Comparison of EuroQol-5D and Short Form-6D Utility Scores in Multiethnic Asian Patients with Psoriatic Arthritis: A Cross-sectional Study

Ying-Ying Leung, May-Ee Png, Hwee-Lin Wee, and Julian Thumboo

ABSTRACT. ***Objective.*** To compare EuroQol-5D (EQ-5D) and Short Form-6D (SF-6D) utility scores in multiethnic Asian patients with psoriatic arthritis (PsA).

Methods. Consecutive patients fulfilling the Classification Criteria for Psoriatic Arthritis attending a rheumatology outpatient clinic were recruited and completed the EQ-5D and SF-6D questionnaires. Comparisons were performed by score distribution, mean, median, and the Outcome Measures in Rheumatology filter: i.e., truth, discrimination, and feasibility.

Results. Eighty-six patients were enrolled (69 English-speaking and 17 Chinese-speaking; male:female ratio 0.91). The score distribution of SF-6D was normal, while that of EQ-5D was bimodal. A ceiling effect was observed in 20% of patients for EQ-5D and none for SF-6D. There were moderate correlations (Spearman's rho = 0.59, $p < 0.0001$) between the 2 scores, but poor agreements on scatterplot, intraclass correlation (ICC 0.43 and standardized ICC 0.21), and Bland-Altman plots. EQ-5D generated lower utility scores than SF-6D in the poorer health subgroup. SF-6D had stronger correlation with the general health status and other external measures of health; and it distinguished better between good and poor general health status, with better effect size and relative efficiency statistics. EQ-5D demonstrated higher patient acceptability.

Conclusion. EQ-5D and SF-6D instruments generated different utility scores in PsA. SF-6D may be superior because of normal scaling distribution and the absence of ceiling and floor effects. SF-6D also had better construct validity and better discrimination of poor health status. More studies are required for cost-utility analysis in PsA. (J Rheumatol First Release March 15 2013; doi:10.3899/jrheum.120782)

Key Indexing Terms:
PSORIATIC ARTHRITIS COMPARATIVE STUDY COST BENEFIT
QUALITY-ADJUSTED LIFE-YEAR

Psoriatic arthritis (PsA) has deleterious effects on joints and skin, causing joint deformities, impaired physical function, and impaired health-related quality of life (HRQOL). The introduction of anti-tumor necrosis factor-α (anti-TNF) therapies has dramatically changed the treatment paradigms for PsA. Given that health resources are finite, this change highlights the importance of cost-utility analysis (CUA), the primary outcome of which is cost per quality-adjusted life-year (QALY). Indirect HRQOL measures, such as the EuroQol-5D (EQ-5D)[1] and the Short form-6D (SF-6D)[2], are commonly used to elicit health state values for calculating QALY. Both instruments measure health in terms of physical function, pain, and mental health. Also, both have a scoring function derived from the statistical modeling of preferences for multideficit health states elicited from the general population of the United Kingdom[3]. Both instruments classify a respondent's self-reported health status according to a specific descriptive or classification system and assign a utility score. A utility score of 1 represents a state of perfect health and a utility score of 0 represents being dead.

Differences between these 2 instruments arise from a combination of differences in descriptive systems and valuations attached to the health states, and these may lead to different utility scores when applied to the same patient. These differences have been noted in other disease groups, including patients with rheumatic diseases[4,5,6,7,8,9]. Although gaps have been noted in the existing literature, to date no conclusion has been drawn with regard to which instrument performs better. This highlights the necessity for

From the Department of Rheumatology and Immunology, Singapore General Hospital; Department of Pharmacy, National University of Singapore; Yong Loo Lin School of Medicine, National University of Singapore; and Duke-National University of Singapore Graduate Medical School, Singapore.

Y-Y. Leung, MBChB, Department of Rheumatology and Immunology, Singapore General Hospital, Duke-National University of Singapore Graduate Medical School; M-E. Png, BSc, Department of Rheumatology and Immunology, Singapore General Hospital; H-L. Wee, PhD, Department of Pharmacy, National University of Singapore; J. Thumboo, MBBS, Department of Rheumatology and Immunology, Singapore General Hospital, Duke-National University of Singapore Graduate Medical School.

Address correspondence to Dr. Y-Y. Leung, Department of Rheumatology and Immunology, Singapore General Hospital, Outram Road, Singapore 169608. E-mail: katyccc@hotmail.com

Accepted for publication January 9, 2013.

Personal non-commercial use only. The Journal of Rheumatology Copyright © 2013. All rights reserved.

Downloaded on April 24, 2024 from www.jrheum.org

such comparisons to be made in a wider spectrum of diseases and sociocultural contexts.

The differences between the EQ-5D and SF-6D have also been observed in PsA cohorts that have poor health status and have undergone anti-TNF therapies[9]. However, the difference between the 2 utility scores has not been demonstrated in the general population of patients with PsA. It should be noted that the SF-6D has an advantage in PsA because it can be derived from the Medical Outcomes Study Short Form-36 (SF-36), which has been commonly used and extensively validated in PsA cohorts[10,11,12]. The aim of our study was to compare the utility scores EQ-5D and SF-6D in a population of multiethnic Asian patients who have PsA, particularly with regard to the aspects of truth, discrimination, and feasibility corresponding to the Outcome Measures in Rheumatology (OMERACT) filter[13]. We compared the 2 instruments for their distribution and agreement, construct validity, discriminant capacity, and acceptability.

