The Journal of Rheumatology Volume 30, no. 10

A comparison of the EuroQol-5D and the Health Utilities Index mark 3 in patients with rheumatic disease.

Nan Luo, Ling-Huo Chew, Kok-Yong Fong, Dow-Rhoon Koh, Swee-Cheng Ng, Kam-Hon Yoon, Sheila Vasoo, Shu-Chuen Li and Julian Thumboo

J Rheumatol 2003;30;2268-2274
http://www.jrheum.org/content/30/10/2268

A Comparison of the EuroQol-5D and the Health Utilities Index Mark 3 in Patients with Rheumatic Disease

NAN LUO, LING-HUO CHEW, KOK-YONG FONG, DOW-RHOON KOH, SWEE-CHENG NG, KAM-HON YOON, SHEILA VASOO, SHU-CHUEN LI, and JULIAN THUMBOO

ABSTRACT. ***Objective.*** To compare the performance of 2 commonly used utility-based health-related quality of life (HRQoL) instruments [the EuroQol-5D (EQ-5D) and Health Utilities Index mark 3 (HUI3)] in patients with rheumatic disease.

Methods. Consecutive outpatients with rheumatic diseases were interviewed twice within 2 weeks using a standard questionnaire containing the EQ-5D, HUI3, and the Medical Outcome Study Short-Form 36 Health Survey (SF-36, used to categorize health status) and assessing clinical and demographic characteristics. EQ-5D and HUI3 utility scores were compared and their construct validity and test-retest reliability were examined by comparing these scores in groups differing in health status and using intraclass correlation coefficients (ICC), respectively.

Results. EQ-5D and HUI3 utility scores in 114 patients differentiated well between varying health states; e.g., patients with higher SF-36 vitality scores had better EQ-5D and HUI3 utility scores (mean: 0.79 for both instruments) than patients with lower vitality scores (mean: 0.68 and 0.69, respectively) ($p < 0.01$ for both comparisons). ICC values for the EQ-5D and HUI3 were 0.64 and 0.75, respectively ($n = 90$, median interval: 7 days). EQ-5D and HUI3 utility scores were similar (mean ± SD: 0.75 ± 0.21 vs 0.76 ± 0.17, $p = 0.647$, paired t test) and showed moderate correlation (Spearman's ρ: 0.45, $p < 0.001$). Differences were present in patients' responses to these 2 instruments: e.g., 12 patients reporting no problems with mobility (EQ-5D item) reported different levels of disability with ambulation (HUI3 item).

Conclusions. The EQ-5D and HUI3 performed equally well in measuring utility-based HRQoL in patients with rheumatic disease, although they measured slightly different, though related, dimensions of health. (J Rheumatol 2003;30:2268–74)

Key Indexing Terms:
QUALITY OF LIFE COMPARATIVE STUDY RHEUMATIC DISEASES SINGAPORE

Health-related quality of life (HRQoL) as measured by utility-based instruments is a useful outcome measure in clinical research. By summarizing a respondent's overall HRQoL into a single utility score, such instruments facilitate the calculation of quality-adjusted life-years (QALY) that are used to conduct cost-utility analysis of clinical interventions[1,2]. The EuroQol-5D (EQ-5D)[3,4] and health utilities index (HUI)[5,6] are 2 utility-based HRQoL instruments that have been widely used in such clinical studies. Both instruments classify a respondent's self-reported health status according to a specific descriptive/classification system and assign the respondent a utility score from a scale on which 1 represents full health and 0 represents being dead. Such scores reflect the preference for or desirability of individual health states from the societal perspective[1].

Both the EQ-5D and HUI have shown great promise in outcome studies of rheumatic disease. The EQ-5D has shown acceptable psychometric properties in patients with rheumatoid arthritis (RA)[7,8] and osteoarthritis (OA)[9,10] and has been used as a HRQoL measure in clinical studies in a variety of rheumatic conditions[11-17]. The HUI Mark 2 (HUI2) has also shown acceptable psychometric properties in patients with systemic lupus erythematosus (SLE)[18], while the HUI Mark 3 (HUI3) has demonstrated construct validity in patients with self-reported arthritis/rheumatism[19] and been used in a clinical trial to evaluate the clinical and

From the Department of Pharmacy, and the Department of Medicine, National University of Singapore; the School of Health Sciences, Nanyang Polytechnic; and the Department of Medicine, National University Hospital, Singapore.

N. Luo, MSc, Research Scholar; S-C. Li, PhD, Assistant Professor, Department of Pharmacy, National University of Singapore; L-H. Chew, BAppSc(NEd), Lecturer, School of Health Sciences, Nanyang Polytechnic; K-Y. Fong, FRCP (Edin), Associate Professor; D-R. Koh, PhD, Associate Professor; K-H. Yoon, MRCP (UK), Assistant Professor; J. Thumboo, FRCP (Edin), Associate Professor, Department of Medicine, National University of Singapore and Department of Medicine, National University Hospital; S-C. Ng, MMed (Internal Medicine), Visiting Consultant; S. Vasoo MRCP (UK), Registrar, Department of Medicine, National University Hospital.

