interval. The strength of agreement between the 2 measurements was estimated for the patients who reported no change from baseline in RA and overall health status. The instrument scores were considered to be continuous, and agreement was therefore estimated using the intraclass correlation coefficient, 2-way mixed (ICC). To estimate the random error of each instrument, i.e., the variation in repeated measurements on the same subject, we plotted the differences between the first and second measurement against their means. To test the null hypothesis (mean difference = 0) we used the 1-sample t-test. The mean difference ± 1.96 * SD is labeled the coefficient of repeatability (CR) and the resulting interval, "the 95% limits of agreement." Both can be interpreted as estimates of the level of random error[30]. Internal consistency was investigated in P2 for SF-36, RAQoL, and HAQ (baseline scores) as these instruments consist of one (RAQoL and HAQ) or more (SF-36) multi-item scales. The analyses were done using Cronbach's alpha coefficients[31]. ICC and Cronbach's alpha values above 0.90 are considered suitable for individual comparisons, while values above 0.70 are acceptable for group comparisons[32].

Responsiveness: a responsive measurement instrument is able to detect clinically important changes over time, even if those changes are small[33]. In our study, the patient-reported changes in RA and overall health over time were considered to be clinically important and were thus used as external indicators for change. The patient-reported changes in RA were thought to mainly reflect disease activity; we thus primarily expected these changes to be illustrated in instruments covering aspects of fatigue, pain, and physical function. P2 was divided into 3 subgroups (better, no change, worse) according to the patient's self-reported change in RA and general health at 6 months from baseline. The score changes (baseline-6 months) were calculated for the subgroups, and normality of the score changes was confirmed. We used ANOVA statistics to test for differences in the mean score changes in the subgroups. If a significant p value was reached, we explored the difference(s) further, adjusting for multiple comparisons by Tukey's method. We expected all 3 subgroups to differ significantly in mean score changes. To examine the responsiveness further, we applied a distribution-based approach, the standardized response mean (SRM), estimated as the ratio of the mean score changes to the SD of that change. SRM was calculated in the 3 subgroups for all instruments and scales, and the values were categorized as small (< 0.5), medium (0.5–0.8), and large (> 0.8).

Statistical analysis. The data were entered into an Access database, and SPSS was used for the statistical analyses. The choice of statistics was based on the distribution of data, and a p value ≤ 0.05 was chosen as the level of statistical significance.

Ethics. The P1 data were gathered according to a protocol approved by the national health authorities and ethics committees in both participating counties, and the study was carried out in accordance with the Declaration of Helsinki. The P2 participants were all informed orally and in writing about the objectives of our study and ensured that their decision to complete the questionnaires was voluntary and would not affect any future treatment decisions.

RESULTS

Patients. P1: Hvidovre and Gråsten outpatient clinics each included 100 patients with RA, thus in total 200 patients were eligible for the study. Seventy-seven percent were women, median age 59 years (range 19–87), median disease duration 6 years (range 0–58), median DAS28 3.10 (range 1.21–6.47), 72% were positive for IgM rheumatoid factor, and 60% had bone erosions on conventional radiographs of hands and feet.

P2: Hvidovre outpatient clinic considered 167 patients with RA for inclusion. Twelve declined to participate, while 5 were excluded. Thus 150 patients were eligible for the study. One hundred forty-four (96%), 135 (94%), and 123 (85%) patients returned the baseline, 2-week, and 6-month questionnaires, respectively. The baseline characteristics were 80% women, median age 60 years (range 22–82). No differences in sex, age, or baseline HRQOL scores were found between dropouts and participants.

One hundred thirty-three patients returned baseline and 2-week questionnaires; 87 (65%) of these reported no changes from baseline in RA and overall health status. One hundred eighteen patients returned both the baseline and the 6-month followup questionnaire; 47 (40%), 23 (19%), and 26 (22%) reported no change, deterioration, or improvement since baseline in RA and overall health status, respectively. The median (range) number of days between responses was 17 (15–21) (baseline to 2 weeks), and 183 (162–281) (baseline to 6 months).

Questionnaire analysis. In P1, the scores for RAQoL, HAQ, EQ-5D, 15D, all 3 VAS, and SF-36 mental health and social functioning scales were skewed toward the better end of the scale, which can be observed as differences in mean and median score values in Table 1. The SF-36 physical and emotional role limitation scores were non-normal in both P1 and P2, while the rest of the instrument scores in P2 did not display major deviations from the normal distribution.

Table 1. Baseline HRQOL scores of P1 and P2 after imputation of the missing data.

