

# Validating and Assessing the Sensitivity of the Health Assessment Questionnaire-Disability Index-derived Short Form-6D in Patients with Early Aggressive Rheumatoid Arthritis

SOGOL S. AMJADI, PAUL M. MARANIAN, HAROLD E. PAULUS, ROBERT M. KAPLAN, VEENA K. RANGANATH, DANIEL E. FURST, PUJA P. KHANNA, and DINESH KHANNA, for the Western Consortium of Practicing Rheumatologists

**ABSTRACT.** *Objective.* New methodologies allow the scores for the Health Assessment Questionnaire-Disability Index (HAQ-DI) to be translated into preferences/utility scores. We evaluated the construct validity of the HAQ-DI-derived Short Form-6D (SF-6D) score and assessed its responsiveness to change over 6- and 12-month followup periods in patients with early aggressive rheumatoid arthritis (RA). *Methods.* Patients (n = 277) participating in an RA observational study completed self-reported measures of symptoms and the HAQ-DI at baseline and at 6 and 12 months. Total Sharp scores, C-reactive protein, and erythrocyte sedimentation rate were assessed along with clinical data. Construct validity was assessed by examining the association between SF-6D score and patient-reported and clinical measures using Spearman correlation coefficients. The responsiveness of SF-6D to change was assessed using patient and physician assessments of the disease as clinical anchors. The magnitude of responsiveness was calculated using SF-6D effect size (ES). *Result.* Mean SF-6D scores were 0.690, 0.720, and 0.723 at baseline and 6 and 12-month followup, respectively. Baseline patient-reported measures had moderate to high correlations with baseline SF-6D ( $r = 0.43$  to  $0.52$ ); whereas clinical measures had negligible to low correlations with SF-6D ( $r = 0.001$  to  $0.32$ ). ES was moderate for the groups that were deemed to have improved (ES  $0.63$ – $0.75$ ) but negligible to small for those that did not (ES  $0.13$ – $0.46$ ). *Conclusion.* Our data support the validity and responsiveness of the HAQ-DI derived SF-6D score in an early RA cohort. These results support the use of the HAQ-DI derived SF-6D in RA cohorts and clinical trials lacking preference-based measures. (J Rheumatol First Release April 15 2009; doi:10.3899/jrheum.080959)

## Key Indexing Terms:

SHORT FORM-6D

HEALTH UTILITY

EARLY RHEUMATOID ARTHRITIS

HEALTH ASSESSMENT QUESTIONNAIRE-DISABILITY INDEX

PREFERENCE BASED MEASURES

COST-EFFECTIVENESS ANALYSIS

Rheumatoid arthritis (RA) is a chronic disorder that primarily involves the joints. As in many other chronic diseases, RA and/or its treatment may have detrimental effects on

health-related quality of life (HRQOL). In general, there are 2 ways to assess HRQOL. These methods include health status and health utility (preference based) assessments<sup>1,2</sup>. Health status measures describe a person's ability to function in one or more domains (e.g., physical functioning and/or mental well-being). Currently, one of the most commonly used disease-specific health status instruments in RA is the Health Assessment Questionnaire-Disability Index (HAQ-DI)<sup>3</sup>. It measures health status by assessing the patient's ability to function physically, and includes questions that involve the function of both upper and lower extremities. HAQ-DI scores are associated with work productivity, disability, and mortality<sup>4,5</sup> in RA.

Preference based measures assess the value or desirability of a state of health against an external metric. They allow direct comparison of health status by integrating multiple pieces of information into a single summary number scaled

---

From the Division of Rheumatology, University of California Los Angeles (UCLA) School of Medicine; and Department of Health Services, UCLA School of Public Health, Los Angeles, California, USA.

P.P. Khanna was supported by a National Institutes of Health Award (T32 AR 053463). D. Khanna was supported by a National Institutes of Health Award (NIAMS K23 AR053858-01A1).

S.S. Amjadi, MS; P.M. Maranian, MS; H.E. Paulus, MD; V.K. Ranganath, MD; D. Furst, MD; P.P. Khanna, MD, MPH; D. Khanna, MD, MS, Division of Rheumatology, UCLA School of Medicine; R.M. Kaplan, PhD, Department of Health Services, UCLA School of Public Health.

Address reprint requests to Dr. D. Khanna, Division of Rheumatology, Department of Medicine, David Geffen School of Medicine, 1000 Veteran Avenue, Room 32-59, Rehabilitation Center, Los Angeles, CA 90095.

E-mail: dkhanna@mednet.ucla.edu

Accepted for publication December 30, 2008.

between 2 anchor states, usually “dead” (0.0) and “perfect health” (1.0)<sup>6</sup>. Preference based measures are used as weights in calculating quality-adjusted life-years (QALY). QALY take into account both quantity and quality of life (QOL) in a single metric, calculated as the arithmetic product of life expectancy and the QOL of the remaining life-years. A year of perfect health is worth 1.0 QALY, a year of life in less than perfect health is worth less than 1.0 QALY, and being dead is worth 0.0 QALY. At a policy level, QALY are incorporated into decision and cost-effectiveness (cost-utility) analyses of healthcare interventions<sup>6</sup>. Preference based measures are obtained either directly (via face-to-face interview with patients) or indirectly. Direct health utilities are usually ascertained via face-to-face interviews, with computer-assisted administration being the state of the art. The most common health utility measures are the standard gamble (SG), time tradeoff (TTO), and rating scale (RS)<sup>6</sup>. Indirect health utilities such as EuroQol use population-assigned weights to calculate utility scores for particular health states from health status instruments. The ease of administration (self-administered) of these indirect measures enables them to be used in national surveys, and as the source of QOL weightings in economic evaluations.