MATERIALS AND METHODS

Data source and collection. From June 2010 through October 2011, we recruited consecutive patients with PsA [based on the Classification Criteria for Psoriatic Arthritis[14]] who attended an outpatient clinic at the rheumatology center at Singapore General Hospital. Patients of different ethnicities (Chinese, Malay, Indian, and others) were recruited. All participants completed identical and validated questionnaires, either the Singapore (English) or the Singapore (Simplified Chinese) version of the SF-36 Health Survey, version 2.0 (SF-36v2), and the EQ-5D, and provided sociodemographic data. The study was reviewed and approved by the Institutional Review Board of SingHealth. Before entry into the study, participants were informed of its nature and purpose and each participant signed an informed consent form.

Instruments. The EQ-5D is a standardized tool to measure health state. It has a 20 cm visual analog scale (EQ-VAS), which records the respondent's self-rated health as a score from 0 to 100 (0 representing worst imaginable health state and 100 best imaginable health state), and a descriptive system comprising 5 health domains (mobility, self-care, usual activities, pain/discomfort, and anxiety/depression). The 5-domain descriptive system classifies 243 different health states using 3 levels of severity for each domain (no problems, some problems, extreme problems). The EQ-5D tariff was estimated using the time tradeoff (TTO) method in a sample of 3395 respondents from the UK general population[1]. It consists of a set of numbers that indicates the level of HRQOL for each EQ-5D health state, on a scale from 1 (full health) to 0 [dead; range –0.594 to 1, where negative values are valued as worse than dead (WTD)]. The EQ-5D is commonly used in quantifying the influence of medical interventions on HRQOL. By comparing the difference in EQ-5D health states before and after treatment, analysts calculated the treatment effectiveness in QALY[15]. Both the English and Chinese versions of the EQ-5D have been validated in Singaporean patients with rheumatic diseases[16,17] and measurement equivalence has been demonstrated for the Singapore English and Chinese versions[18]. Favorable CUA for anti-TNF therapy in PsA has been reported[19].

The SF-6D is a classification tool for describing health as derived from 7 of the 8 health domains that are covered by the SF-36v2 health survey. It has 6 multilevel domains: physical functioning, role participation (combined role-physical and role-emotional), social functioning, bodily pain, mental health, and vitality. Each of the 6 domains has 4–6 levels of response, thus the SF-6D describes 18,000 health states. A preference tariff was estimated using the standard gamble (SG) method to obtain utility values from the UK general population on 249 of the possible health states[2]. The resulting SF-6D index, which ranges from 0.296 (worst health state) to 1.0 (best health state), was used in the assessment of QALY and the cost-effectiveness of various healthcare interventions. A few studies in PsA have demonstrated improvement in SF-6D scores with anti-TNF treatment beyond the minimal clinically important difference (MID) of 0.03[20,21]. The English and Chinese versions of the SF-6D have been demonstrated to be equivalent in the Singaporean population including patients with rheumatic diseases[22].

Statistical analysis. Collected data were entered into a Microsoft Excel spreadsheet and analyzed using SPSS software, version 17.0 (SPSS Inc.) and Stata/SE software version 11.0 (StataCorp LP). All statistical tests were 2-tailed and conducted at a 5% level of significance.

EQ-5D and SF-6D utility scores were calculated using the scoring algorithms developed by Dolan[1] and Brazier, *et al*[2], respectively. Missing data in SF-36v2 were estimated according to protocol. We compared the distribution, mean, median, and agreement between the utility scores. For distribution, we compared the mean ± SD and median [interquartile range (IQR)] utility scores generated by the 2 instruments. Ceiling and floor effects were considered present if > 15% of participants responded giving the highest and lowest possible scores[23]. The within-subject differences in the 2 utility scores were compared by t test. A limit of 0.03 between scores was chosen based on the smallest estimate of the published results of MID for the EQ-5D and SF-6D[7,24]. To assess the degree of agreement between the 2 utility scores, we used the interclass correlation coefficient (ICC; 2-way random-effects model with absolute agreement) and a Bland-Altman plot[25]. An ICC > 0.7 suggests an acceptable level of agreement[26]. As the lower bound of the 2 utility scores differs, we standardized the utility scores linearly to fit the range of 0–1 (based on the theoretical possible range).

For the aspect of truth, we compared the construct validity using the Spearman's rank correlation between the 2 utility scores and the SF-6D summary scores, physical component summary (PCS), mental component summary (MCS), and SF-6D general health (SF-GH). The SF-GH is the participant's response to the first question of the SF-36: "In general, would you say your health is — ?" and it is not included in the calculation of the SF-6D.

For the aspect of discrimination or discriminant capacity, we examined the discriminatory capacities of the 2 utility scores to distinguish between participants with contrasting health states. Participants with different levels of impairment were classified according to the SF-GH. The SF-GH was analyzed with the following categories: "excellent/very good," "good," and "fair/poor." The ability to differentiate between SF-GH "excellent/very good" versus "good" and "good" versus "fair/poor" subgroups was calculated using 1-way ANOVA. The effect size was calculated as the standardized mean difference described by Cohen[27] (i.e., the difference in mean scores divided by the pooled SD). The effect size was categorized as small (0.2–0.5), moderate (> 0.5–0.8), or large (> 0.8). The influence of sociodemographic factors was evaluated by t test or 1-way ANOVA. The ability of the utility scores to detect difference between health status of SF-GH was tested by the relative effective statistics (RE). This is calculated as the ratio of the square of t statistics of the SF-6D utility score to the square of t statistics of the EQ-5D. An RE > 1.0 indicates that SF-6D is more efficient than the EQ-5D in detecting the difference. The reverse is true if the RE is < 1.0. We evaluated the RE of the 2 utility scores in differentiating health status according to SF-GH, "excellent/very good" versus "good" and "good" versus "fair and poor"[28]. For "feasibility," we reported the proportion of missing data for each utility score.

