EXTENDED REPORT

Responsiveness of EQ-5D and SF-6D in patients with early arthritis: results from the ESPOIR cohort

Cécile Gaujoux-Viala,[1,2] Anne-Christine Rat,[1,3] Francis Guillemin,[1,3] René-Marc Flipo,[4] Patrice Fardellone,[5] Pierre Bourgeois,[2] Bruno Fautrel[2]

► An additional figure is published online only. To view these files, please visit the journal online (http://ard.bmj.com/content/early/recent).

[1]Lorraine University, Paris Descartes University, EA 4360 Apemac, Nancy, France
[2]Pierre et Marie Curie University (UPMC) - Paris 6, AP-HP Pitié Salpêtrière Hospital, Department of Rheumatology, 75013 Paris, France
[3]INSERM, CIC-EC CIE6, Nancy, France
[4]Department of Rheumatology, Lille University 2, Lille, France
[5]Department of Rheumatology, Amiens University, Amiens, France

Correspondence to
Cécile Gaujoux-Viala, Université Paris 6 -Pierre & Marie Curie, AP-HP, Groupe hospitalier Pitié-Salpêtrière, Service de Rhumatologie, 83 boulevard de l'Hôpital 75651 Paris cedex 13, France;
cecilegaujoux@yahoo.fr

Received 10 October 2011
Accepted 27 January 2012
Published Online First
14 March 2012

ABSTRACT

Objectives The revolution of early aggressive treatments for early arthritis (EA) has fuelled the search for better approaches to establishing their cost–utility ratio. The authors aimed to compare the responsiveness of the EQ-5D and the SF-6D in a large prospective cohort of patients with EA.

Methods EQ-5D and SF-6D utility measures were assessed in 813 patients with EA over 2 years. Responsiveness was analysed by the standardised response mean (SRM) and effect size between baseline and 6, 12 and 24 months for the entire sample and subgroups by disease evolution (increase or decrease in Disease Activity Score for 28 joints). Bootstrap methods were used to estimate 95% CI.

Results The EQ-5D provided larger absolute mean change estimates with greater variance than the SF-6D, whatever the direction of change. At 12 months, the SF-6D was more sensitive to change with improved condition than the EQ-5D: SRM 0.83 (0.82 to 0.84) versus 0.57 (0.56 to 0.58). In contrast, the EQ-5D was more sensitive to change with deteriorated condition than the SF-6D: SRM −0.20 (−0.23 to −0.18) versus −0.11 (−0.14 to −0.08). Results were similar for 6 and 24 months.

Conclusions The SF-6D was more responsive than the EQ-5D with improved EA condition. Confidence in the relative cost-effectiveness of two treatments would be better with the SF-6D because of its smaller variance. The SF-6D provided more conservative cost-effectiveness ratios than the EQ-5D and may be more appropriate for trials of biological treatments for patients with EA.

INTRODUCTION

Preference-based measures of health have become important for estimating health states in calculations of quality-adjusted life years (QALYs), an essential component of cost–utility analysis. The EQ-5D[1] and the SF-6D[2] are indirect preference-based health-related quality of life (HRQoL) instruments that are increasingly being used for an economic evaluation of clinical interventions and health programs. Although the theoretical concept of utility implies that one specific health state has one utility score, regardless of how it is measured, different instruments can give different scores.[3] The explosion of drug development for rheumatoid arthritis (RA) and the revolution of early aggressive treatments for the disease have fuelled the search for better approaches to establishing cost-effectiveness in early arthritis (EA), but consensus on the choice of utility instrument is lacking. The choice of instrument may affect the results of future studies on new biological agents and their cost-effectiveness. We need a consensus based on the relative merits of the instruments from evidence on their practicality, reliability, construct validity, discriminant validity and responsiveness, as well as their overall suitability for evaluative purposes. Thus, if the measurement properties of the two instruments are close but the elicited utility levels differ, sensitivity analyses using the two measurements of utility could be appropriate for determining cost–utility ratios.