Instruments (best/worst score)	Group P1 (n = 200) Baseline Scores, mean (SD)/median (IQR)	Group P1 (n = 200) % of Patients with Best/Worst Possible Scores	Group P2 (n = 150) Baseline Scores, mean (SD)/median (IQR)	Group P2 (n = 150) % of Patients with Best/Worst Possible Scores
RAQoL (0/30)	9 (7)/8 (3–14)	9.1/0	11 (7)/11 (5–15)	3.5/0
HAQ (0/3)	0.79 (0.77)/0.63 (0–1.38)	25.3/0	0.94 (0.68)/0.88 (0.25–1.38)	9.8/0.7
EQ-5D (1/0)	0.73 (0.19)/ 0.76 (0.66–0.82)	16.8/0.5	0.67 (0.18)/0.71 (0.63–0.78)	5.6/0.7
15D (1/0)	0.88 (0.09)/0.89 (0.83–0.95)	11.0/0	0.84 (0.09)/0.85 (0.79–0.90)	2.8/0
SF-36 (100/0)				
Physical functioning	60 (27)/65 (40–85)	5.2/2.1	53 (24)/55 (35–74)	0.7/2.1
Role physical	47 (41)/50 (0–100)	27.6/33.5	29 (38)/0 (0–50)	16.4/52.1
Bodily pain	56 (25)/62 (32–74)	8.1/2.5	58 (21)/52 (41–72)	5.6/0
General health	55 (23)/55 (40–72)	1.4/1.4	61 (13)/60 (50–70)	0/0
Vitality	58 (25)/60 (40–80)	4.1/1.0	48 (24)/45 (30–69)	0.7/1.4
Social functioning	84 (24)/100 (75–100)	55.3/0.5	49 (11)/50 (50–50)	0.7/0
Role emotional	67 (40)/100 (33–100)	51.6/19.8	50 (45)/50 (0–100)	38.6/40.7
Mental health	79 (19)/84 (72–92)	11.3/0	67 (9)/66 (63–71)	0/0
VAS pain (0/100)	32 (25)/26 (10–50)	3.9/0	39 (26)/37 (15–55)	1.4/0
VAS fatigue (0/100)	39 (30)/33 (10–65)	4.5/0	45 (30)/48 (17–70)	2.2/0.7
VAS global RA (0/100)	33 (26)/27 (9–52)	5.6/0	40 (25)/40 (19–58)	0/0

SD: standard deviation; IQR: interquartile range; HRQOL: health-related quality of life; RA: rheumatoid arthritis; HAQ: Health Assessment Questionnaire; EQ: EuroQol; SF-36: Medical Outcomes Study Short Form-36 Health Survey; VAS: visual analog scale.

However, the 15D scores and the HAQ scores tended to cluster around the better end. For all instruments, the P1 scores were better than the P2 scores, indicating a better health status in the former sample (Table 1).

Missing data: some items were returned with more missing values than would be expected by chance (0.5%–2%) according to Fayers and Machin[26]. Some of these included the 15D sexual activities dimension (P1: 13%, P2: 14%), the HAQ question regarding bathing in a bathtub (P1: 14.7%, P2: 19%), the VAS global (P1: 11%, P2: 5%), VAS fatigue (P1: 12%, P2: 4%), VAS pain (P1: 10%, P2: 6%), and for P1 the SF-36 questions regarding limitations in work or daily activities as a result of either emotional (8%–10%) or physical (6%–8%) problems (data not shown). The baseline scores of P1 and P2 after imputation of missing items are shown in Table 1.

Validity: known-groups validity. While all differences between low and moderate DAS28 were visualized by large and significant score gradients, not all instruments were able to significantly discriminate between moderate and high DAS28. This effect was less marked for the score gradients over the VAS for self-reported current arthritis activity groups (Figure 1A-1D and Table 2). The HAQ, RAQoL, and SF-36 physical functioning and bodily pain scales discriminated well between patients receiving disability pension versus those who did not, as illustrated by medium effect size values (0.5–0.8). No score differences in the presence/absence of bone erosions groups reached statistical significance (Table 2).

Convergent and discriminant validity. In 5 of the 6 defined health aspect groups (physical function, pain, fatigue, global RA, overall health), we found the hypothesized associations reflected in correlation coefficients above 0.70, indicating that the instruments in question may be measuring the same construct (Table 3; results from the mental/social function group are not shown). The RAQoL was moreover strongly correlated with instruments in the physical function (–0.65 to 0.81), pain (–0.72 to 0.75), fatigue (0.75 to 0.78), and overall health (–0.69 to –0.83) groups, indicating that the scale identifies most aspects of RA. This effect was also present for the HAQ, although to a lesser degree, as the correlations with instruments in the fatigue and mental/social function groups were weaker. The EQ-5D and the 15D showed moderate to strong correlations with instruments in the other groups, indicating an ability to identify broad aspects of health. The SF-36 physical functioning, bodily pain, and vitality scales were strongly correlated with their related measures, while the physical role limitations, general health perceptions, mental health, social functioning, and emotional role limitations scales all revealed moderate correlations with their related measures.

Reliability: 87 patients reported no change from baseline in RA and overall health status. Agreement between the baseline and 2-week measurements as estimated by ICC was strongest for HAQ and RAQoL, while 15D, VAS global RA and fatigue, SF-36 vitality, and bodily pain followed closely (Table 4). The score differences between the 2 measurements did not deviate notably from the normal distribution, and the 95% limits of agreement could therefore be estimated using means and standard deviations. The mean differences were significantly different from zero in 15D and the SF-36 bodily pain, general health perceptions, and social