Address reprint requests to Dr. J. Thumboo, Department of Rheumatology and Immunology, Singapore General Hospital, Outram Road, Singapore, 169608. E-mail: julian_thumboo@sgh.com.sg

Submitted December 16, 2002; revision accepted March 28, 2003.

economic outcomes of a treatment paradigm for patients with knee OA[20,21]. Further, both instruments discriminate well between respondents with and without arthritis/rheumatism in population surveys[19,22,23].

However, there are gaps in our knowledge about the roles of these 2 instruments in studying patients with rheumatic disease. First, it is not clear how comparable EQ-5D and HUI3 scores are in such patients, or whether one instrument has better psychometric properties than the other. To the best of our knowledge, there are no comparative studies of these instruments in patients with rheumatic disease, and the few published studies comparing the EQ-5D and HUI3 in patients with non-rheumatic diseases have yielded conflicting results. In a study of patients with intermittent claudication, investigators found patients had similar EQ-5D and HUI3 scores when they were in better health but differing scores when they were in worse health[24]. In a population survey, however, investigators found that mean EQ-5D and HUI3 scores differed in healthier populations but were similar in less healthy populations[23]. The mean HUI3 score of respondents with self-reported arthritis/rheumatism was higher than the mean EQ-5D score (0.71 vs 0.68, n = 341)[23]. Second, to our knowledge, validity and reliability of the HUI3 in patients with physician-diagnosed rheumatic disease have not been reported. Third, it is unclear which instrument is more suitable for use in patients with rheumatic disease. Although both instruments can be used for QALY calculation/cost-utility analysis, it would not be ideal to use more than one such instrument in clinical studies, where the patient burden of completing such instruments is a major concern.

Our primary aim was therefore to compare the performance of the EQ-5D and HUI3 in patients with rheumatic disease by comparing utility scores, test-retest reliability, and construct validity of these instruments. A secondary aim was to further assess the validity and reliability of the HUI3 in patients with physician-diagnosed rheumatic disease.

MATERIALS AND METHODS

Patients and study design. Consecutive patients seen during a 2-week period in February 2001 at the rheumatology outpatient clinic of a tertiary referral hospital in Singapore (an urban, multi-ethnic country with a population of 3 million, located in South-East Asia) were recruited for this institutional review board approved study. Inclusion criteria were physician diagnosis of a rheumatic disease and literacy in English or Chinese. After providing written consent, each patient was interviewed in English or Chinese by a trained nurse interviewer using a standardized, pre-tested questionnaire (available in identical English or Chinese versions) including the EQ-5D self-report questionnaire (EQ-5D), the health descriptive system of HUI Mark 3 (HUI3), the Medical Outcome Study Short-Form 36 Health Survey (SF-36)[25], and assessing socio-demographic characteristics, acute and chronic medical conditions, and self-reported pain [using a 10 cm pain visual analog scale (VAS)]. Patients were also examined for fibromyalgia tender points by one trained nurse examiner[26]. Followup telephone interviews (3 attempts) using the EQ-5D and the HUI3 descriptive systems were conducted 1 week after the baseline interview.

Instruments. The EQ-5D self-report questionnaire consists of a descriptive system and a visual analog scale (EQ-VAS)[4]. The descriptive system has 5 dimensions (i.e., mobility, self-care, usual activities, pain/discomfort, and anxiety/depression), with 3 response options for each dimension (i.e., no problems, moderate problems, and extreme problems). The EQ-VAS is a vertical, graduated 20 cm thermometer (0-100 points), with 100 representing best imaginable health state and 0 representing worst imaginable health state. Respondents classify and rate their health on the day of the survey. Their responses to the 5-dimension descriptive system can be converted into a utility score by using an EQ-5D value set derived from a general population[4]. A widely used value set was developed by Dolan using a representative sample of the UK general population[27] that generates utility scores for all the 243 possible health states defined by the EQ-5D descriptive system. Utility scores range from –0.59 to 1.00, with 0 being dead and 1.00 the state of full health.

Singaporean English[28] and Chinese[29] versions of EQ-5D used in this study were developed using EuroQol Group guidelines and input. The Singaporean English EQ-5D differs from the original English version in only one respect: the word "box" in the instructions for the EQ-VAS was replaced with "BLACK BOX" to improve compliance with the instruction to link the box representing a respondent's health state to the EQ-VAS. A similar revision was adopted for the Singaporean Chinese EQ-5D. Both language versions have been conceptually and psychometrically validated[28,29] and a preliminary analysis suggests that language of administration does not generally influence responses to these Singaporean EQ-5D versions[30].