RESULTS

Characteristics of participants. Eighty-six participants (69 English-speaking and 17 Chinese-speaking) were enrolled into the study. Characteristics of participants in each ethnic group and in the total sample are shown in Table 1. The

Personal non-commercial use only. The Journal of Rheumatology Copyright © 2013. All rights reserved.

Downloaded on April 24, 2024 from www.jrheum.org

Table 1. Participant characteristics and sociodemographics.

Characteristic	Total, n = 86	Chinese, n = 46	Indian, n = 27	Malay and Others, n = 13
Age, yrs, mean ± SD	49.06 ± 13.38	50.78 ± 12.95	49.48 ± 13.55	42.08 ± 13.33
Female, n (%)	45 (52.3)	25 (54.35)	15 (55.56)	5 (38.46)
Duration of illness, yrs, mean ± SD	7.12 ± 8.18	6.53 ± 7.01	8.42 ± 10.16	6.62 ± 7.97
Married, n (%)	55 (64.0)	33 (71.74)	13 (28.26)	8 (17.39)
Occupation, n (%)				
Employee/employer/self-employed	55 (64.0)	29 (63.04)	18 (66.67)	8 (61.54)
Retired	14 (16.3)	10 (21.74)	2 (7.41)	2 (15.38)
Homemaker	9 (10.5)	3 (6.52)	4 (14.81)	2 (15.38)
Unemployed	5 (5.8)	3 (6.52)	2 (7.41)	0 (0.00)
Student/national service	3 (3.5)	1 (2.17)	1 (3.70)	1 (7.69)
SF-General health, n (%)				
Excellent/very good	9 (10.47)	5 (10.87)	2 (7.41)	2 (15.38)
Good	36 (41.86)	16 (34.78)	11 (40.74)	9 (69.23)
Fair/poor	41 (47.67)	25 (54.35)	14 (51.85)	2 (15.38)
SF-36-PCS, mean ± SD	36.89 ± 13.99	37.23 ± 13.89	34.77 ± 15.17	40.09 ± 11.91
SF-36-MCS, mean ± SD	43.48 ± 11.35	43.51 ± 11.76	41.35 ± 10.20	47.78 ± 11.74
EQ-5D-VAS, mean ± SD	67.25 ± 18.71	68.96 ± 19.09	60.67 ± 18.80*	75.00 ± 13.25
EQ-5D-utility, mean ± SD	0.74 ± 0.24	0.77 ± 0.22	0.67 ± 0.27	0.76 ± 0.19
SF-6D-utility, mean ± SD	0.68 ± 0.13	0.69 ± 0.13	0.65 ± 0.14	0.70 ± 0.14

* $p < 0.05$. SF: Short form-6D; PCS: physical component summary; MCS: mental component summary; VAS: visual analog scale; EQ-5D: EuroQol-5D.

HRQOL among participants with PsA was much lower than that found among the healthy "normal" population: both the norm-based SF-36 PCS and MCS were below the norm mean of 50. The characteristics of PsA participants from different ethnicities were generally similar, except that the EQ-VAS was poorer among Indian participants with PsA.

EQ-5D and SF-6D distribution and agreement. The score distribution for the SF-6D was normal (skewness = 0.27, kurtosis = –0.37, p = 0.10 by the Kolmogorov-Smirnov test), while that of the EQ-5D was bimodal (skewness = –1.62, kurtosis = 2.80, $p < 0.001$ by Kolmogorov-Smirnov test; Figure 1). The 3 most commonly reported EQ-5D profiles were 11111 (20%), 11121 (29.4%), and 11122 (10.6%), whereas the reported SF-6D profiles were spread across all states, none of which was reported by > 2 participants (2.4%). A ceiling effect was observed in EQ-5D (range –0.014 to 1.0), where 20% of participants responded with the highest possible score and 2.3% of participants had negative scores for EQ-5D corresponding to the WTD state. No ceiling or floor effects were observed for the SF-6D (range 0.355 to 1.0). The mean (± SD) EQ-5D and SF-6D utility scores were 0.74 (± 0.24) and 0.68 (± 0.13), respectively (p = 0.001). The median EQ-5D and SF-6D utility scores were 0.8 (IQR 0.09) and 0.64 (IQR 0.18). There was a mean difference of 0.05 (± 0.2) between the utility scores, which was higher than the MID of 0.03 that we chose for comparison. The paired utility scores are presented graphically as a scatterplot for the entire population (Figure 2); the Spearman product-moment correlation coefficient between EQ-5D and SF-6D was 0.59 ($p < 0.0001$). The deviation from the 45-degree line was evident, particularly in the low end of the utility scores. In Bland-Altman plots, EQ-5D scores were systemically lower than the SF-6D in subjects with lower averaged utility scores (Figure 2). Poor agreement between EQ-5D and SF-6D utility scores was demonstrated with the low ICC (0.43, 95% CI 0.23 to 0.59) for the entire population. The ICC was even lower when it was standardized (0.21, 95% CI –0.09 to 0.50). There were wide limits of agreement on the Bland-Altman plot (1.96 SD from –0.35 to 0.46; Figure 2).