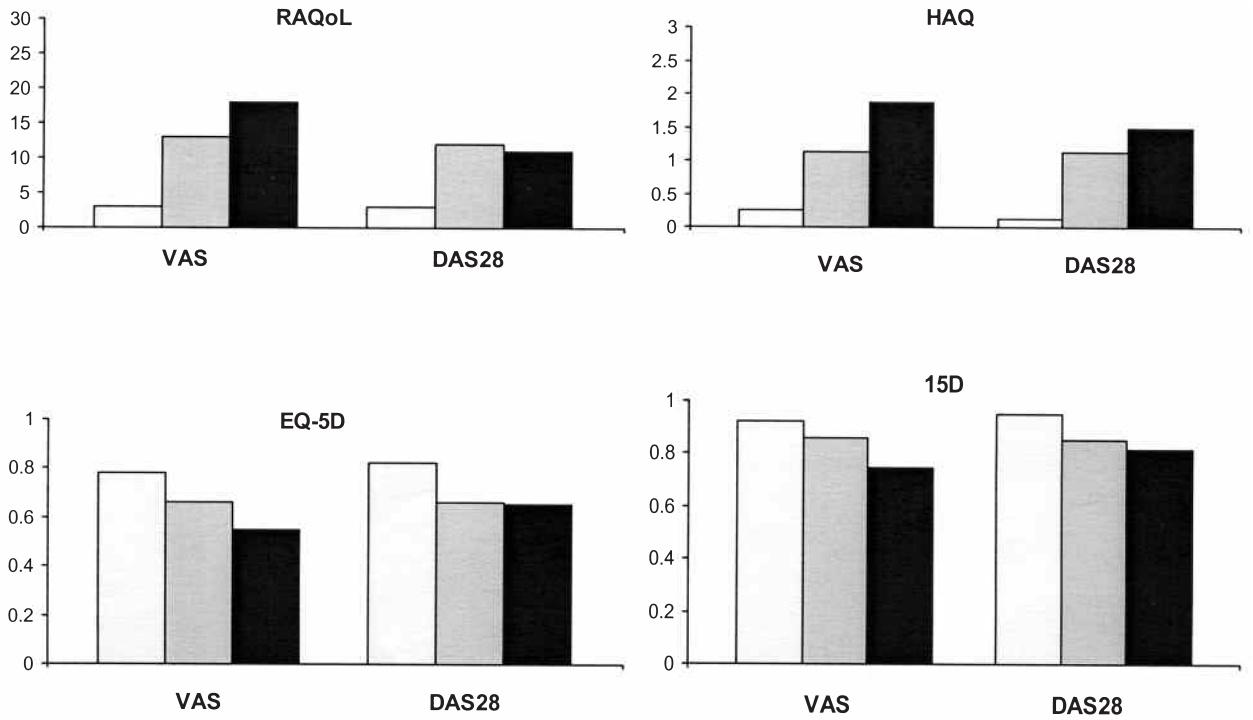

Figure 1. Known-groups validity as illustrated by median score comparisons (RAQoL, HAQ, EQ-5D, and 15D) in patients with low, moderate, and high DAS28 and VAS for self-reported current arthritis activity. VAS for self-reported current arthritis activity (0–10): low, 0–3.3 (white columns); moderate, 3.4–6.6 (gray columns); high, 6.7–10.0 (black columns). Disease activity score based on 28-joint count (DAS28): low, < 3.2 (white); moderate, 3.2–5.1 (gray); high, > 5.1 (black).

Table 2. Known-groups validity. HRQOL scores for patients with known DAS28, VAS for self-reported current arthritis activity, bone erosion, and disability pension status in P1 (n = 200).

Instruments (best/worst score)	DAS 28 Low (< 3.2), n = 89	DAS 28 Moderate (3.2–5.1), n = 73	DAS 28 High (> 5.1), n = 12	VAS for Arthritis Activity Low (0–3.3), n = 115	VAS for Arthritis Activity Moderate (3.4–6.6), n = 43	VAS for Arthritis Activity High (6.7–10.0), n = 18	Presence of Bone Erosions No, n = 63	Presence of Bone Erosions Yes, n = 119	Presence of Bone Erosions Effect Size	Disability Pension No, n = 120	Disability Pension Yes, n = 58	Disability Pension Effect Size
	Median HRQOL score			Median HRQOL score			Mean HRQOL score (SD)			Mean HRQOLscore (SD)		
RAQoL (0/30)	3	12*	11	3	13*	18**	8 (7)	8 (7)	0.00	7 (6)†	11 (8)	**0.71**
HAQ (0/3)	0.13	1.13*	1.50	0.25	1.13*	1.88**	0.69 (0.76)	0.86 (0.78)	0.22	0.62 (0.73)†	1.12 (0.77)	**0.66**
EQ-5D (1/0)	0.82	0.66*	0.66	0.78	0.66*	0.55**	0.77 (0.19)	0.72 (0.19)	0.27	0.76 (0.20)†	0.70 (0.17)	0.33
15D (1/0)	0.95	0.85*	0.82	0.92	0.86*	0.75**	0.89 (0.10)	0.88 (0.09)	0.06	0.89 (0.10)†	0.87 (0.09)	0.30
SF-36 (100/0)												
Physical functioning	80	48*	25**	75	50*	37	64 (28)	57 (27)	0.24	67 (26)†	49 (23)	**0.69**
Role physical	75	25*	0	75	0*	0	48 (40)	46 (43)	0.04	52 (42)	40 (40)	0.29
Bodily pain	74	41*	22**	74	41*	31**	58 (23)	57 (25)	0.01	61 (25)†	50 (22)	**0.50**
General health	70	45*	30	62	45*	25	56 (22)	55 (24)	0.04	57 (23)	54 (25)	0.12
Vitality	75	45*	35	70	45*	25**	58 (27)	58 (25)	0.00	60 (26)	54 (25)	0.24
Social functioning	100	88*	75	100	88*	75**	82 (22)	85 (24)	–0.12	85 (22)	81 (26)	0.17
Role emotional	100	67*	0**	100	33*	33	66 (41)	65 (41)	0.02	68 (40)	61 (44)	0.15
Mental health	92	80*	72	92	76*	58**	78 (21)	80 (18)	–0.10	80 (20)	79 (19)	0.03
VAS pain (0/100)	1.2	4.6*	6.7	1.3	5.1*	7.5**	3.2 (2.8)	3.2 (2.4)	0.00	2.8 (2.4)†	4.0 (2.6)	0.47
VAS fatigue (0/100)	1.5	5.8*	4.9	2.0	6.4*	8.1**	3.8 (3.2)	3.9 (2.9)	0.05	3.7 (3.0)	4.2 (2.9)	0.20
VAS global RA (0/100)	1.0	4.7*	6.4	1.3	5.2*	7.8**	3.2 (2.8)	3.3 (2.6)	0.04	3.0 (2.7)	3.8 (2.5)	0.30

Statistically significant differences between * low/moderate and ** moderate/high DAS28 and VAS for self-reported present arthritis activity (Kruskall-Wallis and Mann-Whitney tests for all but SF-36 physical functioning, bodily pain, general health, and vitality, in which ANOVA was used). † Statistically significant differences between receivers/nonreceivers of disability pension (2-sample t-test). Effect size is calculated as the difference in mean scores divided by the pooled standard deviations. Effect size values for the dichotomous variables are considered small (< 0.5), medium (0.5–0.8), or large (> 0.8). Values in bold type indicate medium effect sizes. DAS: Disease Activity Score. For other abbreviations, see Table 1.