The Short Form-6D (SF-6D)<sup>7</sup> is an indirect preference based measure that is derived from responses on the Medical Outcomes Short Form-36 (SF-36), a widely used generic health status instrument<sup>8</sup>. Brazier, *et al*<sup>7</sup> developed the SF-6D, which is based on 11 SF-36 items, by asking a UK general population to report preferences for a sample of the SF-6D health states using a standard gamble technique. Although the SF-36 has 8 domains, the SF-6D has reduced this to 6 domains (physical function, role limitation, social function, pain, mental health, and vitality). Based on econometric modeling of the observed preferences, they constructed a model for estimating mean preferences for all possible SF-6D health states. The scoring algorithm produces scores ranging from 0.29 to 1.00. Although clinical trials in RA often incorporate the SF-36, which can be used to calculate the SF-6D<sup>9</sup>, they are usually limited to one or 2 years and do not represent the general RA population. On the other hand, observational studies in RA provide a unique perspective for assessing longterm outcomes such as joint replacement, cardiovascular morbidity, and mortality associated with RA<sup>10</sup>. To assess the cost-effectiveness of interventions these problems necessitate (e.g., joint replacement, treatment of cardiovascular disease, etc.), one needs a preference based measure to assess QALY. However, many longterm observational studies in RA have not included any preference based measures. As stated by Bansback, *et al*, “Because new programs and treatments in RA are competing alongside other disease areas for funding, it is important for the rheumatology community to be able to demonstrate the value of their interventions to policy makers” [page 964]<sup>11</sup>. Consequently, it is useful to find ways to convert the

more traditional RA-related health status instruments (e.g., HAQ-DI) to preference based measures (the SF-6D).

Bansback, *et al*<sup>11</sup> recently developed several linear regression models to map the SF-6D from the HAQ-DI in patients with RA from the UK and Canada. In this study, we employed the model developed by Bansback, *et al* to estimate SF-6D scores from the HAQ-DI in US subjects participating in an early RA observational cohort. The aims of our study were to evaluate convergent and divergent evidence for the construct validity of the HAQ-DI derived SF-6D score. In addition, we assessed the responsiveness of the HAQ-DI derived SF-6D to changes in other patient reported measures such as patient global assessment over 6 and 12-month followup periods.

## MATERIALS AND METHODS

**Patients and data.** Patients included in this study are part of a longterm observational study involving the Western Consortium of Practicing Rheumatologists (CPR), a regional consortium of rheumatology practices in the western US and Mexico<sup>12,13</sup>. The consortium physicians participating in the study were mainly from community and university practices in California, Idaho, New Mexico, Oregon, Utah, Colorado, Washington, Wyoming, and Guadalajara, Mexico.

Since 1993, 323 patients have been enrolled into the study. Inclusion criteria for the CPR cohort included a diagnosis of early RA, no previous treatment with disease modifying antirheumatic drugs (DMARD), rheumatoid factor-seropositive (titer  $\geq 1:80$  or  $\geq 40$  IU), and  $\geq 6$  swollen joints and  $\geq 9$  tender joints. The consortium rheumatologists assessed patient disease status at study entry (baseline), 6 months, 1 year, and yearly thereafter. Using standard methods, detailed physician assessment included all of the core set outcomes measures required to calculate the Disease Activity Score (DAS), including 28 tender and swollen joint counts and acute-phase reactant measures, as well as 0–100 mm visual analog scales (VAS) for global, pain, fatigue, and arthritis severity assessments. In addition, study visits included radiographs of the hands, wrists, and forefeet, and the total Sharp score was calculated<sup>14</sup>. At each scheduled physician visit, blood specimens were collected for C-reactive protein (CRP); erythrocyte sedimentation rate (ESR) was determined when clinically indicated, in rheumatologist’s office or local laboratory.

Patients were also asked to complete a detailed questionnaire at study entry and every 6 months thereafter for the duration of the study. The questionnaires evaluated changes in demographics, health, medication, pain and global VAS, the HAQ-DI, and the Center for Epidemiological Studies-Depression scale (CES-D).

**Patient reported measures.** The HAQ-DI is a 20-item arthritis-targeted measure assessing upper and lower extremity functioning<sup>15</sup>. The HAQ-DI score is computed by summing the highest item score in each of the 8 domains and dividing the sum by 8, yielding a score from 0 (no disability) to 3 (severe disability). The original HAQ-DI includes an additional grade of difficulty for patients using assistive/adaptive devices such as a cane or a walker.

In addition to completing the HAQ-DI, patients completed 4 VAS as part of their patient questionnaires: patient global assessment of their arthritis (PGA), overall pain, overall fatigue, and overall arthritis severity; patients were asked to indicate by placing a vertical mark on the line how fatigue, pain, or arthritis interfered with their lives “during the past week.” Their rheumatologists also completed a physician global assessment. All scales ranged from 0 to 100 mm, where 0 indicated no symptoms and 100 very severe symptoms.

**Predicting mean SF-6D using HAQ-DI.** The SF-6D<sup>7</sup> derives preference based scores from the SF-36 by using population based utilities for SF-36

health states. The SF-6D revises the SF-36 into a 6-dimensional health state classification system: physical function, role limitations, social function, pain, mental health, and vitality; the general health scale items are not incorporated and 2 scales measuring role limitations due to physical and emotional problems are collapsed into a "role limitations" dimension. An SF-6D health state is defined by selecting one level from each dimension. A total of 18,000 health states are thus defined. The SF-6D is scored from 0.29 to 1.00, where 0.29 represents worst possible health and 1.00 is perfect health<sup>7</sup>.