The HUI3 is the latest member of the HUI family. The descriptive system of the HUI3 comprises 8 dimensions (i.e., vision, hearing, speech, ambulation, dexterity, emotion, cognition, and pain), with each dimension having 5 to 6 response levels. The system is able to define 972,000 unique health states. Using multi-attribute utility theory[31], Feeny and colleagues[32] developed a utility function that assigns a utility score ranging from –0.36 to 1.00 (0 representing being dead and 1.00 the state of full health) to each of these health states, based on a survey of the general population of Hamilton, Ontario, Canada. The English and Chinese[33] versions of the HUI3 descriptive system were used in our study, with a recall period of 4 weeks.

The SF-36 is a validated[34] 36-item instrument measuring perceived health in 8 dimensions: physical functioning, role limitations due to physical problems, bodily pain, general health, vitality, social functioning, role limitations due to emotional problems, and mental health, with higher scores (range 0-100) reflecting better perceived health. Respondents rate their health status in terms of these dimensions for the past 4 weeks. The UK English[25] and Hong Kong Chinese[35] versions of the SF-36 were used in this study. Both versions have been validated for use in Singapore and have very similar psychometric properties[36].

Data analysis. EQ-5D and HUI3 utility scores were calculated using scoring functions developed by Dolan[27] and Feeny and colleagues[32], respectively. Construct validity of the 2 instruments was investigated by examining known-groups and divergent construct validity[37]. We hypothesized that EQ-5D and HUI3 scores would be higher in patients with better HRQoL (measured using the SF-36)[38] or health status (measured using several clinical variables)[8,23,38]. To test this hypothesis, we dichotomized patients into subgroups with better or worse HRQoL/health status and compared EQ-5D and HUI3 scores in these groups using unpaired t tests and Mann-Whitney U tests of significance as appropriate. Patients were dichotomized using the median point of SF-36 scale scores, pain VAS scores, or the presence of tender points, or of acute or chronic medical conditions. The construct validity of these 2 instruments was further investigated by examining the correlation between their utility scores and SF-36 scores. Based on the literature, we anticipated a strong correlation between EQ-5D and HUI3 utility scores (convergent construct validity)[23,39], while poor to moderate correlations were expected between EQ-5D and SF-36 scores and between HUI3 and SF-36 scores (divergent construct validity)[9,38,40]. Revicki and Kaplan[40] found that profile-based instruments (e.g. the SF-36) are only moderately correlated with utility-based instruments such as the EQ-5D (correlation coefficient: 0.3 to 0.6). For purposes

of comparison, we adopted this definition of moderate correlation for this study. Test-retest reliability of EQ-5D and HUI3 scores was examined using single measure intraclass correlation coefficients (ICC).

To investigate the comparability of EQ-5D and HUI3 scores, we compared the mean scores for the entire group and for subgroups with better or worse health status using paired t tests and Wilcoxon rank-sum tests. We also examined agreement between responses to corresponding dimensions of these instruments (i.e., EQ-5D mobility item vs HUI3 ambulation item, EQ-5D pain/discomfort item vs HUI3 pain item, and EQ-5D anxiety/depression item vs HUI3 emotion item). Statistical analyses were performed using SPSS for Windows version 11.0.1 (SPSS Inc, Illinois, USA).

RESULTS

Characteristics of patients. One hundred and fourteen patients (OA: 24; RA: 49; SLE: 31; and spondyloarthropathy: 10) completed baseline interviews, and 94 patients (82.5%) completed followup telephone interviews (median interval: 7 days, interquartile range: 7-12 days, English questionnaire completed by 66 patients at baseline and 52 patients at followup). As expected, English-speaking patients were younger (mean age: 44 vs 57 yrs, $p < 0.001$) and better educated (> 6 years of schooling: 87.9% vs 35.4%, $p < 0.001$). There were no missing data for baseline interviews for the EQ-5D or HUI3; 4 patients had 1 missing item for either instrument in followup telephone interviews. Typical administration time was 40 min for baseline and 20 min for followup interviews.

The mean (SD) age of patients was 49 (16.4) yrs; the majority were female (81.6%) and ethnic Chinese (Chinese: 79.8%, Indian: 15.8%, Malay: 4.4%), with 55 patients (48.2%) reporting at least one chronic medical condition and 51 (44.7%) reporting acute medical conditions over the preceding 4 weeks. Mean (SD) and median (interquartile) pain VAS, EQ-VAS, and tender point scores at baseline were 3.9 (2.6) and 3.9 (1.9-6.0), 66.4 (15.5) and 67.1 (58.8-77.3), and 3.7 (4.5) and 2 (0-6), respectively.