Aspect of truth. For construct validity, the Spearman's rho correlation between the EQ-5D and SF-36v2 summary scores and SF-GH were moderate (0.37–0.45), while the correlations for the SF-6D with SF-GH status were good (0.57–0.84; Table 2). There were similar correlations between the 2 utility scores and the EQ-VAS. At the domain level, there were moderate correlations between the EQ-5D and SF-6D domains that measure similar constructs. The Spearman's rho were 0.53 ($p < 0.001$) between SF-6D pain and EQ-5D pain/discomfort, and 0.48 ($p < 0.001$) between SF-6D mental health and EQ-5D anxiety/depression. SF-6D physical functioning correlated weakly with EQ-5D mobility, self-care, and usual activities (r = 0.33–0.39), which reflects that these domains measured different aspects of HRQOL.

Aspect of discrimination. The utility scores did not differ by sex, age, marital status, ethnicity, languages, or education level. For discriminant capacity, the SF-6D distinguished between participants with "better" or "poor" health status, with strong effect sizes. The EQ-5D distinguished partici-

Personal non-commercial use only. The Journal of Rheumatology Copyright © 2013. All rights reserved.

Downloaded on April 24, 2024 from www.jrheum.org

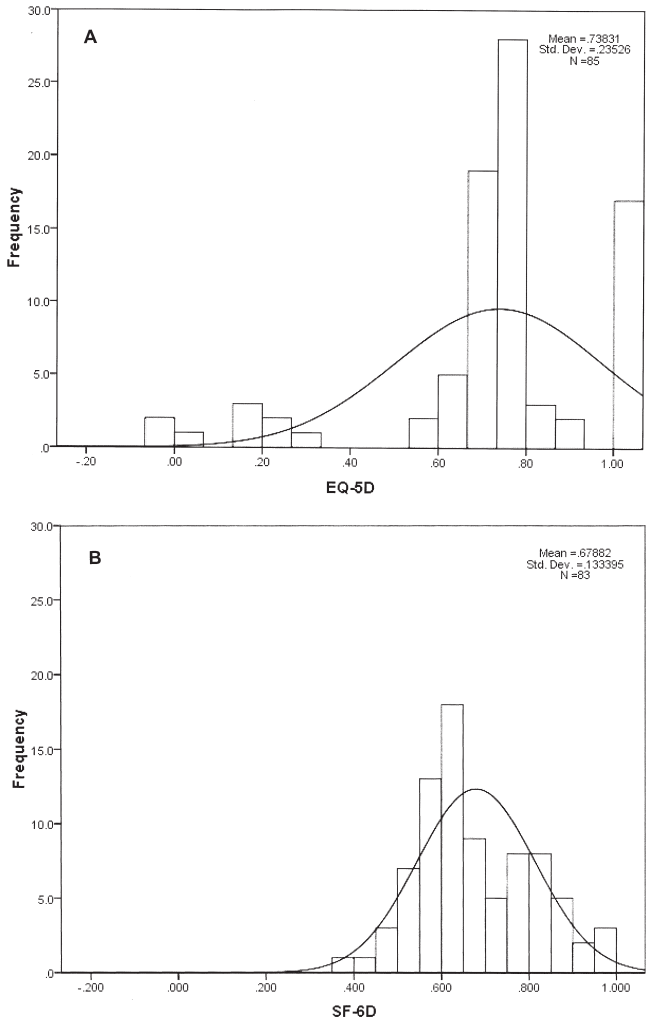

Figure 1. A. Score distribution of the EQ-5D was normal (skewness = 0.27, kurtosis = –0.37, p = 0.10, Kolmogorov-Smirnov test). B. Score distribution of the SF-6D was bimodal (skewness = –1.62, kurtosis = 2.80, p < 0.001, Kolmogorov-Smirnov test).