Bansback, *et al* developed several linear regression models to estimate the relationship between the HAQ-DI and SF-6D<sup>11</sup>. They used 2 models to predict the SF-6D from HAQ-DI. Model 1 used the 8 HAQ-DI domain scores and treated them as continuous variables. In Model 2, the HAQ-DI domains were treated as ordinal variables. Both models displayed acceptable and very similar statistical fit. However, Model 1, by treating each domain score as a continuous variable, assumes the intervals between response levels are the same, which may not be completely valid. On the other hand, Model 2, by treating each level of the domain score as an ordinal variable, does not make this assumption and therefore is less restrictive<sup>11</sup>. In our study we obtained similar results using both models. Predicted SF-6D under Model 1 and Model 2 were, respectively, 0.675 and 0.690 at baseline, 0.718 and 0.720 at 6 months, and 0.722 and 0.723 at 12 months. Since Model 2 conforms better to an ordinal HAQ-DI scale, we calculated the results using Model 2 as our prediction model.

**Statistical analysis.** Descriptive statistics for continuous variables are presented as means and standard deviations, and for categorical variables as proportions.

**Construct validity.** We examined the association between baseline SF-6D and other baseline patient reported and clinical measures using Spearman correlation coefficients. We also assessed the association between change in SF-6D and change in the other patient reported and clinical measures from baseline to 6 months and from baseline to 12 months. A correlation from 0.00 to 0.20 was interpreted as no correlation; 0.21–0.40 as low correlation; 0.41–0.60 as moderate correlation; 0.61–0.80 as marked correlation; and 0.81–1.00 as high correlation<sup>16</sup>. Based on previous reports that showed moderate to high correlation between HRQOL and other patient reported measures, and low to negligible correlations between HRQOL and clinical measures<sup>17,18</sup>, we hypothesized that SF-6D scores would have at least moderate correlation ( $r > 0.40$ ) with PGA, pain VAS, and fatigue VAS, and low to negligible correlation ( $r < 0.40$ ) with disease severity, physician global assessment, ESR, CRP, and Sharp score.

The ability of baseline SF-6D to discriminate baseline PGA, pain VAS, fatigue VAS, arthritis severity VAS, and physician global assessment was assessed by classifying each of the VAS into 3 categories: mild (0.0–33.0), moderate (33.1–66.0), and severe (66.1–100.0)<sup>19</sup>. Differences among mild, moderate, and severe categories were evaluated for each variable using one-way ANOVA.

**Responsiveness in change.** We used PGA, patient reported pain, fatigue and disease severity VAS, and the physician global assessment VAS as clinical anchors to assess the responsiveness to change<sup>20</sup>. We divided our group into 2 categories: patients with improvement from baseline to 6 months and patients with no improvement from baseline to 6 months (the same was done at Month 12) based on clinical anchors. Improvement was defined as a decrease in the VAS scores by  $\geq 10$  mm from baseline to 6-month followup and from baseline to 12-month followup. A cutoff of 10 mm on a 0–100 mm scale was based on previous studies<sup>21–23</sup>, where a change of 10 mm on a 100 mm scale is consistent with minimally important difference<sup>20</sup>. In order to assess the responsiveness to change of SF-6D at 6 months, we estimated the SF-6D effect size (ES) by taking the change in mean SF-6D from baseline to 6 months and dividing the result by the standard deviation at baseline (SD = 0.06). The same was done to calculate the ES of SF-6D at 12 months. According to Cohen's rule, an ES of 0.20–0.49 represents a small change, 0.50–0.79 a medium change, and  $\geq 0.80$  a large change<sup>24</sup>.

In order to calculate quality-adjusted life-years (QALY) we plotted the mean SF-6D at baseline, 6 months, and 12 months. The mean QALY was

calculated by estimating the area under the path for each individual patient who had data available at baseline and 6 and 12-month visits ( $n = 177$ ). The area under the path is equal to the sum of the areas under consecutive SF-6D measurements, and the area under SF-6D measurements is obtained by multiplying the duration of the SF-6D in months by the average score of SF-6D. We used the following formula<sup>25</sup> to assess average QALY:  $[(0.5 \times (\text{SF-6D at baseline} + \text{SF-6D at 6 months}) \times 6) + (0.5 \times (\text{SF-6D at 6 months} + \text{SF-6D at 12 months}) \times 6)]/12$ . We assumed that the SF-6D changes between measurements at baseline, 6 months, and 12 months were smooth and gradual.

We also assessed the proportion of subjects with floor and ceiling effects (percentages of respondents scoring at the lowest and highest possible scale level). Computations were achieved using SAS System Release 8.2 software (SAS Institute, Cary, NC, USA).

## RESULTS

Two hundred seventy-seven patients had data available to compute the HAQ-DI and the SF-6D at baseline and formed the study sample. The subjects were mainly Caucasian (79.4%) and female (76.9%), with a mean  $\pm$  standard deviation (SD) age of  $51.1 \pm 13.2$  years and mean disease duration of  $8.6 \pm 10.2$  months; 85% of patients had disease duration  $\leq 12$  months. The PGA, physician global assessment (0–100 mm VAS), and DAS were  $42.2 \pm 23.5$ ,  $49.3 \pm 21.5$ , and  $6.0 \pm 1.1$ , respectively, representing moderate–severe disease (Table 1). There were no floor and ceiling effects observed for SF-6D score at baseline.