EQ-5D: Construct validity, reliability, and correlation with SF-36 scores. Mean (SD) and median (interquartile) EQ-5D scores were 0.75 (0.21) and 0.79 (0.73-0.79), respectively. Responses to the EQ-5D health descriptive system are summarized in Table 1. Pain/discomfort was the dimension most patients (78.1%) reported. The system classified patients into 16 unique health states, with 19 patients classified as having full health. As expected, patients with better health status had higher mean and median EQ-5D scores than patients with worse health status, supporting known-groups construct validity of the EQ-5D ($p < 0.05$ for 9 of 12 comparisons, Table 3). Patients with acute or chronic medical conditions or tender points had lower EQ-5D scores than patients without these conditions; however, these differences did not reach statistical significance. Correlations between EQ-5D and SF-36 scores ranged from 0.23 to 0.55 (Spearman's ρ, $p < 0.05$ for all), with the SF-36 role-emotional and bodily pain scales showing the lowest and highest correlations, respectively. Test-retest reliability of EQ-5D scores as measured by the intraclass correlation coefficient was 0.64 [95% confidence interval (CI): 0.50-0.74].

Table 1. Responses to EQ-5D dimensions.

Dimension	n	%
Mobility		
No problems	90	78.9
Some problems	24	21.2
Extreme problems	0	0
Self-care		
No problems	112	98.2
Some problems	2	1.8
Extreme problems	0	0
Usual activities		
No problems	93	81.6
Some problems	19	16.7
Extreme problems	2	1.8
Pain/discomfort		
No pain/discomfort	25	21.9
Moderate pain/discomfort	83	72.8
Extreme pain/discomfort	6	5.3
Anxiety/depression		
No anxiety/depression	70	61.4
Moderate anxiety/depression	41	36.0
Extreme anxiety/depression	3	2.6

HUI3: Construct validity, reliability, and correlation with SF-36 scores. Mean (SD) and median (interquartile) HUI3 scores were 0.76 (0.17) and 0.79 (0.68-0.88), respectively. Responses to the HUI3 health descriptive system are summarized in Table 2. As seen with the EQ-5D, pain was the dimension most patients (89.5%) reported. The HUI3 system classified patients into 72 unique health states, with 4 patients classified as having full health. As expected, patients with better health status had higher mean and median HUI3 scores than patients with worse health status ($p < 0.05$ for 10 of 12 comparisons, Table 3), supporting known-groups construct validity of the HUI3. Patients with acute medical conditions had lower scores than patients without these conditions, but this did not reach statistical significance. Patients with and without chronic medical conditions had the same mean HUI3 scores. Correlations between HUI3 and SF-36 scores ranged from 0.29 to 0.49 (Spearman's ρ, $p < 0.01$ for all), with the SF-36 physical functioning and bodily pain scales showing the lowest and highest correlations, respectively. Test-retest reliability of HUI3 scores as measured by the intraclass correlation coefficient was 0.75 (95% CI: 0.65-0.83).

Comparison of the EQ-5D and HUI3. EQ-5D and HUI3 scores for all patients were remarkably similar, with mean scores differing by 0.01 points (0.75 vs 0.76, $p = 0.647$, paired t test) and median scores being identical (0.79 points). Similarly, EQ-5D and HUI3 scores were very similar when compared in subgroups of patients categorized by SF-36 scores and clinical variables (Table 3). In comparing these subgroups, the difference between mean

Table 2. Responses to HUI3 dimensions. In each dimension, a value of 1 indicates full ability, while higher values indicate progressively greater levels of disability.

Dimension	n	%
Vision		
1	46	40.4
2	61	53.5
3	4	3.5
4	2	1.8
5	1	0.9
6	0	0
Hearing		
1	109	95.6
2	4	3.5
3	1	0.9
4–6	0	0
Speech		
1	110	96.5
2	4	3.5
3–5	0	0
Ambulation		
1	90	78.9
2	16	14.0
3	7	6.1
4	1	0.9
5–6	0	0
Dexterity		
1	88	77.2
2	25	21.9
3	0	0
4	1	0.9
5–6	0	0
Emotion		
1	59	51.8
2	44	38.6
3	10	8.8
4	1	0.9
5	0	0
Cognition		
1	74	64.9
2	12	10.5
3	23	20.2
4	3	2.6
5	2	1.8
6	0	0
Pain		
1	12	10.5
2	58	50.9
3	36	31.6
4	8	7.0
5	0	0

EQ-5D and HUI3 scores was no larger than 0.03 points, while the difference between median EQ-5D and HUI3 scores was no larger than 0.06 points ($p > 0.05$ for all, paired t tests and Wilcoxon rank-sum tests).

In patients with better HRQoL or health status, differences between mean EQ-5D and HUI3 scores did not follow any particular pattern; while in patients with worse HRQoL or health status, mean EQ-5D scores were generally lower than mean HUI3 scores (by up to 0.03 points). Median EQ-5D scores were generally lower than median HUI3 scores for patients differing in HRQOL or health status (by up to 0.06 points).