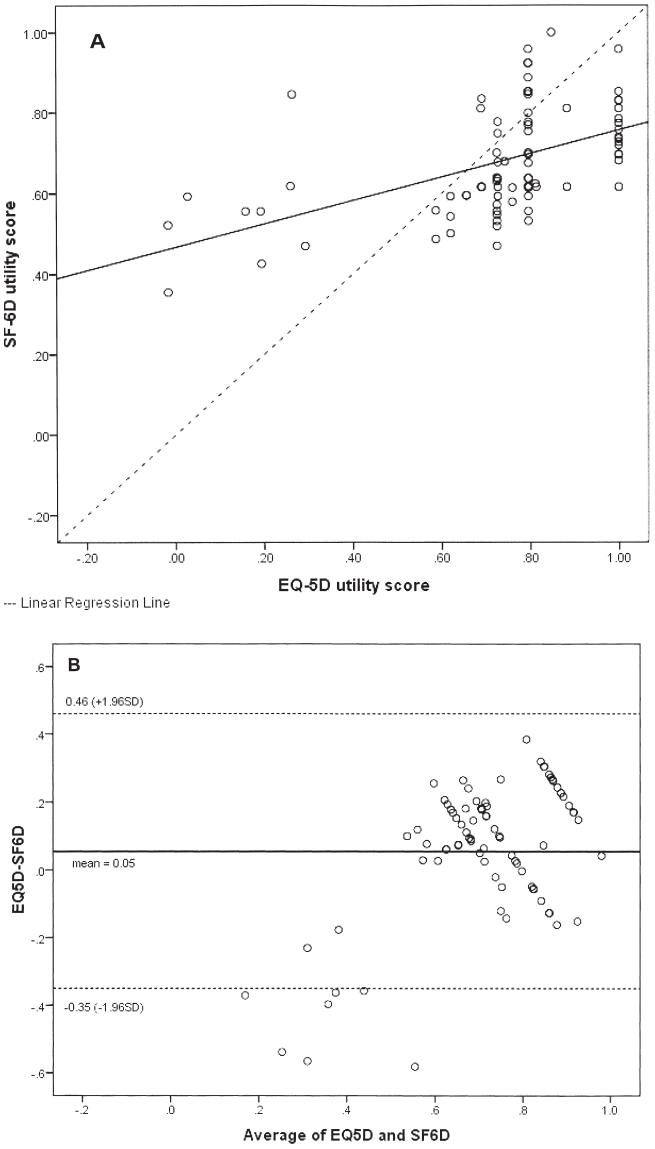

Figure 2. A. Paired utility scores for the entire population; Spearman product-moment correlation coefficient between EQ-5D and SF-6D was 0.59 (p < 0.0001). Deviation from the 45-degree line can be seen, particularly in the low end of the utility scores. B. In Bland-Altman plot, EQ-5D scores were systemically lower than the SF-6D in subjects with lower averaged utility scores.

pants with "good" versus "fair/poor" health status with only moderate effect size (Table 3). Considering the difference between "excellent/very good" and "good" of the SF-general health, the RE score was 0.98, implying that the SF-6D was as efficient as the EQ-5D in differentiating these 2 health status. Also, the RE was 1.95 for differentiating "good" from "fair/poor" of the SF-GH. This implied that SF-6D was 95% more efficient than EQ-5D in identifying patients with "good" or "poor" general health status.

Aspect of feasibility. The EQ-5D has 5 items, which is much fewer than the SF-6D, which is derived from the 36 items of the SF-36. The SF-6D individual items are rarely presented in isolation. The acceptability in terms of completion rate is higher in the EQ-5D. Missing raw data was 1.2% for the EQ-5D. It was 9.3% for SF-6D, and was then reduced to 3.5% by estimating the missing data according to the SF-36v2 protocol.

DISCUSSION

Using an appropriate and valid utility index is a major issue in cost-utility analysis. Therefore, it is important to understand the performance of different indirect utility instruments in various diseases. In this study, we presented and compared utility data in terms of distribution, agreement, and the OMERACT filter, which comprises truth, discrimination, and feasibility, in a cohort of multiethnic Asian patients with PsA recruited from a secondary and tertiary referral rheumatology center. We found a 0.05 difference in the utility scores generated by the 2 instruments, which is

Personal non-commercial use only. The Journal of Rheumatology Copyright © 2013. All rights reserved.

Downloaded on April 24, 2024 from www.jrheum.org

Table 2. Correlation of EuroQol-5D (EQ-5D) and Short form-6D (SF-6D) with SF-general health and SF-36 summary scores.

	Spearman's Rho				
	SF-6D Utility	SF-GH	PCS	MCS	EQ-VAS
EQ-5D utility	0.594*	–0.440*	0.445*	0.371*	0.494*
SF-6D utility	—	–0.569*	0.843*	0.623*	0.538*

* $p < 0.0001$. PCS: physical component summary of SF-36v2; SF-36: Medical Outcomes Study Short Form-36; MCS: mental component summary of SF-36v2; SF-GH: SF-36-general health; VAS: visual analog scale.

Table 3. Discriminant capacity of EQ-5D and SF-6D utility scores.

Measure		Utility Score, mean (SD)	p*	Effect Size
EQ-5D	SF-general health			
	Excellent/very good	0.91 ± 0.11		0.91**
	Good	0.79 ± 0.15		
	Fair/poor	0.65 ± 0.28	0.001	0.62***
SF-6D	SF-general health			
	Excellent/very good	0.82 ± 0.09		0.94
	Good	0.72 ± 0.12		—
	Fair/poor	0.61 ± 0.12	< 0.0001	0.92

* One-way ANOVA. ** Comparison between excellent/very good versus good subgroup. *** Comparison between good versus fair/poor subgroup. EQ-5D: EuroQol-5D; SF-6D: Short form-6D.

higher than the smallest published MID value of 0.03 for the SF-6D[7,24]. In a review comparing 8 longitudinal studies across 11 patient groups, the mean MID for the SF-6D was 0.041 and the mean MID for the EQ-5D was 0.074[29]. We acknowledge that the difference we found between the 2 utility scores was lower than the upper limits of MID, yet this difference was substantial.