The HAQ-DI scores were  $1.18 \pm 0.70$  at baseline,  $0.78 \pm 0.65$  at 6 months, and  $0.72 \pm 0.68$  at 12 months. Corresponding mean SF-6D scores were  $0.690 \pm 0.056$  ( $n = 277$ ),  $0.720 \pm 0.053$  ( $n = 206$ ), and  $0.723 \pm 0.057$  ( $n = 211$ ). The distributions of SF-6D scores at baseline, 6, and 12 months are shown in Figure 1. Because we recorded data every 6 months, we were able to calculate average QALY over a period of 12 months. The mean QALY during the first 12 months was  $0.72 \pm 0.05$  (Figure 2).

**Construct validity.** Table 2 reports the Spearman correlation between HAQ-DI derived SF-6D scores and several patient reported and clinical measures. As expected, clinical measures such as ESR and total Sharp score had low to negligible correlations with SF-6D ( $r = 0.001$  for Sharp score and  $r = -0.14$  for ESR). Among clinical measures, CRP had the highest correlation with SF-6D at baseline ( $r = -0.31$ ). Baseline patient reported measures such as PGA and CES-D had at least moderate correlations with SF-6D ( $r > 0.40$ ), with PGA having the highest correlation with SF-6D ( $r = -0.52$ ).

**Discriminant validity of the SF-6D.** SF-6D scores were able to discriminate between mild, moderate, and severe PGA, pain VAS, fatigue VAS, and arthritis severity VAS with F-test  $p$  values  $< 0.0001$  for the overall comparisons (Figure 3). In addition, the SF-6D scores were discriminative of mild and moderate, mild and severe, and moderate and severe scores for each of the VAS assessments ( $p < 0.01$ ), with the exception of moderate versus severe fatigue VAS scores ( $p = 0.06$ ).

Table 1. Baseline characteristics.

Variable	N	Mean $\pm$ SD	Median
Age, yrs	277	51.1 $\pm$ 13.2	51.2
Female, %	213	76.9	—
Ethnicity, %			
Caucasian	220	79.4	—
Latino (Mexican American, Hispanic)	36	13.0	—
Other	21	7.6	—
Disease duration, mo	277	8.6 $\pm$ 10.2	6.0
Center for Epidemiologic Studies-Depression Score (0–60)	230	15.5 $\pm$ 11.4	13.0
C-reactive protein, mg/dl	251	2.1 $\pm$ 8.0	1.4
Erythrocyte sedimentation rate, mm/h	226	40.8 $\pm$ 25.6	35.5
Visual analog scale (0–100)*			
Patient pain	201	58.9 $\pm$ 27.7	62.0
Patient global	244	42.2 $\pm$ 23.5	43.0
Patient fatigue	201	52.6 $\pm$ 24.3	55.0
Patient arthritis severity	201	59.9 $\pm$ 26.3	61.0
Physician global assessment	263	49.3 $\pm$ 21.5	51.0
Tender joint count (0–28)	259	12.9 $\pm$ 7.0	12.0
Swollen joint count (0–28)	259	12.2 $\pm$ 6.5	12.0
Total Sharp score	246	6.67 $\pm$ 9.47	3.63
Disease Activity Score (DAS-28)	192	6.0 $\pm$ 1.1	6.0
Health Assessment Questionnaire-Disability Index (HAQ-DI; 0–3.0) <sup>†</sup>	277	1.18 $\pm$ 0.70	1.25
HAQ-DI derived SF-6D score (0.29–1.0) <sup>††</sup>	277	0.690 $\pm$ 0.056	0.690

\* VAS of 0 represents no symptoms; 100 severe symptoms. <sup>†</sup> HAQ-DI scores of 0 and 3 represent no disability and severe disability, respectively. <sup>††</sup> SF-6D of 0.29 represents worst possible health; 1.0 represents perfect health.

*SF-6D responsiveness to change.* The magnitude of ES was larger for the group that was deemed to have improved (ES > 0.50); patients who improved had an ES of moderate magnitude compared to negligible to low magnitude for patients who did not improve. The largest SF-6D ES was observed for the change in PGA (ES = 0.75) and pain VAS (ES = 0.75) in patients who improved at 12 months (Table 3).

## DISCUSSION

Measuring HRQOL in patients with RA makes it possible to distinguish between the effectiveness of different therapies. The main use of preference measures is to guide decision-making<sup>1,26</sup>. For example, preference measures can serve as “quality-adjustment factors” for calculating QALY in decision and cost-effectiveness analyses<sup>2</sup>. QALY has the potential to influence public policy and resource allocations, as it is an effective way to compare therapeutic interventions within the disease, and even across illnesses. Due to lack of time and resources, few studies have administered preference based measures<sup>27</sup>. Consequently, linear regression models have been developed to estimate the preference based values using other HRQOL measures in other chronic diseases<sup>9,28,29</sup>.

Recently, Bansback, *et al*<sup>11</sup> developed models of the relationship between HAQ-DI and SF-6D using various regression analyses. Their results suggested that the models are helpful in utilizing existing valuation data by offering a

method for researchers who need preference scores, but have not used a preference based measure in their study.

We assessed the construct validity and responsiveness to change of HAQ-DI derived SF-6D scores in a population of patients with early RA. Our patients had a mean SF-6D of 0.69 at baseline, which is very similar to Bansback’s results (UK: 0.62, Canada: 0.68). The small difference between the SF-6D scores is explained by differences in the HAQ-DI scores. Our patients had a lower HAQ-DI than the UK subjects at baseline (1.18 vs 1.41), which resulted in a higher SF-6D (closer to perfect health). This may be due to early disease duration of our cohort; disease duration in the UK was not provided in that report.