Although mean and median EQ-5D and HUI3 scores were very similar, the correlation between EQ-5D and HUI3 scores for all patients was 0.45 for baseline interviews and 0.57 for followup interviews (Spearman's ρ, $p < 0.001$ for both). To better understand the relationship between EQ-5D and HUI3 scores, we studied responses to corresponding dimensions of these instruments. Differences in these responses were noted (Table 4). For example, in baseline interviews, 12 patients reporting no problems with the EQ-5D mobility dimension reported varying levels of disability with the HUI3 ambulation dimension. Correlations between responses to corresponding dimensions of the EQ-5D and HUI3 ranged from 0.33 to 0.52 (Spearman's ρ, $p < 0.001$ for all, Table 4).

DISCUSSION

We compared utility scores and psychometric properties of the EQ-5D and HUI3 in patients with rheumatic disease. Both instruments resulted in very similar mean and median scores, and good levels of construct validity and test-retest reliability. However, correlation between EQ-5D and HUI3 scores and concordance of responses to corresponding items of these 2 instruments were moderate at best, suggesting that these instruments measure slightly different, though related, dimensions of health. To the best of our knowledge, this is one of the few clinical studies comparing these widely used utility-based HRQoL instruments, and is the first such study in patients with physician-diagnosed rheumatic disease.

The EQ-5D and HUI3 both showed acceptable known-groups construct validity and comparable test-retest reliability. Mean and median EQ-5D and HUI3 scores were very similar, suggesting that these instruments provide comparable utility scores at group level for patients with rheumatic disease. Very similar mean EQ-5D and HUI3 scores were also reported for patients treated for intermittent claudication[25] and younger respondents (aged 15-34 years) in a population survey[23]. Mean EQ-5D and HUI3 scores were as hypothesized for 9 and 10 (of 12) comparisons, respectively, of patients with better versus worse health status, suggesting that both instruments have sufficient discriminative power for cross-sectional studies of patients with rheumatic disease. Moreover, both instruments had only weak to moderate correlations with the SF-36, supporting the *a priori* hypothesis of divergent construct validity (i.e., the EQ-5D and HUI3 measure preference for or desirability of health status while the SF-36 measures self-perceived health status[25]). The 1-week test-retest reliability of these 2 instruments (ICC: 0.64 for EQ-5D; 0.75 for HUI3) was better than or close to the recommended level of 0.7 for group comparisons[37]. These results are comparable with those from previous studies of EQ-5D[8,9] and HUI3[41]. It was not

Table 3. Known-groups construct validity of the EQ-5D and HUI3.

Grouping Variable	n	Mean ± SD (median) EQ-5D	Mean ± SD (median) HUI3	Difference‡
SF-36 scores †				
Physical functioning				
≥ 65	60	0.81 ± 0.15 (0.79)	0.79 ± 0.17 (0.85)	0.02
< 65	54	0.69 ± 0.24 (0.73)**	0.72 ± 0.16 (0.76)*	–0.03
Role-physical				
≥ 75	62	0.79 ± 0.18 (0.79)	0.80 ± 0.14 (0.85)	–0.01
< 75	52	0.70 ± 0.23 (0.73)*	0.71 ± 0.19 (0.76)**	–0.01
Bodily pain				
≥ 61	58	0.83 ± 0.11 (0.79)	0.83 ± 0.12 (0.85)	0
< 61	56	0.67 ± 0.25 (0.73)***	0.69 ± 0.18 (0.72)***	–0.02
General health				
≥ 57	63	0.82 ± 0.12 (0.79)	0.81 ± 0.14 (0.85)	0.01
< 57	51	0.67 ± 0.26 (0.73)***	0.70 ± 0.19 (0.76)**	–0.03
Vitality				
≥ 50	74	0.79 ± 0.16 (0.79)	0.79 ± 0.14 (0.85)	0
< 50	40	0.68 ± 0.26 (0.73)**	0.69 ± 0.19 (0.72)**	–0.01
Social functioning				
≥ 75	66	0.81 ± 0.15 (0.79)	0.79 ± 0.15 (0.84)	0.02
< 75	48	0.68 ± 0.26 (0.73)**	0.71 ± 0.18 (0.73)*	–0.03
Role-emotional				
100	73	0.79 ± 0.17 (0.79)	0.80 ± 0.15 (0.85)	–0.01
< 100	41	0.69 ± 0.26 (0.73)*	0.69 ± 0.19 (0.72)***	0
Mental health				
≥ 72	62	0.79 ± 0.15 (0.79)	0.81 ± 0.13 (0.85)	–0.02
< 72	52	0.69 ± 0.26 (0.73)*	0.70 ± 0.19 (0.73)**	–0.01
Pain VAS (cm)†				
< 3.9	57	0.82 ± 0.11 (0.79)	0.83 ± 0.13 (0.85)	–0.01
≥ 3.9	57	0.69 ± 0.26 (0.73)**	0.69 ± 0.18 (0.72)***	0
Tender points present				
No	39	0.79 ± 0.17 (0.79)	0.81 ± 0.15 (0.85)	–0.02
Yes	73	0.73 ± 0.22 (0.76)	0.73 ± 0.18 (0.77)*	0
Chronic medical condition present				
No	59	0.77 ± 0.17 (0.79)	0.76 ± 0.16 (0.80)	0.01
Yes	55	0.74 ± 0.24 (0.79)	0.76 ± 0.18 (0.79)	–0.02
Acute medical condition present				
No	28	0.79 ± 0.23 (0.79)	0.77 ± 0.19 (0.85)	0.02
Yes	86	0.74 ± 0.19 (0.79)	0.76 ± 0.16 (0.79)	–0.02