Table 3. Convergent and discriminant validity. Multitrait-multimethod (MTMM) correlation matrix illustrating the associations of the different HRQOL instruments to assess physical function, pain, fatigue, global RA, and overall health (mental/social function not shown) in P1 (n = 200).

Health Aspects	Instruments	Physical Function			Pain		Fatigue		Global RA		Overall Health		
		HAQ	PF	RP	BP	VAS Pain	VT	VAS Fatigue	VAS Global RA	RAQOL	EQ-5D	15D	GH
Physical function	HAQ	1											
	PF	**–0.769**	1										
	RP	**–0.574**	**0.615**	1									
Pain	BP	–0.714	0.625	0.593	1								
	VAS pain	0.714	–0.645	–0.590	**–0.819**	1							
Fatigue	VT	–0.600	0.566	0.631	0.632	–0.617	1						
	VAS fatigue	0.617	–0.513	–0.558	–0.648	0.706	**–0.724**	1					
Global RA	VAS global RA	0.714	–0.636	–0.573	–0.780	–0.905	–0.617	0.741	1				
	RAQoL	0.814	–0.694	–0.650	–0.723	0.750	–0.745	0.779	**0.815**	1			
Overall health	EQ-5D	–0.791	0.725	0.604	0.727	–0.755	0.650	–0.679	–0.757	–0.760	1		
	15D	–0.741	0.690	0.640	0.690	–0.653	0.749	–0.704	–0.673	–0.830	**0.809**	1	
	GH	–0.508	0.511	0.417	0.618	–0.544	0.572	–0.554	–0.569	–0.688	**0.571**	**0.641**	1

HAQ: Health Assessment Questionnaire; PF: SF-36, physical functioning; RP: SF-36, physical role limitations; BP: SF-36, bodily pain; VT: SF-36, vitality; RAQoL: Rheumatoid Arthritis Quality of Life scale; EQ-5D: EuroQol; 15D: 15 dimensions of health; GH: SF-36, general health perceptions. Correlation coefficients (Spearman's rho). All correlations are statistically significant ($p < 0.01$). Values in bold type indicate correlations expected to exceed 0.7 (convergent validity).

Table 4. Reliability and responsiveness in P2 (n = 150). Reliability given by intraclass correlation coefficients (ICC) and 95% limits of agreement for patients reporting no change after 2 weeks. Responsiveness given by standardized response means (SRM) for patients reporting improvement, no change, or deterioration after 6 months.

Health Aspects	Instruments	Reliability (n = 87)			Responsiveness, SRM (n = 96)		
		ICC (95% CI)	95% Limits of Agreement, mean ± 1.96*SD	% of Maximum Score	Improvement (n = 26)	No Change (n = 47)	Deterioration (n = 23)
Physical function	HAQ	0.97 (0.96–0.98)	0 ± 0.38	13	–0.10	–0.26	0.13
	SF-36, PF	0.88 (0.82–0.92)	0 ± 21	21	0.56	0.04	0.11
	SF-36, RP	0.83 (0.74–0.89)	0 ± 61	61	0.66**	0	–0.43
Pain	SF-36, BP	0.90 (0.84–0.93)	4 ± 25*	25	**0.93****	0.10	–0.50
	VAS pain	0.87 (0.80–0.92)	0 ± 34	34	**–0.95****	0.06	0.60
Fatigue	SF-36, VT	0.91 (0.86–0.94)	0 ± 28	28	**0.94****	0.05	–0.37
	VAS fatigue	0.93 (0.90–0.96)	0 ± 29	29	–0.70**	–0.05	0.17
Global RA	VAS global RA	0.91 (0.86–0.94)	0 ± 28	28	–0.60**	–0.15	0.40
	RAQoL	0.96 (0.94–0.97)	0 ± 5	17	–0.67	–0.15	0.17
Overall health	EQ-5D	0.79 (0.68–0.87)	0 ± 0.27	27	0.65**	0.06	–0.36
	15D	0.93 (0.89–0.96)	0.02 ± 0.08*	8	0.54	0.26	–0.30
	SF-36, GH	0.82 (0.73–0.88)	–3 ± 20*	20	–0.28	–0.50	–0.21
Mental/social function	SF-36, MH	0.55 (0.32–0.71)	0 ± 19	19	–0.25	0.06	0.11
	SF-36, RE	0.66 (0.48–0.78)	0 ± 88	88	0.48	0.19	–0.28
	SF-36, SF	0.52 (0.26–0.68)	3 ± 23*	23	0.48	–0.11	–0.06

HAQ: Health Assessment Questionnaire; PF: SF-36, physical functioning; RP: SF-36, physical role limitations; BP: SF-36, bodily pain; VT: SF-36, vitality; RAQoL: Rheumatoid Arthritis Quality of Life scale; EQ-5D: EuroQol; 15D: 15 dimensions of health; GH: SF-36, general health perceptions; MH: mental health; RE: emotional role limitations; SF: social functioning. The 95% limits of agreement represent the interval: mean difference between baseline and 2 week measurements ± 1.96* SD of that difference. * Mean difference significantly different from zero (1 = sample T-test). Values in bold type indicate large SRM values (> 0.8). ** Significant differences in mean score improvements from baseline using ANOVA and adjusting for multiple comparisons by Tukey's method.

functioning scales (Table 4). The 95% limits of agreement interval was narrowest for 15D (0.02 ± 0.08), while HAQ, RAQoL and SF-36 mental health displayed coefficients of repeatability below 20% of the maximum score of the scale (Table 4 and Figures 2A-2D).