Our study also demonstrated poor agreement and ICC with utility scores generated by the EQ-5D and SF-6D. Similar findings have been reported in cohorts of rheumatoid arthritis (RA)[7], ankylosing spondylitis (AS)[8], inflammatory arthritis[9], and early arthritis[30,31]. A common finding in comparison studies is that EQ-5D tends to generate lower utilities than SF-6D in subgroups with poorer health, and that the reverse is true in the healthier subgroups[8,32,33]. Our study also revealed similar findings in a PsA cohort. There were more prominent deviations from the 45-degree line on the scatterplot and negative values for the difference between EQ-5D and SF-6D in the lower end of average score on a Bland-Altman plot. Adams, *et al*[9] demonstrated similar phenomena in patients with inflammatory arthritis (345 RA and 159 PsA) before they started biological therapy and again 12 months later. At baseline, 12% of the PsA participants reported a negative utility score with EQ-5D, which corresponded to a status of WTD. Significantly lower utility scores with EQ-5D compared to SF-6D were observed, and the participant group with the WTD state might have directly contributed to the large difference (about 2-fold) in QALY gain for a given change in health status. Most criticisms about EQ-5D were not about the instrument itself but the preference-based values to the raw TTO scores and how the WTD status is handled[33,34]. Using the same database, Adams, *et al* illustrated the influence of using a revised scoring method for EQ-5D in CUA analysis[35]. They demonstrated that the revised EQ-5D has a lesser difference between the utility scores and the change in utility after biological treatment. However, that study did not conclude how this new EQ-5D scoring system may alter the utility estimates, and ultimately the results of an economic model[36]. Although the HRQOL of our cohort was much worse than that of the general population, the general health status of our participants was better than that reported by Adams, *et al*, and the proportion of participants with the WTD state in our study cohort was only 2.3%. However, the large difference between the utilities generated by these 2 instruments was still present. This implies that the difference between the 2 utility scores is not just related to the preference-based weights and methods of handling the WTD status. In Singapore, where there is no primary healthcare that supports the care of chronic inflammatory arthritis, the majority of patients with PsA are cared for in secondary and tertiary centers. Hence, we believe that our sample represents the whole spectrum of illness in PsA.

It is well known that different indirect utility instruments may yield different utility scores[37,38]. Many attributed this to the different methods by which these health preference utility scores were derived (TTO for EQ-5D and SG for SF-6D). The SF-6D describes more health states than does EQ-5D (18,000 vs 243 health states) and therefore may capture more health states at the extreme ends of the range and may capture smaller health changes[39]. However, only a minority of these states have been valued by SG, and not all the states were valued by TTO when weights of the EQ-5D were calculated. This means that most of the health states carry utility values that were estimated from the utility function rather than being measured directly. Moreover, it should be emphasized that the utility scales are anchored at 0 and 1, while they are by no means representing variables lying on interval scales. The weak interval properties of various utility scores generated from indirect HRQOL instruments have been illustrated in a large-scale comparison[40]. Indeed, the descriptive systems in different HRQOL differ widely in their coverage of different dimensions of HRQOL and thus the reported differences in utility scores are attributable, in part, to these differences.

There has been debate whether the valuation of a health state may be a better reflection of the true welfare value that is associated with health[41], while criticisms are that utility scores and general health evaluations are basically measure-

Personal non-commercial use only. The Journal of Rheumatology Copyright © 2013. All rights reserved.

Downloaded on April 24, 2024 from www.jrheum.org

ments of different constructs. In AS, Boonen, *et al*[8] have shown that the EQ-5D, SF-6D, and EQ-VAS correlated well with external health measurements, with only moderate agreement; the disagreement between EQ-5D and EQ-VAS was more prominent in the subgroup with poorer health. There is, however, growing interest using the general health measure as a composite measurement of disease activity in PsA. The PsA index, which consisted of patient global assessment, skin global assessment, and physician global assessment, was found to explain > 90% of the variance in the baseline scores of the Group for Research and Assessment of Psoriasis and Psoriatic Arthritis (GRAPPA) Composite Exercise project[42], and was taken forward for further evaluation in the GRAPPA 2010 and OMERACT 11 meetings[43]. In PsA, patient global assessment (PGA) has been shown to have reasonable reliability[44] and construct validity[45]. We observed better ICC and limits of agreement for the SF-6D with EQ-VAS (data not shown), which may imply that the SF-6D provides a better reflection of general health. Although the EQ-VAS is a similar general health measurement, it has not been evaluated as a valid measure of general health status or composite disease activity score in PsA. Therefore, this finding requires further validation.

The measurement of HRQOL in PsA is under investigation, and there have been limited data on using indirect HRQOL measures for CUA in PsA. Our study adds information to the literature on the comparison of performance of these 2 utilities in PsA. In terms of "truth," the SF-6D utility score performed better, with higher correlation with SF-36 summary scores and SF-GH. Caution in interpretation is needed, in that the SF-6D and summary scores were derived from the same instrument, the SF-36. However, SF-GH was not included in the calculation of the SF-6D. In terms of "discrimination," the SF-6D utilities performed better in differentiating participants with poorer health status in terms of effect sizes and the RE scores. Gaujoux-Viala, *et al*[30] demonstrated similar systematic disagreement between EQ-5D and SF-6D in a prospective study on 813 patients with early arthritis, especially in patients with worse clinical outcomes. The SF-6D was shown to have better RE statistics[30]. In their longitudinal study over 2 years, the same group of authors also recently demonstrated better responsiveness for improvement with the SF-6D in standardized response mean and effect sizes[31].

For "feasibility," the SF-6D adds more missing values than the EQ-5D, although the missing values could be minimized by substituting with the average from the other items in the same subscale. For distribution, the SF-6D was better in terms of having normal scaling distribution and the absence of ceiling and floor effects. However, further research is required to determine which instrument performs better.