Statistical significance derived from parametric and non-parametric tests for each pair of known-groups were very similar, therefore only p values for parametric 2 sample t tests are reported. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$; † Cut-off values are median scores; ‡ Difference = (EQ-5D score – mean HUI3 score), $p > 0.05$ for all (paired t tests).

surprising that the observed test-retest reliability of the EQ-5D was lower than that for the HUI3, as the recall period for the EQ-5D was the day of the survey (i.e., 1 day), while that for the version of the HUI3 used in this study was 4 weeks. The use of a 1-week test-retest period in this study thus resulted in no overlap in recall period for the EQ-5D but did lead to an overlapping recall period for the HUI3. Therefore, changes in HRQoL occurring over the 1 week period would have a greater influence on EQ-5D than on HUI3 scores, resulting in comparatively lower test-retest reliability for the EQ-5D.

Although the EQ-5D and HUI3 showed similar mean/median scores and psychometric properties, differences were observed in EQ-5D and HUI3 scores for individual patients, as suggested by the moderate correlation between utility scores. The magnitude of correlation in this study was similar to that in patients treated for intermittent claudication (Spearman's ρ: 0.54)[24] but lower than that in patients with a variety of medical conditions (Spearman's ρ: 0.65)[39] and in a community based study (Spearman's ρ: 0.69)[23]. Also, agreement of responses to corresponding dimensions of the 2 instruments were moderate (Spearman's ρ: 0.33-0.52), and patients reporting no problems for a given dimension with one instrument reported significant problems in the same dimension for the other instrument (as illustrated for the EQ-5D mobility and HUI3 ambulation items in Results). These findings are similar to those in

Table 4. Responses to corresponding EQ-5D and HUI3 dimensions for baseline interviews: cross-tabulation and correlations.

EQ-5D Dimension	HUI3 Dimension						
	1	2	3	4	5	6	ρ*
				HUI3: ambulation			
EQ-5D mobility							0.35
No problems	78	7	4	1	0	0	
Some problems	12	9	3	0	0	0	
Extreme problems	0	0	0	0	0	0	
				HUI3: pain			
EQ-5D: pain/discomfort							0.52
No pain/discomfort	10	14	0	1	0	—	
Moderate pain/discomfort	2	43	34	4	0	—	
Extreme pain/discomfort	0	1	2	3	0	—	
				HUI3: emotion			
EQ-5D: anxiety/depression							0.33
No anxiety/depression	44	24	2	0	0	—	
Moderate anxiety/depression	14	20	7	0	0	—	
Extreme anxiety/depression	1	0	1	1	0	—	

* Spearman's correlation coefficients, $p < 0.001$ for all.

patients in a community-based survey comparing these 2 instruments, which found correlations between corresponding items for these 2 instruments of 0.50 to 0.61[23]. Also in that study, 20 of 55 respondents (36%) with bronchitis and emphysema reported problems with the EQ-5D mobility dimension but only 6 respondents (11%) reported difficulties with the HUI3 ambulation dimension. There are several possible reasons for the differences in EQ-5D and HUI3 utility scores for individual patients seen in this study. First, our data and data in the literature suggest that the EQ-5D and HUI3 measure slightly different, though related, dimensions of health. The EQ-5D defines health in terms of functioning, while the HUI3 does so in terms of ability. For example, the EQ-5D mobility item inquires about problems in walking about while the HUI3 ambulation item inquires about use of walking equipment and help from another person. Thus, a discordant response may occur if a patient feels no problem in walking about even though he or she has to use a walking aid.

Second, scoring methods of the 2 instruments were developed using different methodologies. The EQ-5D used a time trade-off (TTO) technique for eliciting health utility and regression analysis for data modeling, while the HUI3 used rating scale (RS) and standard gamble (SG) techniques for utility elicitation and multi-attribute utility theory[31] for data modeling. Systematic differences have been shown in utility scores elicited using RS, TTO, or SG[42-45]. Third, utility scores were calculated using scoring functions derived from different populations, namely the UK for the EQ-5D and Canada for the HUI3. Fourth, differences in scaling (with EQ-5D scores having a lower bound than HUI3 scores) could have resulted in these differences.

Given the above, it appears that both the EQ-5D and HUI3 are suitable utility-based HRQoL instruments for use in patients with rheumatic disease. The decision regarding which instrument to use in an individual study would depend on several factors (which differ between these instruments), including the preferred approach to define health status (functioning/ability), HRQoL dimensions and time frame of interest, overall respondent burden, and availability of a utility function for the country/language of interest.