Several studies, mainly cross-sectional, have compared EQ-5D and SF-6D scores for patients with RA;[4–7] a common finding by the two measures is the presence of small but significant differences between utility estimates.[8–10] However, few comparisons for EA exist. In a previous study,[11] we demonstrated a systematic disagreement between EQ-5D utility and SF-6D utility for EA, especially for patients with worse clinical outcomes: the utility was systematically lower with the EQ-5D. We have also assessed their construct validity and discrimination capacity.[11]

One other important aspect of validity is the ability of a measure to reflect the change in patients' condition over time. The EQ-5D and the SF-6D can detect some degree of change for patients with RA.[12–15] In two studies of North American populations, the SF-6D appeared more responsive to health improvement in patients with RA than was the EQ-5D.[13 14] However, other results have been conflicting. In patients with different rheumatological conditions (51% RA), the EQ-5D was more responsive to improvement than was the SF-6D.[15] Harrison *et al*[16] examined the responsiveness of the EQ-5D and the SF-6D in cohorts of patients with early inflammatory disease or established RA. The SF-6D did not respond well to deteriorated conditions in patients with established severe RA, and its use for patients with severe progressive disease may be inappropriate. In EA, data were limited. Only data on very early inflammatory arthritis (4–11 weeks' duration, n=182) included in the Steroids in Very Early Arthritis randomised controlled trial (RCT) of intramuscular steroid treatment versus placebo were available.[17] Furthermore, the authors did not differentiate patients by disease activity, functional ability or disease evolution.[16]

A recent review of the use of generic utility measures for RA recommended more head-to-head

Table 1 Overview of the instrument properties of the EQ-5D and the SF-6D

	Domains (levels for each domain (n))	Questions (n)	Possible health states (n)	Valuation technique	Score range
EQ-5D	Mobility (3) Usual activities (3) Self-care (3) Pain/discomfort (3) Anxiety/depression (3)	5	243	TTO	−0.59 to 1.00
SF-6D	Physical function (6) Role limitation (4) Social function (5) Pain (6) Mental health (5) Vitality (5)	11 from the SF-36*	18 000	Standard gamble	0.296 to 1.00

*The SF-6D can also be calculated using the 12-Item Short Form Health Survey.[24]
TTO, time tradeoff; SF-36, 36-Item Short Form Health Survey.

Table 2 Characteristics of patients with early arthritis included in the ESPOIR cohort at baseline (n=813)

Characteristic	
Age (years)	48.1±12.6
Female sex, n (%)	624 (76.7)
Years of education, n (%)	
<6	101 (12.4)
6–12	457 (56.2)
>12	255 (31.4)
Employed, n (%)	481 (59.2)
Married or living together, n (%)	594 (73)
DAS28*	5.11±1.31
HAQ score	0.98±0.68
Erythrocyte sedimentation rate	29.4±24.6
C-reactive protein level†	20.3±32.4
Rheumatoid factor†, n (%)	372 (45.8)
Anti-CCP2 antibodies†, n (%)	315 (38.8)
Van-der-Heijde-modified Sharp score‡	5.97±10.14
EQ-5D score	0.518±0.306
SF-6D score	0.582±0.114

Data are expressed as mean±SD unless indicated otherwise.
*Among patients, 9.5% had a DAS28 of ≤3.2, 44.6% had a DAS28 of 3.2–5.1 and 45.9% had a DAS28 of >5.1.
†Baseline C-reactive protein level (normal <10 mg/L), IgM and IgA rheumatoid factors (ELISA, Menarini, France; positive >9 UI/ml) and anti-CCP2 antibodies (ELISA, DiaSorin, France; positive >50 U/ml) were detected in all patients by the same technique in a central laboratory (Paris-Bichat).
‡Of 715 patients, 160 (22.3%) had erosions on hands and/or feet at baseline.
ESPOIR, Étude et Suivi des Polyarthrites Indifférenciées Récentes; DAS28, Disease Activity Score for 28 joints; HAQ, Health Assessment Questionnaire.

comparisons of the measures in longitudinal studies across the spectrum of RA disease severity.[18]

We aimed to compare the responsiveness of the EQ-5D and the SF-6D in a large group of patients with EA followed over 2 years.