SF-6D scores had both convergent and divergent construct validity (Table 2); SF-6D scores had moderate association with HRQOL measures and no to low association with clinical measures. Marra, *et al* utilized the data from the Canadian RA population used to develop HAQ-DI derived SF-6D in a different report<sup>30</sup>, and assessed construct validity of indirect preference based measures (including SF-6D) and RA related variables. Similar to our results, they found moderate to high correlations between baseline SF-6D and baseline patient reported outcomes (pain VAS and patient global VAS). However, they also found moderate correlations between baseline SF-6D score and baseline tender and swollen joint counts ( $r = 0.47$  to  $0.53$ ), whereas we found negligible to small correlations between SF-6D and clinical



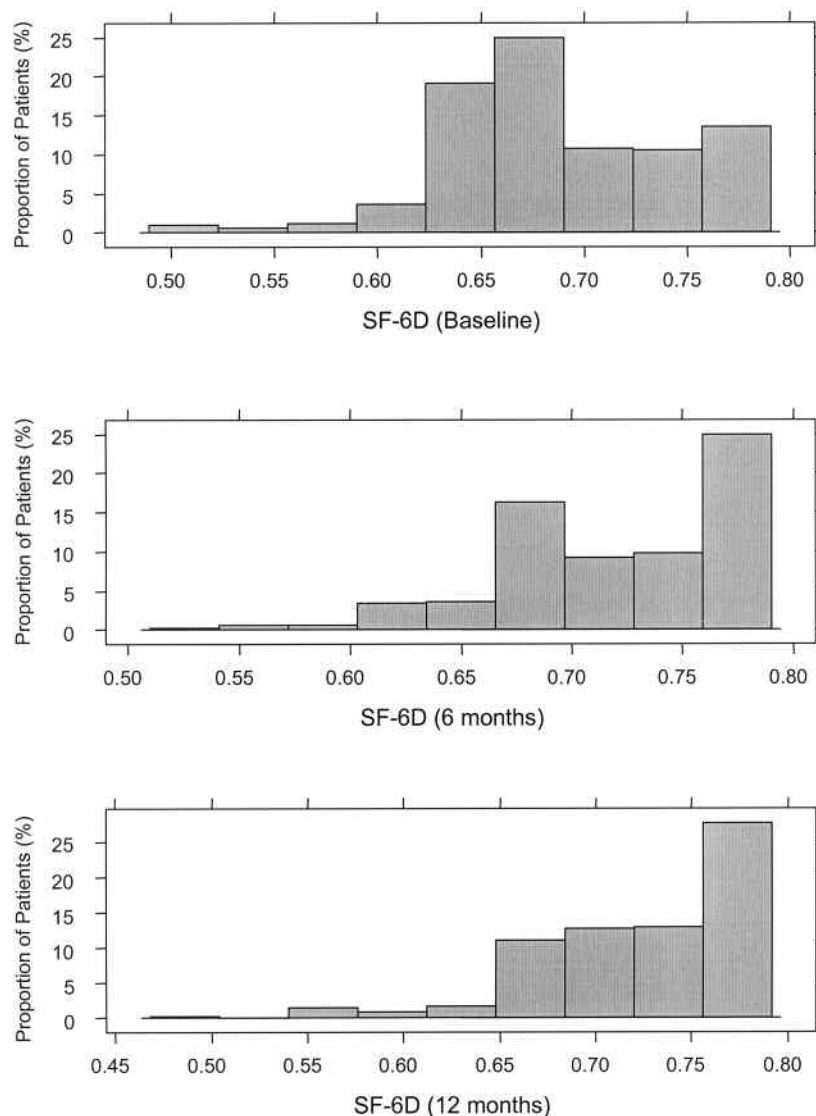


Figure 1. Histogram illustrating the distribution of SF-6D scores (0.29–1.00) at baseline and 6 and 12-month followup visits.

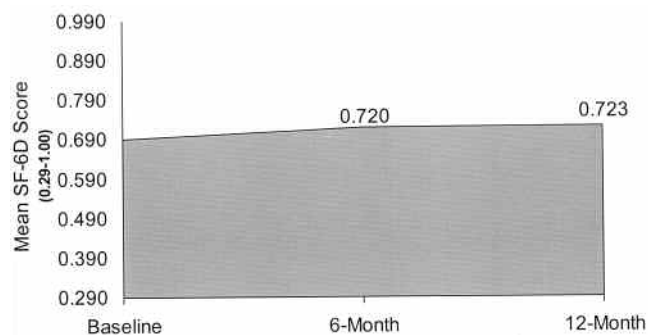


Figure 2. Mean QALY over 12 months: area under the curve represents the mean QALY over 12 months of followup.

measures such as tender and swollen joint counts, ESR, and radiographic damage. The differences may be related to the estimation of SF-6D; Marra, *et al*<sup>30</sup> estimated SF-6D directly from SF-36 using the formula from Brazier, *et al*<sup>7</sup>, whereas we derived SF-6D from the HAQ-DI using the model described by Bansback, *et al*<sup>11</sup>. In addition, we found lower correlations over time, compared to baseline. This is to be expected, as change scores inflate error variance, thereby attenuating the correlations<sup>31</sup>.

The SF-6D scores at baseline were able to discriminate between mild, moderate, and severe PGA, pain VAS, fatigue VAS, arthritis severity VAS, and physician global assessment VAS, with the exception of moderate versus severe

Table 2. Construct validity: correlation between SF-6D, disease related visual analog scales (VAS), and other clinical measures.