Limitations of this study include the assessment of outpatients seen at a tertiary referral hospital. As such, our results may not be generalizable to severely disabled patients with rheumatic disease. Scoring functions for utility scores were derived from 2 different populations, as population-based scoring norms are not available for the Singapore population. Therefore, the observed difference in utility scores in this study may partially reflect differences in utility between the 2 populations (previously reported for the EQ-5D[46]). To address this concern, we hope to develop a Singaporean utility value set for the EQ-5D and HUI3. Another possible limitation is the use of both English and Chinese versions of these instruments, as subtle differences in semantics between the 2 language versions might introduce additional variability. However, previous research has not shown substantial disparity in psychometric properties[28,29] or mean scores[30] for these versions of the EQ-5D, and analyses of EQ-5D and HUI3 utility scores in patients completing either language version yielded similar results (not shown) as the pooled analysis.

In conclusion, we found that the EQ-5D and HUI3 performed equally well in measuring utility-based HRQoL in patients with rheumatic disease, although they measured slightly different, though related, dimensions of health.

ACKNOWLEDGMENTS

The authors would like to thank staff nurses from the Nanyang Polytechnic Advanced Diploma in Nursing Course for their help in interviewing patients and the EuroQol Group and Health Utilities Inc. for allowing us to use the EQ-5D and HUI3, respectively.

REFERENCES

1. Gold MR, Patrick DL, Torrance GW, et al. Identifying and valuing outcomes. In: Gold MR, Siegel JE, Russell LB, Weinstein MC, editors. Cost-Effectiveness in health and medicine. New York: Oxford University Press; 1996:82-134.
2. Drummond MF, O'Brien BJ, Stoddart GL, Torrance GW. Cost-utility analysis. In: Methods for the economic evaluation of health care programmes. 2nd ed. Oxford: Oxford University Press; 1997:139-99.
3. Brooks R, with the EuroQol Group. EuroQol: the current state of play. Health Policy 1996;37:53-72.
4. Rabin R, de Charro F. EQ-5D: a measure of health status from the EuroQol Group. Ann Med 2001;33:337-43.
5. Feeny D, Furlong W, Boyle M, Torrance GW. Multi-attribute health status classification systems. Health utilities index. Pharmacoeconomics 1995;7:490-502.
6. Furlong WJ, Feeny DH, Torrance GW, Barr RD. The health utilities index (HUI) system for assessing health-related quality of life in clinical studies. Ann Med 2001;33:375-84.