The internal consistency as estimated by Cronbach's alpha coefficients on P2 baseline scores was highest for HAQ (0.95) and RAQoL (0.90), while the SF-36 physical functioning (0.89), vitality (0.89), physical role limitations (0.87), bodily pain (0.86), and mental health (0.86) scales exceeded 0.85 (data not shown).

Responsiveness: 96 patients were divided into the 3 sub-

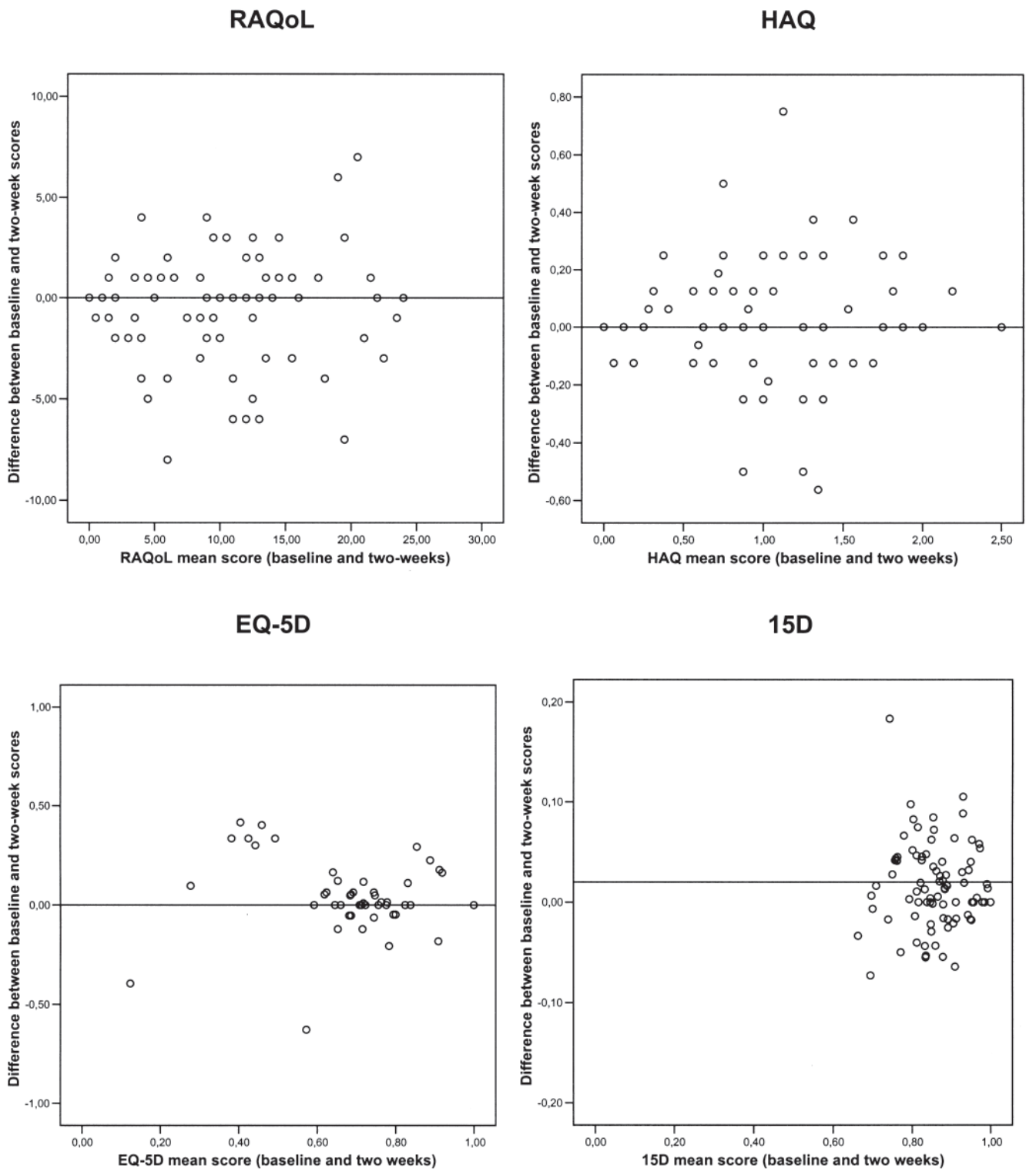

Figure 2. Bland-Altman plots of differences between baseline and 2-week scores on mean score (baseline and 2 wks) for RAQoL, HAQ, EQ-5D, and 15D. The mean difference is marked with a horizontal line.

groups (better, n = 26; no change, n = 47; worse, n = 23). Using ANOVA statistics we found significant differences between the 3 subgroups for all instruments, except the HAQ and SF-36 physical functioning, general health perceptions, social functioning, and mental health dimensions, which were then omitted from further testing. Further exploration revealed statistically significant differences between mean scores changes in the "worse" and "better" subgroups

for all instruments examined and between mean scores changes in the "no changes" and "better" subgroups in all instruments examined except the 15D, RAQoL, and SF-36 emotional role limitations scale.