PATIENTS AND METHODS

Patients

Between December 2002 and March 2005, 813 patients with EA from 14 French regional centres were included in the Étude et Suivi des Polyarthrites Indifférenciées Récentes (ESPOIR) cohort.[19] Inclusion criteria were as follows: age of 18–70 years, more than two swollen joints for >6 weeks and <6 months, suspected or confirmed diagnosis of RA and not taking disease-modifying anti-rheumatic drugs or steroids (except if <2 weeks). Patients were followed every 6 months during the first 2 years. At baseline and on each visit, data on clinical and biological variables, including data for the Disease Activity Score for 28 joints (DAS28), a composite index of disease activity, were collected.[20] Furthermore, on each visit, patients completed self-administered patient-reported outcome measures, including a functional ability questionnaire, the Health Assessment Questionnaire (HAQ)[21] and HRQoL questionnaires, the EQ-5D and the 36-Item Short Form Health Survey.[22] The protocol for the ESPOIR cohort study was approved by the ethics committee of Montpellier, France. All patients gave their signed informed consent before inclusion in the study.

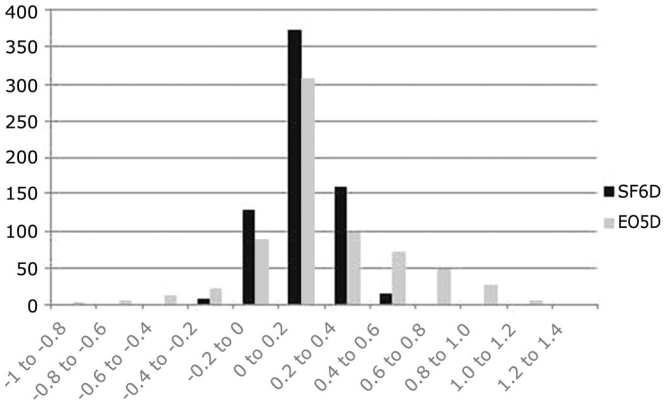

Figure 1 Frequency distribution of changes in SF-6D and EQ-5D utility scores over 2 years.

Utility measurement

Assessment of utility involves assigning a numeric value from 0 to 1 for health states, where 0 indicates death and 1 indicates a state of perfect health. The values reflect the preference for a health state in a situation of choice, which includes uncertainty or sacrifice (eg, life years). Methods such as standard gamble and time tradeoff (TTO) may be used to measure health states directly, but they are less suitable for clinical research and less widely used for feasibility reasons. Instead, indirect utility assessment techniques, namely the EQ-5D and the SF-6D, have been developed. Indirect health utility assessments involve population-assigned weights to calculate utility scores for particular health states from multidomain health status questionnaires completed by patients[23] (table 1).

Statistical analysis

EQ-5D and SF-6D utility scores were calculated using the scoring algorithms developed by Dolan[1] and Brazier *et al*,[2] respectively, using UK population weights. Descriptive statistics (mean±standard deviation, median IQR, minimum and maximum) and distributions of the changes in EQ-5D and SF-6D utility scores were computed. Ceiling and floor effects were compared and considered present if >15% of the respondents had the highest or lowest possible score.[25] To investigate whether the changes in the EQ-5D and the SF-6D are valid measures of change in EA health status, we used Spearman's product–moment correlation to compare the change scores for the two instruments with those for external measures of health (the HAQ and DAS28) from baseline to 6, 12 or 24 months.