	Baseline		Change from Baseline to 6 mo		Change from Baseline to 12 mo	
	SF-6D	N	SF-6D	N	SF-6D	N
Disease duration, mo	0.17	277	—	—	—	—
(p)	(0.05)					
Patient global assessment (PGA)*,	-0.52	243	-0.35	181	-0.42	178
(p)	(< 0.0001)		(< 0.0001)		(< 0.0001)	
Pain VAS*	-0.50	201	-0.33	124	-0.40	125
(p)	(< 0.0001)		(0.0002)		(< 0.0001)	
Fatigue VAS*	-0.43	201	-0.19	120	-0.19	122
(p)	(< 0.0001)		(0.037)		(0.034)	
Arthritis severity VAS*	-0.50	201	-0.39	120	-0.28	124
(p)	(< 0.0001)		(< 0.001)		(0.001)	
CES-D*	-0.43	230	-0.28	162	-0.27	162
(p)	(< 0.0001)		(0.0004)		(0.0006)	
Physician global assessment,	-0.32	263	-0.25	157	-0.29	162
(p)	(< 0.0001)		(0.002)		(0.0001)	
Swollen joint count	-0.19	258	-0.23	169	-0.14	162
(p)	(0.003)		(0.003)		(0.074)	
Tender joint count	-0.19	258	-0.17	169	-0.17	162
(p)	(0.002)		(0.024)		(0.034)	
Total Sharp score	0.001	245	-0.03	137	-0.005	123
(p)	(0.985)		(0.707)		(0.958)	
DAS28	-0.30	192	-0.49	84	-0.17	74
(p)	(< 0.0001)		(< 0.0001)		(0.146)	
ESR	-0.14	226	-0.27	99	-0.13	101
(p)	(0.033)		(0.006)		(0.184)	
CRP	-0.31	250	-0.36	125	-0.36	129
(p)	(< 0.001)		(< 0.0001)		(< 0.0001)	

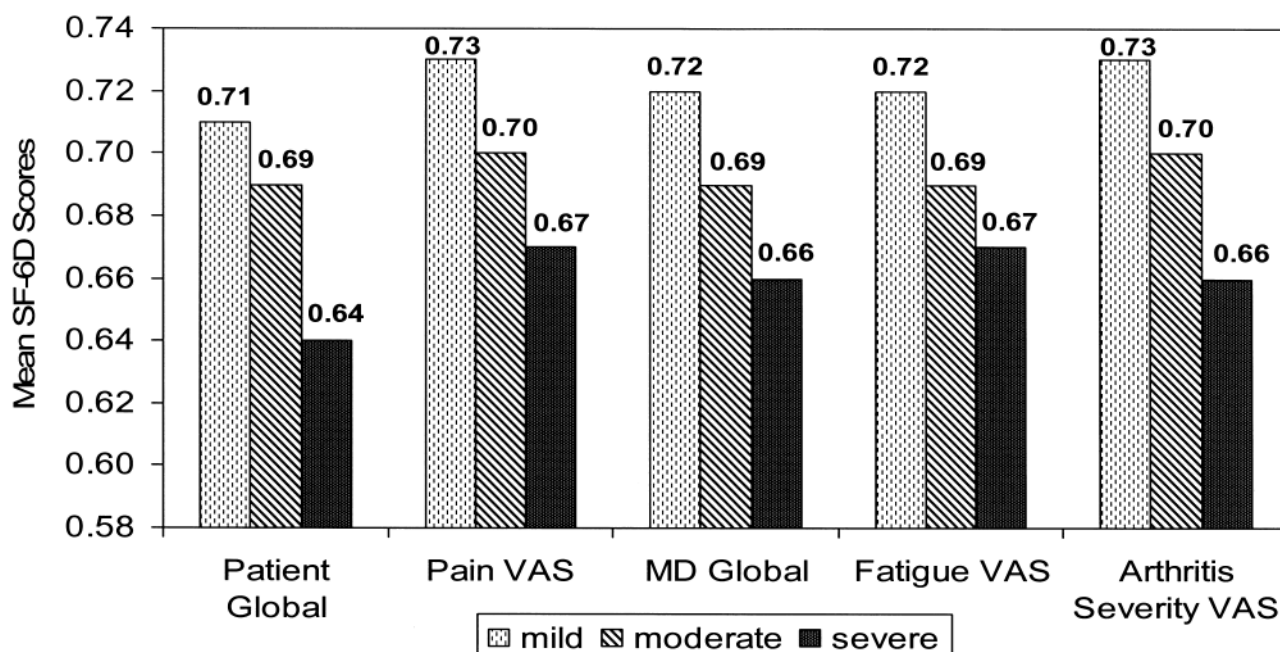
\* Reported by patients.

fatigue VAS ( $p = 0.059$ ). A similar finding was seen in another analysis that assessed SF-6D in scleroderma<sup>17</sup>; the SF-6D scores at baseline were able to differentiate between mild, moderate, and severe patient global assessment.

The ability of HRQOL instruments to detect clinically important changes is crucial to their usefulness in determining the effectiveness of different therapies<sup>32</sup>. The magnitude of responsiveness as measured by these instruments is useful in assessing treatment efficiency and in estimating sample size for future study designs<sup>33</sup>. In our study, SF-6D scores were able to discriminate between patients who were deemed to have improved and those who did not improve. SF-6D scores had a larger magnitude of ES for the improved group (both from baseline to 6 months and from baseline to 12 months, with ES ranging from 0.62 to 0.74) compared to the group with no improvement (ES ranging from 0.13 to 0.45). Overall, the SF-6D had the largest magnitude of change to PGA and pain VAS at 12 months. Previous studies found that the minimally important difference — the smallest difference in scores that patients perceive as beneficial<sup>34</sup> — in SF-6D for different arthritides ranges from 0.030 to 0.037<sup>17,30,35</sup>. In our study, the mean differences in SF-6D scores at 6 and 12 months were 0.030 and 0.033, respectively; these are minimally important differences and

are thus clinically meaningful. In addition, SF-6D scores increased over the 12-month period, suggesting that treatment of early RA in our cohort resulted in higher preferences by RA patients for their current health states.