The responsiveness as reflected by SRM estimates can be seen in Table 4. The pain and fatigue domains of the HRQOL instruments (SF-36, bodily pain, VAS pain, and SF-36 vitality) showed good responsiveness to improvement, while responsiveness to deterioration was limited. In the physical function domain, only the SF-36 physical role limitations scale yielded a statistically significant response to patient-reported improvement from baseline. The HAQ showed no response to patient-reported change in any direction.

DISCUSSION

Ours is the first study to compare the validity, reliability, and responsiveness of a broad array of health status and quality of life measurement instruments in patients with RA. We investigated 3 generic (SF-36, EQ-5D, 15D) and 2 disease-specific instruments (RAQoL, HAQ) along with 3 VAS scales (fatigue, pain, global RA) in 2 subpopulations of outpatients with RA. Our hypothesis was that the disease-specific instruments would perform better than the generic instruments.

We examined construct validity, reliability in terms of repeatability and internal consistency, and responsiveness defined as the ability to detect patient-reported changes in RA and overall health. Our methods are based on traditional psychometric theory and are described in detail by Fayers and Machin[26].

Testing for known-groups validity, we observed the strongest relationship between disease severity and the SF-36 physical functioning and bodily pain scales, as they managed to successfully discriminate between the different DAS28, VAS for self-reported current arthritis activity, and disability pension groups. The RAQoL and HAQ followed closely, but failed to discriminate between patients with moderate and high DAS28. However, the findings must be regarded with caution because of the limited sample size in the group of patients with severe DAS28. Moreover, since the VAS for current arthritis activity is self-reported, we were not surprised to observe the most consistent discriminative ability in this category. In a validation study of SF-36, Kosinski, *et al*[34] also report superior known-groups validity of the physical functioning and bodily pain scales, while Kvien, *et al*[29] compared the SF-36 with disease-specific measures [modified HAQ, Arthritis Impact Measurement Scales (AIMS2), and VAS] and found equal abilities to detect differences in disease activity. Several studies have found good known-groups validity of the RAQoL[8-10]; the only other study to compare with generic measures favored the RAQoL and HAQ over EQ-5D, the Health Utilities Index-2 (HUI2), HUI3, and SF-6D in sensitivity to self-reported RA severity and control[35]. In our study, the instruments generally discriminated less well between the patients with moderate and high disease activity than between those with low and moderate disease activity. This drawback may be of lesser importance in the future, as treatment options continue to improve and thereby provide decreased disease activity levels for the majority of patients. None of the instruments could differentiate between patients with erosive and nonerosive disease. A possible explanation for this is that we did not take the extent of erosions into account, which may have resulted in an erosive group including a wide range of erosion states. Moreover, Pincus and Sokka have pointed out that radiographic findings are only weakly correlated with measures of pain and functional status[36]. Our findings are, however, in contrast with a study by Ruta, *et al*[37], in which both the physical and mental component scores managed to discriminate between patients with and without joint erosions.

The convergent and discriminant validity analysis revealed a large number of strong correlations, indicating a high degree of interrelatedness of the investigated health aspects. The RAQoL and the HAQ were highly associated with 4 of the 5 defined health aspect groups; nevertheless they did not convincingly stand out from EQ-5D and 15D. This is in agreement with other studies of the RAQoL that reported strong associations with relevant Nottingham Health Profile sections[8,10] and the SF-36[9]. Convergent validity was confirmed in the SF-36 physical functioning, vitality, and bodily pain scales, while the remaining 5 scales did not fit into our model. Thus, the social functioning and mental health scales showed stronger correlations with the vitality measures than with each other, while the tendencies for the general health, physical, and emotional role limitation scales are more unclear. Kvien, *et al*[29] found similar correlation patterns; Ruta, *et al*[37] support the convergent validity of the physical functioning and bodily pain scales, while Talamo, *et al*[38] support the convergent validity of the physical functioning scales by a strong correlation (0.72) with the HAQ.

The RAQoL and HAQ displayed excellent repeatability and internal consistency as illustrated by ICC and Cronbach's alpha values above 0.90. This is in concordance with other studies revealing test-retest correlation coefficients between 0.90 and 0.94 (RAQoL) and 0.96 and 0.97 (HAQ)[8,10,14]. Two SF-36 scales (vitality, bodily pain), 15D, and VAS fatigue and global RA also exceeded an ICC of 0.90, which according to Nunally and Bernstein[32] qualifies for evaluation at the level of the individual. Using the Bland-Altman approach, the level of random error was less than 20% of the maximum score only for 15D, HAQ, RAQoL, and the SF-36 mental health scale. This has to be taken into account in clinical practice, since the change in scores at the individual level must exceed the level of random error in order to reflect a real difference in health state or quality of

life. Russell, *et al*[39] also estimated the random error of the EQ-5D, SF-36, and VAS pain using the Bland-Altman approach. Their results are comparable to ours for VAS pain, SF-36 physical functioning, vitality, and mental health scales, while they report greater random error for EQ-5D.