There are several limitations to our study. First, the construct validity assessments are limited to the SF-GH. No other assessments of disease activity, such as joint count or Health Assessment Questionnaire data, were collected. Second, this was a cross-sectional study and the discrimination evaluation is limited to how the utility scores differentiate different health states, instead of a change in status with treatment over time. Third, we did not address the reliability for these utility scores in PsA. Fourth, our study cohort consisted of PsA patients with a long duration of illness, recruited from a single center, which limits the study's generalizability to patients with early PsA. Our sample size was relatively small and this may affect the interpretation of ceiling and floor effects. Moreover, HRQOL measures are heavily influenced by comorbidities that may introduce inaccuracy. Finally, the scoring algorithms used for both instruments were developed from the UK general population, because no such algorithm was available in Singapore at the time of the study.

Despite these limitations, we have demonstrated in a cohort of patients with PsA that the SF-6D performed slightly better in terms of construct validity and discrimination than the EQ-5D, which is shorter and more feasible in clinical practice. The SF-6D also had normal distribution and lack of ceiling effects. However, the 2 instruments yielded different utility scores in PsA. This would have a great effect on QALY estimates, and it highlights the importance of choosing the appropriate instrument for cost-effectiveness evaluation. Additional research is needed to determine whether the EQ-5D or the SF-6D is the better instrument for cost-utility analysis in PsA.

ACKNOWLEDGMENT

The authors appreciate the support of the Duke-NUS/SingHealth Academic Medicine Research Institute and the editorial services of Jon Kilner, MS, MA, Pittsburgh, PA, USA, and Taara Madhavan, Associate, Duke-NUS Graduate Medical School, Singapore.

REFERENCES

1. Dolan P. Modeling valuations for EuroQol health states. Med Care 1997;35:1095–108.
2. 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.
3. 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.
4. Brazier J, Roberts J, Tsuchiya A, Busschbach J. A comparison of the EQ-5D and SF-6D across seven patient groups. Health Econ 2004;13:873–84.
5. Conner-Spady B, Suarez-Almazor ME. Variation in the estimation of quality-adjusted life-years by different preference-based instruments. Med Care 2003;41:791–801.
6. Gerard K, Nicholson T, Mullee M, Mehta R, Roderick P. EQ-5D versus SF-6D in an older, chronically ill patient group. Appl Health Econ Health Policy 2004;3:91–102.
7. Marra CA, Woolcott JC, Kopec JA, Shojania K, Offer R, Brazier JE, 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.

Personal non-commercial use only. The Journal of Rheumatology Copyright © 2013. All rights reserved.