Our study is not without limitations. First, our results are applicable for informing health policy decisions and not for individual preferences, as we used mean scores of the SF-6D. Second, Bansback's Canadian study population had moderate disease with a mean duration of  $13.98 \pm 11.64$  years at baseline, whereas our patients had more aggressive disease ( $\text{DAS } 6.0 \pm 1.1$ ), with 85% having baseline disease duration of  $\leq 12$  months. Thus, the validity of these models needs to be assessed in patients with milder RA and shorter disease duration. Another limitation is that Bansback's models are somewhat limited. The Bansback group created several translation models, but each model had an  $R^2$  for predicting SF-6D values of only about 0.50. Although an  $R^2$  of 0.50 is respectable, it still means that only about half of the variance in SF-6D scores can be known on the basis of HAQ-DI responses. Even though translations are attractive, investigators may still be better advised to select utility based outcome measures such as the EuroQol and Quality of Well Being scales when measuring outcomes for cost-effectiveness analyses<sup>36</sup>. Lastly, we assessed only the construct



Global Assessment	ANOVA		Linear contrasts (p-values)		
	F-statistic	p-value	mild v. moderate	mild v. severe	moderate v. severe
Patient Global	36.47	<.0001	<.0001	<.0001	<.0001
Pain VAS	23.32	<.0001	0.0018	<.0001	0.0003
Physician Global	19.32	<.0001	<.0001	<.0001	0.0069
Fatigue VAS	14.02	<.0001	0.0002	<.0001	0.0565
Arthritis Severity VAS	26.97	<.0001	0.0089	<.0001	<.0001

Figure 3. Discrimination of the SF-6D among mild, moderate and severe patient global assessment, pain, fatigue, arthritis severity, and physician global assessment on visual analog scales.

Table 3. Responsiveness of SF-6D to change in patient global, pain, fatigue, and arthritis severity and physician global visual analog scales (VAS: 0-100 mm).

Assessment	SF-6D Effect Size: 6 mo		SF-6D Effect Size: 12 mo	
	≥ 10 mm Decrease in VAS	< 10 mm Decrease in VAS	≥ 10 mm Decrease in VAS	< 10 mm Decrease in VAS
Patient global (n)	0.66 (202)	0.33 (75)	0.75 (199)	0.29 (78)
Pain VAS (n)	0.64 (230)	0.24 (47)	0.75 (225)	0.13 (52)
Fatigue VAS (n)	0.66 (219)	0.36 (58)	0.63 (213)	0.46 (64)
Arthritis severity VAS (n)	0.66 (228)	0.17 (49)	0.71 (233)	0.18 (44)
Physician global (n)	0.65 (213)	0.31 (64)	0.67 (224)	0.34 (53)

validity of the HAQ-DI derived SF-6D and did not assess the criterion validity, as SF-36 was not administered in this observational study. A more accurate measure of validity would be to assess the criterion validity, which requires the instrument in question (HAQ-DI derived SF-6D in this case) to correlate with an instrument that is considered the “gold standard” (observed SF-6D in this case).

Our results provide support for the validity of HAQ-DI derived SF-6D scores in patients with early RA over a period of 12 months. In addition, the results show that SF-6D scores are responsive to changes in HRQOL measures. Our study supports the use of the HAQ-DI derived SF-6D in RA observational cohorts where no preference based measure has been obtained.

## ACKNOWLEDGMENT

The Western Consortium of Practicing Rheumatologists: Robert Shapiro, Maria W. Greenwald, H. Walter Emori, Fredrica E. Smith, Craig W. Wiesenhutter, Charles Boniske, Max Lundberg, Anne MacGuire, Jeffrey Carlin, Robert Ettlinger, Michael H. Weisman, Elizabeth Tindall, Karen Kolba, George Krick, Melvin Britton, Rudy Greene, Ghislaine Bernard Medina, Raymond T. Mirise, Daniel E. Furst, Kenneth B. Wiesner, Robert F. Willkens, Kenneth Wilske, Karen Basin, Robert Gerber, Gerald Schoepflin, Marcia J. Sparling, George Young, Philip J. Mease, Ina Oppliger, Douglas Roberts, J. Javier Orozco Alcala, John Seaman, Martin Berry, Ken J. Bulpitt, Grant Cannon, Gregory Gardner, Allen Sawitzke, Andrew Lun Wong, Daniel O. Clegg, Timothy Spiegel, Wayne Jack Wallis, Mark Wener, Robert Fox.