In general, responsiveness to patient-reported improvement was better than responsiveness to deterioration. Two studies on SF-36[37] and EQ-5D versus HAQ[40] are in agreement with our findings, while a study of SF-36, AIMS2, and MHAQ by Hagen, *et al*[15] shows equally good responsiveness to improvement and deterioration. As expected, the instruments measuring pain and fatigue were the most responsive to the relatively short-term changes in our study. We could not, however, show responsiveness in the instruments covering physical function, which suggests these are more related to longterm changes, such as structural joint damage. Contrary to what we had expected, the disease-specific instruments (RAQoL and HAQ) did not show superior ability in detecting patient-reported changes in RA and overall health. These findings are in contrast to a study by Marra, *et al*[14], where the RAQoL and HAQ performed better than the EQ-5D when using a patient transition question. However, when the authors applied the VAS for self-reported disease activity as an external indicator of change, the responsiveness statistic was boosted, and the difference between the generic and disease-specific measures was evened out. Other studies using patient-reported indicators for change did not show superior responsiveness of disease-specific over generic measures. Using clinical indicators for change, Wells, *et al*[11] studied 40 patients with RA who initiated oral methotrexate, and found the HAQ to be more responsive (SRM > 0.8) to the treatment effect than the RAQoL and SF-36 after 6 months' therapy. One limitation to our responsiveness analysis is the small sample size, which prevents firm conclusions. Moreover, we lack clinical measures to assess whether the patients improved or not, which in turn means that our results are not directly transferable to clinical studies, where the traditional endpoints are disease activity measures. Data regarding disease duration is equally important, because of its influence on the reversibility of the HAQ[41].

A high number of missing answers to some questions led us to suspect that they were not unanswered at random. To explain this, we consider that questions regarding sexual activities are always controversial, and people choose not to answer for various reasons (15D). A bathtub is not standard bathroom inventory in Danish homes (HAQ). More than half of the P1 population had retired due either to age or to disability, which could have reduced their response rate to questions regarding work (SF-36, physical and emotional role limitation scales).

The RAQoL revealed no real flaws: it had good construct validity, reliability, and responsiveness. The HAQ showed overall good psychometric properties, yet the responsiveness was disappointingly low. The SF-36 physical functioning, bodily pain, and vitality scales seemed to be well suited for patients with RA, while the performance of the remaining scales was more variable. The 15D displayed good psychometric properties, but the ceiling effect was marked and missing answers were a problem in the sexual activities dimension. The EQ-5D was easy to complete, construct validity and responsiveness acceptable, while reliability was fairly poor.

Comparison of the validity, reliability, and responsiveness of 3 generic and 2 disease-specific instruments showed, surprisingly, that the disease-specific instruments did not perform conclusively better than the generic. Each instrument revealed strengths and weaknesses, which prevented the recommendation of one instrument for all purposes in favor of the others. In future studies of HRQOL in RA, the choice of instrument should therefore be guided by the specific purpose of the study.

ACKNOWLEDGMENT

We thank Tuulikki Sokka for allowing us to use the QUEST data and for giving valuable comments on the manuscript. Medical student Lykke Ørnbjerg aided with data collection.

REFERENCES

1. Felson DT, Anderson JJ, Boers M, et al. American College of Rheumatology. Preliminary definition of improvement in rheumatoid arthritis. Arthritis Rheum 1995;38:727-35.
2. Buchbinder R, Bombardier C, Yeung M, Tugwell P. Which outcome measures should be used in rheumatoid arthritis clinical trials? Clinical and quality-of-life measures' responsiveness to treatment in a randomized controlled trial. Arthritis Rheum 1995;38:1568-80.
3. Tugwell P, Wells G, Strand V, et al. Clinical improvement as reflected in measures of function and health-related quality of life following treatment with leflunomide compared with methotrexate in patients with rheumatoid arthritis: sensitivity and relative efficiency to detect a treatment effect in a twelve-month, placebo-controlled trial. Leflunomide Rheumatoid Arthritis Investigators Group. Arthritis Rheum 2000;43:506-14.
4. Guyatt GH, Feeny DH, Patrick DL. Measuring health-related quality of life. Ann Intern Med 1993;118:622-9.
5. Hurst NP, Jobanputra P, Hunter M, Lambert M, Lochhead A, Brown H. Validity of Euroqol — a generic health status instrument — in patients with rheumatoid arthritis. Economic and Health Outcomes Research Group. Br J Rheumatol 1994;33:655-62.
6. Wolfe F, Hawley DJ. Measurement of the quality of life in rheumatic disorders using the EuroQol. Br J Rheumatol 1997;36:786-93.
7. Fries JF, Spitz P, Kraines RG, Holman HR. Measurement of patient outcome in arthritis. Arthritis Rheum 1980;23:137-45.
8. Thorsen H, Hansen TM, McKenna SP, Sorensen SF, Whalley D. Adaptation into Danish of the Stanford Health Assessment Questionnaire (HAQ) and the Rheumatoid Arthritis Quality of Life Scale (RAQoL). Scand J Rheumatol 2001;30:103-9.
9. Tijhuis GJ, de Jong Z, Zwinderman AH, et al. The validity of the Rheumatoid Arthritis Quality of Life (RAQoL) questionnaire. Rheumatology Oxford 2001;40:1112-9.
10. de Jong Z, van der Heijde D, McKenna SP, Whalley D. The reliability and construct validity of the RAQoL: a rheumatoid arthritis-specific quality of life instrument. Br J Rheumatol 1997;36:878-83.