Downloaded on April 24, 2024 from www.jrheum.org

8. Boonen A, van der Heijde D, Landewé R. How do the EQ-5D, SF-6D and the well-being rating scale compare in patients with ankylosing spondylitis? Ann Rheum Dis 2007;66:771-7.
9. Adams R, Walsh C, Veale D, Bresnihan B, FitzGerald O, Barry M. Understanding the relationship between the EQ-5D, SF-6D, HAQ and disease activity in inflammatory arthritis. Pharmacoeconomics 2010;28:477-87.
10. Husted JA, Gladman DD, Farewell VT, Long JA, Cook RJ. Validating the SF-36 health survey questionnaire in patients with psoriatic arthritis. J Rheumatol 1997;24:511–7.
11. Leung YY, Ho KW, Zhu TY, Tam LS, Kun EW, Li EK. Testing scaling assumptions, reliability and validity of Medical Outcomes Study Short-form 36 health survey in psoriatic arthritis. Rheumatology 2010;49:1495–501.
12. Leung YY, Tam LS, Kun EW, Ho KW, Li EK. Comparison of 4 functional indexes in psoriatic arthritis with axial or peripheral disease subgroups using Rasch analyses. J Rheumatol 2008;35:1613-21.
13. Boers M, Brooks P, Strand CV, Tugwell P. The OMERACT filter for Outcome Measures in Rheumatology. J Rheumatol 1998;25:198-9.
14. Taylor W, Gladman D, Helliwell P, Marchesoni A, Mease P, Mielants H, et al. Classification criteria for psoriatic arthritis: Development of new criteria from a large international study. Arthritis Rheum 2006;54:2665-73.
15. The PRET project: Preparation for the re-valuation of the EQ-5D tariff. Sheffield: The University of Sheffield; 2011. (Internet. Accessed January 25, 2013.) Available from: http://www.shef.ac.uk/scharr/sections/heds/mvh/pret
16. Luo N, Chew LH, Fong KY, Koh DR, Ng SC, Yoon KH, et al. Validity and reliability of the EQ-5D self-report questionnaire in Chinese-speaking patients with rheumatic diseases in Singapore. Ann Acad Med Singapore 2003;32:685-90.
17. Luo N, Chew LH, Fong KY, Koh DR, Ng SC, Yoon KH, et al. Validity and reliability of the EQ-5D self-report questionnaire in English–speaking Asian patients with rheumatic diseases in Singapore. Qual Life Res 2003;12:87-92.
18. Luo N, Chew LH, Fong KY, Koh DR, Ng SC, Yoon KH, et al. Do English and Chinese EQ-5D versions demonstrate measurement equivalence? An exploratory study. Health Qual Life Outcomes 2003;1:7.
19. Bansback NJ, Ara R, Barkham N, Brennan A, Fraser AD, Conway P, et al. Estimating the cost and health status consequences of treatment with TNF antagonists in patients with psoriatic arthritis. Rheumatology 2006;45:1029–38.
20. Strand V, Sharp V, Koenig AS, Park G, Shi Y, Wang B, et al. Comparison of health-related quality of life in rheumatoid arthritis, psoriatic arthritis and psoriasis and effects of etanercept treatment. Ann Rheum Dis 2012;71:1143-50.
21. Saad A, Ashcroft DM, Watson KD, Symmons DP, Noyce PR, Hyrich KL, on behalf of the British Society for Rheumatology Biologics Register. Improvements in quality of life and functional status in patients with psoriatic arthritis receiving anti–tumor necrosis factor therapies. Arthritis Care Res 2010;62:345-53.
22. Wee HL, Cheung YB, Fong KY, Luo N, Machin D, Thumboo J. Are English- and Chinese-language versions of the SF-6D equivalent? A comparison from a population-based study. Clin Ther 2004;26:1137-48.
23. Veenhof C, Bijlsma JW, van den Ende CH, van Dijk GM, Pisters MF, Dekker J. Psychometric evaluation of osteoarthritis questionnaires: A systematic review of the literature. Arthritis Rheum 2006;55:480-92.
24. Barton GR, Sach TH, Avery AJ, Jenkinson C, Doherty M, Whynes DK, et al. A comparison of the performance of the EQ-5D and SF-6D for individuals aged > or = 45 years. Health Econ 2008;17:815-32.
25. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet 1986;1:307–10.
26. Fayers PM, Machin D. Quality of life: The assessment, analysis and interpretation of patient-reported outcomes. Chichester: John Wiley & Sons Ltd.; 2007.
27. Cohen J. Statistical power for the behavioral sciences. 2nd ed. Hillsdale, NJ: Erlbaum Associates; 1988.
28. Petrou S, Hockley C. An investigation into the empirical validity of the EQ-5D and SF-6D based on hypothetical preferences in general population. Health Econ 2005;14:1169-89.
29. 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.
30. Gaujoux-Viala C, Rat AC, Guillemin F, Flipo RM, Fardellone P, Bourgeois P, et al. Comparison of the EQ-5D and the SF-6D utility measures in 813 patients with early arthritis: Results from the ESPOIR cohort. J Rheumatol 2011;38:1576-84.
31. Gaujoux-Viala C, Rat AC, Guillemin F, Flipo RM, Fardellone P, Bourgeois P, et al. Responsiveness of EQ-5D and SF-6D in patients with early arthritis: Results from the ESPOIR cohort. Ann Rheum Dis 2012;71:1478-83.
32. Seymour J, McNamee P, Scott A, Tinelli M. Shedding new light onto the ceiling and floor? A quantile regression approach to compare EQ-5D and SF-6D responses. Health Econ 2010;19:683-96.
33. Marra CA, Marion SA, Guh DP, Najafzadeh M, Wolfe F, Esdaile JM, et al. Not all "quality-adjusted life years" are equal. J Clin Epidemiol 2007;60:616.
34. Patrick DL, Starks HE, Cain KC, Uhlmann RF, Pearlman RA. Measuring preferences for health states worse than death. Med Decis Making 1994;14:9–18.
35. Adams R, Craig BM, Walsh CD. The impact of a revised EQ-5D population scoring on preference-based utility scores in an inflammatory arthritis cohort. Value Health 2011;14:921-7.
36. Craig BM, Busschbach JJ. Toward a more universal approach in health valuation. Health Econ 2011;20:864-75.
37. Conner-Spady B, Suarez-Almazor ME. Variation in the estimation of quality-adjusted life-years by different preference-based instruments. Med Care 2003;41:791-801.
38. Wee HL, Machin D, Loke WC, Li SC, Cheung YB, Luo N, et al. Assessing differences in utility scores: A comparison of four widely used preference-based instruments. Value Health 2007;10:256-65.
39. Bryan S, Longworth L. Measuring health-related utility: Why the disparity between EQ-5D and SF-6D? Eur J Health Econ 2005;6:253-60.
40. 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.
41. Parkin D, Devlin N. Is there a case for using visual analogue scale valuations in cost-utility analysis? Health Econ 2006;15:653–64.
42. Helliwell PS, Fitzgerald O, Strand CV, Mease PJ. Composite measures in psoriatic arthritis: A report from the GRAPPA 2009 annual meeting. J Rheumatol 2011;38:540-5.
43. Coates LC, Mumtaz A, Helliwell PS, Mease PJ, Callis Duffin K, Krueger GG, et al. Development of a disease severity and responder index for psoriatic arthritis (PsA) — Report of the OMERACT 10 PsA Special Interest Group. J Rheumatol 2011;38:1496-501.
44. Cauli A, Gladman DD, Mathieu A, Olivieri I, Porru G, Tak PP, et al. Patient global assessment in psoriatic arthritis: A multicenter GRAPPA and OMERACT study. J Rheumatol 2011;38:898–903.
45. Leung YY, Ho KW, Zhu TY, Tam LS, Kun EW, Li EK. Construct validity of the modified numeric rating scale of patient global assessment in psoriatic arthritis. J Rheumatol 2012;39:844-8.

Personal non-commercial use only. The Journal of Rheumatology Copyright © 2013. All rights reserved.

Downloaded on April 24, 2024 from www.jrheum.org