## REFERENCES

1. Tsevat J, Weeks JC, Guadagnoli E, et al. Using health-related quality-of-life information: clinical encounters, clinical trials, and health policy. *J Gen Intern Med* 1994;9:576-82.
2. Khanna D, Tsevat J. Health-related quality of life — an introduction. *Am J Manag Care* 2007;13 Suppl 9:S218-S223.
3. Fries JF, Spitz P, Kraines RG, Holman HR. Measurement of patient outcome in arthritis. *Arthritis Rheum* 1980;23:137-45.
4. Wolfe F, Kleinheksel SM, Cathey MA, Hawley DJ, Spitz PW, Fries JF. The clinical value of the Stanford Health Assessment Questionnaire Functional Disability Index in patients with rheumatoid arthritis. *J Rheumatol* 1988;15:1480-8.
5. Wolfe F, Mitchell DM, Sibley JT, et al. The mortality of rheumatoid arthritis. *Arthritis Rheum* 1994;37:481-94.
6. Torrance GW. Utility approach to measuring health-related quality of life. *J Chron Dis* 1987;40:593-603.
7. Brazier J, Roberts J, Deverill M. The estimation of a preference-based measure of health from the SF-36. *J Health Econ* 2002;21:271-92.
8. Ware JE J, Kosinski M, Keller S. SF-36 physical and mental health summary scales: A user's manual. Boston: Health Institute, New England Medical Center; 1994.
9. Kaplan RM, Groessl EJ, Sengupta N, Sieber WJ, Ganiats TG. Comparison of measured utility scores and imputed scores from the SF-36 in patients with rheumatoid arthritis. *Med Care* 2005;43:79-87.
10. Wolfe F, Michaud K, Gefeller O, Choi HK. Predicting mortality in patients with rheumatoid arthritis. *Arthritis Rheum* 2003;48:1530-42.
11. Bansback N, Marra C, Tsuchiya A, et al. Using the Health Assessment Questionnaire to estimate preference-based single indices in patients with rheumatoid arthritis. *Arthritis Rheum* 2007;57:963-71.
12. Khanna D, Oh M, Furst DE, et al. Evaluation of the preliminary definitions of minimal disease activity and remission in an early seropositive rheumatoid arthritis cohort. *Arthritis Rheum* 2007;57:440-7.
13. Khanna D, Wu H, Park G, et al. Association of tumor necrosis factor alpha polymorphism, but not the shared epitope, with increased radiographic progression in a seropositive rheumatoid arthritis inception cohort. *Arthritis Rheum* 2006;54:1105-16.
14. Sharp JT. Measurement of structural abnormalities in arthritis using radiographic images. *Radiol Clin North Am* 2004;42:109-19.
15. Fries JF, Spitz P, Kraines RG, Holman HR. Measurement of patient outcome in arthritis. *Arthritis Rheum* 1980;23:137-45.
16. Franzblau A. A primer of statistics for non-statisticians. New York: Harcourt Brace & World; 1958.
17. Khanna D, Furst DE, Wong WK, et al. Reliability, validity, and minimally important differences of the SF-6D in systemic sclerosis. *Qual Life Res* 2007;16:1083-92.
18. Khanna D, Yan X, Tashkin DP, et al. Impact of oral cyclophosphamide on health-related quality of life in patients with active scleroderma lung disease: Results from the scleroderma lung study. *Arthritis Rheum* 2007;56:1676-84.
19. Khanna D, Furst DE, Wong WK, et al. Reliability, validity, and minimally important differences of the SF-6D in systemic sclerosis. *Qual Life Res* 2007;16:1083-92.
20. Hays RD, Farivar S, Liu H. Approaches and recommendations for estimating minimally important differences for health-related quality of life measures. *COPD* 2005;2:63-7.
21. Wells G, Li T, Maxwell L, Maclean R, Tugwell P. Determining the minimal clinically important differences in activity, fatigue, and sleep quality in patients with rheumatoid arthritis. *J Rheumatol* 2007;34:280-9.
22. Strand V, Kelman A. Outcome measures in osteoarthritis: randomized controlled trials. *Curr Rheumatol Rep* 2004;6:20-30.
23. Khanna D, Pope JE, Khanna PP, et al. The minimally important difference for the fatigue visual analog scale in patients with rheumatoid arthritis followed in an academic clinical practice. *J Rheumatol* 2008;35:2339-43.
24. Cohen J. A power primer. *Psychol Bull* 1992;112:155-9.
25. Cost-benefit analysis. In: Drummond M, Sculpher M, Torrance GW, O'Brien BJ, Stoddart G, editors. *Methods for the economic evaluation of health care programmes*. 3rd ed. Oxford: Oxford Medical Publications; 2005:211-45.
26. Patrick DL, Erickson P. Health status and health policy: quality of life in health care evaluation and resource allocation. New York: Oxford University Press; 1993.
27. Kopec JA, Willison KD. A comparative review of four preference-weighted measures of health-related quality of life. *J Clin Epidemiol* 2003;56:317-25.
28. Brazier JE, Kolotkin RL, Crosby RD, Williams GR. Estimating a preference-based single index for the Impact of Weight on Quality of Life-Lite (IWQOL-Lite) instrument from the SF-6D. *Value Health* 2004;7:490-8.
29. Marra CA, Marion SA, Guh DP, et al. Not all "quality-adjusted life years" are equal. *J Clin Epidemiol* 2007;60:616-24.
30. 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.
31. Cronbach L, Furby L. How should we measure "change" — or should we? *Psychol Bull* 1970;74:68-80.
32. Hays RD, Hadorn D. Responsiveness to change: an aspect of validity, not a separate dimension. *Qual Life Res* 1992;1:73-5.
33. Hays RD. Reliability and validity (including responsiveness). In: Fayers P, Hays RD, editors. *Assessing quality of life in clinical trials*. 2nd ed. New York: Oxford; 2005:25-39.
34. Jaeschke R, Singer J, Guyatt GH. Measurement of health status. Ascertaining the minimal clinically important difference. *Control Clin Trials* 1989;10:407-15.
35. Walters SJ, Brazier JE. Comparison of the minimally important difference for two health state utility measures: EQ-5D and SF-6D. *Qual Life Res* 2005;14:1523-32.
36. Torrance GW. Preferences for health outcomes and cost-utility analysis. *Am J Manag Care* 1997;3 Suppl:S8-20.