7. 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.
8. 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.
9. Fransen M, Edmonds J. Reliability and validity of the EuroQol in patients with osteoarthritis of the knee. Rheumatology 1999;38:807-13.
10. Brazier JE, Harper R, Munro J, Walters SJ, Snaith ML. Generic and condition-specific outcome measures for people with osteoarthritis of the knee. Rheumatology 1999;38:870-7.
11. Wolfe F, Hawley DJ. Measurement of the quality of life in rheumatic disorders using the EuroQol. Br J Rheumatol 1997;36:786-93.
12. Kobelt G, Eberhardt K, Jonsson L, Jonsson B. Economic consequences of the progression of rheumatoid arthritis in Sweden. Arthritis Rheum 1999;42:347-56.
13. Wolfe F, Kong SX, Watson DJ. Gastrointestinal symptoms and health related quality of life in patients with arthritis. J Rheumatol 2000;27:1373-8.
14. Wang C, Mayo NE, Fortin PR. The relationship between health related quality of life and disease activity and damage in systemic lupus erythematosus. J Rheumatol 2001;28:525-32.
15. Suarez-Almazor ME, Conner-Spady B. Rating of arthritis health states by patients, physicians, and the general public. Implications for cost-utility analyses. J Rheumatol 2001;28:648-56.
16. Sokoll KB, Helliwell PS. Comparison of disability and quality of life in rheumatoid and psoriatic arthritis. J Rheumatol 2001;28:1842-6.
17. Pipitone N, Scott DL. Magnetic pulse treatment for knee osteoarthritis: a randomised, double-blind, placebo-controlled study. Curr Med Res Opin 2001;17:190-6.
18. Moore AD, Clarke AE, Danoff DS, et al. Can health utility measures be used in lupus research? A comparative validation and reliability study of 4 utility indices. J Rheumatol 1999;26:1285-90.
19. Grootendorst P, Feeny D, Furlong W. Health utilities index mark 3: evidence of construct validity for stroke and arthritis in a population health survey. Med Care 2000;38:290-9.
20. Raynauld JP, Torrance GW, Band PA, et al. A prospective, randomized, pragmatic, health outcomes trial evaluating the incorporation of hylan G-F 20 into the treatment paradigm for patients with knee osteoarthritis (Part 1 of 2): clinical results. Osteoarthritis Cartilage 2002;10:506-17.
21. Torrance GW, Raynauld JP, Walker V, et al. A prospective, randomized, pragmatic, health outcomes trial evaluating the incorporation of hylan G-F 20 into the treatment paradigm for patients with knee osteoarthritis (Part 2 of 2): economic results. Osteoarthritis Cartilage 2002;10:518-27.
22. Mittmann N, Trakas K, Risebrough N, Liu BA. Utility scores for chronic conditions in a community-dwelling population. Pharmacoeconomics 1999;15:369-76.
23. Statistics Canada. A head-to-head comparison of two generic health status measures in the household population: McMaster health utilities index (mark 3) and the EQ-5D (internal document). Ottawa: Statistics Canada; 2000.
24. Bosch JL, Hunink MG. Comparison of the health utilities index mark 3 (HUI3) and the EuroQol EQ-5D in patients treated for intermittent claudication. Qual Life Res 2000;9:591-601.
25. Ware JE, Snow KK, Kosinski M, Gandek B. SF-36 health survey manual and interpretation guide. Boston: The Health Institute, New England Medical Center; 1993.
26. Wolfe F, Smythe HA, Yunus MB, et al. The American College of Rheumatology 1990 criteria for the classification of fibromyalgia: report of the multicenter criteria committee. Arthritis Rheum 1990;33:160-72.
27. Dolan P. Modeling valuations for EuroQol health states. Med Care 1997;35:1095-108.
28. Luo N, Chew LH, Fong KY, et al. Validity and reliability of the EQ-5D self-reported questionnaire in English-speaking Asian patients with rheumatic diseases in Singapore. Qual Life Res 2003;12:87-92.
29. Luo N, Chew LH, Fong KY, et al. Validity and reliability of the EQ-5D self-reported questionnaire in Chinese-speaking patients with rheumatic diseases in Singapore. Ann Acad Med Singapore (in press).
30. Luo N, Chew LH, Fong KY, et al. Are scores of Singaporean English and Chinese EQ-5D versions equivalent? An exploratory study [abstract]. Qual Life Res 2002;11:647.
31. Keeney RL, Raiffa H. Decisions with multiple objectives: preference and value tradeoffs: Second Edition. New York: Cambridge University Press; 1993.
32. Feeny D, Furlong W, Torrance GW, et al. Multi-attribute and single-attribute utility functions for the health utilities index mark 3 system. Med Care 2002;40:113-28.
33. Wang Q, Chen G. The health status of the Singaporean population as measured by a multi-attribute health status system. Singapore Med J 1999;40:389-96.
34. Wagner AK, Gandek B, Aaronson NK, et al. Cross cultural comparisons of the content of SF-36 translations across 10 countries: results from the international quality of life assessment project. J Clin Epidemiol 1998;51:925-32.
35. Lam CL, Gandek B, Ren XS, Chan MS. Tests of scaling assumptions and construct validity of the Chinese (HK) version of the SF-36 Health Survey. J Clin Epidemiol 1998;51:1139-47.
36. Thumboo J, Fong KY, Machin D, et al. A community-based study of scaling assumptions and construct validity of the English (UK) and Chinese (HK) SF-36 in Singapore. Qual Life Res 2001;10:175-88.
37. Fayers PM, Machin D. Quality of life: Assessment, analysis and interpretation. Chichester: John Wiley & Sons; 2000.
38. Brazier J, Jones N, Kind P. Testing the validity of the Euroqol and comparing it with the SF-36 health survey questionnaire. Qual Life Res 1993;2:169-80.
39. Hawthorne G, Richardson J, Day NA. A comparison of the assessment of quality of life (AQoL) with four other generic utility instruments. Ann Med 2001;33:358-70.
40. Revicki DA, Kaplan RM. Relationship between psychometric and utility-based approaches to the measurement of health-related quality of life. Qual Life Res 1993;2:477-87.
41. Boyle MH, Furlong W, Feeny D, Torrance GW, Hatcher J. Reliability of the health utilities index—mark III used in the 1991 cycle 6 Canadian General Social Survey Health Questionnaire. Qual Life Res 1995;4:249-57.
42. Read JL, Quinn RJ, Berwick DM, Fineberg HV, Weinstein MC. Preferences for health outcomes. Comparison of assessment methods. Med Decis Making 1984;4:315-29.
43. Bass EB, Steinberg EP, Pitt HA, et al. Comparison of the rating scale and the standard gamble in measuring patient preferences for outcomes of gallstone disease. Med Decis Making 1994;14:307-14.
44. Martin AJ, Glasziou PP, Simes RJ, Lumley T. A comparison of standard gamble, time trade-off, and adjusted time trade-off scores. Int J Technol Assess Health Care 2000;16:137-47.
45. Morimoto T, Fukui T. Utilities measured by rating scale, time trade-off, and standard gamble: review and reference for health care professionals. J Epidemiol 2002;12:160-78.
46. Tsuchiya A, Ikeda S, Ikegami N, et al. Estimating an EQ-5D population value set: the case of Japan. Health Econ 2002;11:341-53.