11. Wells G, Boers M, Shea B, et al. Sensitivity to change of generic quality of life instruments in patients with rheumatoid arthritis: preliminary findings in the generic health OMERACT study. OMERACT/ILAR Task Force on Generic Quality of Life. Life Outcome Measures in Rheumatology. International League of Associations for Rheumatology. J Rheumatol 1999;26:217-21.
12. Wiebe S, Guyatt G, Weaver B, Matijevic S, Sidwell C. Comparative responsiveness of generic and specific quality-of-life instruments. J Clin Epidemiol 2003;56:52-60.
13. Salaffi F, Stancati A, Carotti M. Responsiveness of health status measures and utility-based methods in patients with rheumatoid arthritis. Clin Rheumatol 2002;21:478-87.
14. Marra CA, Rashidi AA, Guh D, et al. Are indirect utility measures reliable and responsive in rheumatoid arthritis patients? Qual Life Res 2005;14:1333-44.
15. Hagen KB, Smedstad LM, Uhlig T, Kvien TK. The responsiveness of health status measures in patients with rheumatoid arthritis: comparison of disease-specific and generic instruments. J Rheumatol 1999;26:1474-80.
16. Sokka T, Kautiainen H, Toloza S, et al. QUEST-RA: Quantitative clinical assessment of patients with rheumatoid arthritis seen in standard rheumatology care in 15 countries. Ann Rheum Dis 2007;66:1491-6.
17. Bjorner JB, Thunedborg K, Kristensen TS, Modvig J, Bech P. The Danish SF-36 Health Survey: translation and preliminary validity studies. J Clin Epidemiol 1998;51:991-9.
18. Szende A, Oppe M, Devlin N, editors. EQ-5D value sets: Inventory, comparative review and user guide. EuroQol Group Monographs. Vol. 2. New York: Springer; 2007.
19. Wittrup-Jensen K. Measurement and valuation of health-related quality of life [thesis]. University of Southern Denmark; 2006.
20. Sintonen H. The 15D instrument of health-related quality of life: properties and applications. Ann Med 2001;33:328-36.
21. Fries JF, Spitz PW, Young DY. The dimensions of health outcomes: the Health Assessment Questionnaire, disability and pain scales. J Rheumatol 1982;9:789-93.
22. McHorney CA, Ware JE Jr, Lu JF, Sherbourne CD. The MOS 36-item Short-Form Health Survey (SF-36): III. Tests of data quality, scaling assumptions, and reliability across diverse patient groups. Med Care 1994;32:40-66.
23. McHorney CA, Ware JE Jr, Raczek AE. The MOS 36-Item Short-Form Health Survey (SF-36): II. Psychometric and clinical tests of validity in measuring physical and mental health constructs. Med Care 1993;31:247-63.
24. Ware JE Jr, Sherbourne CD. The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. Med Care 1992;30:473-83.
25. Bruce B, Fries JF. The Stanford Health Assessment Questionnaire: a review of its history, issues, progress, and documentation. J Rheumatol 2003;30:167-78.
26. Fayers P, Machin D. Quality of life: the assessment, analysis and interpretation of patient-reported outcomes. 2nd ed. West Sussex: Wiley; 2007.
27. Prevoo ML, van 't Hof MA, Kuper HH, van Leeuwen MA, van de Putte LB, van Riel PL. Modified disease activity scores that include twenty-eight-joint counts. Development and validation in a prospective longitudinal study of patients with rheumatoid arthritis. Arthritis Rheum 1995;38:44-8.
28. Cohen J. Statistical power for the behavioral sciences. 2nd ed. Hillsdale, NJ: Erlbaum Associates; 1988.
29. Kvien TK, Kaasa S, Smedstad LM. Performance of the Norwegian SF-36 Health Survey in patients with rheumatoid arthritis. II. A comparison of the SF-36 with disease-specific measures. J Clin Epidemiol 1998;51:1077-86.
30. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet 1986;1:307-10.
31. Cronbach L. Coefficient alpha and the internal structure of tests. Psychometrika 1951;16:297-334.
32. Nunally J, Bernstein I. Psychometric theory. New York: McGraw-Hill, Inc.; 1994.
33. Guyatt GH, Deyo RA, Charlson M, Levine MN, Mitchell A. Responsiveness and validity in health status measurement: a clarification. J Clin Epidemiol 1989;42:403-8.
34. Kosinski M, Keller SD, Ware JE Jr, Hatoum HT, Kong SX. The SF-36 Health Survey as a generic outcome measure in clinical trials of patients with osteoarthritis and rheumatoid arthritis: relative validity of scales in relation to clinical measures of arthritis severity. Med Care 1999;37 Suppl:MS23-39.
35. Marra CA, Woolcott JC, Kopec JA, et al. A comparison of generic, indirect utility measures (the HUI2, HUI3, SF-6D, and the EQ-5D) and disease-specific instruments (the RAQoL and the HAQ) in rheumatoid arthritis. Soc Sci Med 2005;60:1571-82.
36. Pincus T, Sokka T. Quantitative measures for assessing rheumatoid arthritis in clinical trials and clinical care. Best Pract Res Clin Rheumatol 2003;17:753-81.
37. Ruta DA, Hurst NP, Kind P, Hunter M, Stubbings A. Measuring health status in British patients with rheumatoid arthritis: reliability, validity and responsiveness of the short form 36-item health survey (SF-36). Br J Rheumatol 1998;37:425-36.
38. Talamo J, Frater A, Gallivan S, Young A. Use of the short form 36 (SF-36) for health status measurement in rheumatoid arthritis. Br J Rheumatol 1997;36:463-9.
39. Russell AS, Conner-Spady B, Mintz A, Maksymowych WP. The responsiveness of generic health status measures as assessed in patients with rheumatoid arthritis receiving infliximab. J Rheumatol 2003;30:941-7.
40. Hurst NP, Kind P, Ruta D, Hunter M, Stubbings A. Measuring health-related quality of life in rheumatoid arthritis: validity, responsiveness and reliability of EuroQol (EQ-5D). Br J Rheumatol 1997;36:551-9.
41. Aletaha D, Smolen J, Ward MM. Measuring function in rheumatoid arthritis: Identifying reversible and irreversible components. Arthritis Rheum 2006;54:2784-92.