Table 3 Correlations of change in outcome measures

	ΔSF-6D			ΔDAS28			ΔHAQ			ΔPhysical component of the SF-36			ΔMental component of the SF-36			ΔPain at rest		
Outcome	**Δ6**	**Δ12**	**Δ24**	**Δ6**	**Δ12**	**Δ24**	**Δ6**	**Δ12**	**Δ24**	**Δ6**	**Δ12**	**Δ24**	**Δ6**	**Δ12**	**Δ24**	**Δ6**	**Δ12**	**Δ24**
ΔSF-6D	–	–	–	−0.41	−0.47	−0.47	−0.55	−0.56	−0.59	0.57	0.60	0.64	0.60	0.62	0.69	−0.25	−0.29	−0.30
ΔEQ-5D	0.49	0.57	0.55	−0.37	−0.41	−0.40	−0.52	−0.53	−0.55	0.41	0.50	0.51	0.37	0.37	0.41	−0.23	−0.24	−0.29

Spearman's product–moment correlation; Δ, change; Δ6, change between month 0 and month 6; Δ12, change between month 0 and month 12; Δ24, change between month 0 and month 24.
DAS28, Disease Activity Score for 28 joints; HAQ, Health Assessment Questionnaire; SF-36, 36-Item Short Form Health Survey.

Table 4 Mean change and responsiveness at 1 year of the EQ-5D and the SF-6D for subgroups categorised by disease activity and functional disability at baseline.

	Responsiveness by disease activity					Responsiveness by functional disability			
DAS28	**Utility**	**Mean change (SD)**	**SRM* (95% CI)**	**ES* (95% CI)**	**HAQ**	**Utility**	**Mean change (SD)**	**SRM* (95% CI)**	**ES* (95% CI)**
≤3.2	SF-6D	0.061 (0.120)	0.52 (0.50 to 0.55)	0.63 (0.59 to 0.66)	≤1	SF-6D	0.072 (0.124)	0.59 (0.58 to 0.60)	0.73 (0.71 to 0.74)
	EQ-5D	0.079 (0.190)	0.43 (0.40 to 0.46)	0.42 (0.39 to 0.44)		EQ-5D	0.059 (0.231)	0.27 (0.26 to 0.28)	0.29 (0.28 to 0.30)
3.2–5.1	SF-6D	0.073 (0.118)	0.61 (0.60 to 0.63)	0.68 (0.67 to 0.69)	1–2	SF-6D	0.119 (0.130)	0.90 (0.89 to 0.91)	1.39 (1.37 to 1.40)
	EQ-5D	0.086 (0.243)	0.35 (0.34 to 0.36)	0.36 (0.35 to 0.37)		EQ-5D	0.243 (0.342)	0.71 (0.70 to 0.72)	0.80 (0.79 to 0.82)
>5.1	SF-6D	0.121 (0.135)	0.88 (0.87 to 0.90)	1.15 (1.14 to 1.17)	>2	SF-6D	0.176 (0.114)	1.62 (1.57 to 1.67)	1.95 (1.91 to 2.00)
	EQ-5D	0.230 (0.369)	0.62 (0.61 to 0.63)	0.70 (0.69 to 0.71)		EQ-5D	0.507 (0.403)	1.32 (1.29 to 1.36)	1.84 (1.78 to 1.91)

*SRM, ES and 95% CI values were obtained by bootstrap methods.
DAS28, Disease Activity Score for 28 joints; HAQ, Health Assessment Questionnaire; SRM, standardised response mean; ES, effect size.

Responsiveness

The responsiveness of the EQ-5D and the SF-6D was assessed by use of effect size (ES) and the standardised response mean (SRM) at 12 months for the entire sample and for subgroups categorised by disease activity at baseline (DAS28 ≤3.2, 3.2–5.1 and >5.1), by functional ability at baseline (HAQ ≤1, 1–2 and >2) and by disease activity evolution (DAS28 change <0 (disease activity improvement) or >0 (disease activity deterioration)). Both ES and SRM provide a ratio of signal (mean change) to noise (SD). ES is calculated as the ratio of the mean change between baseline and follow-up assessment to the SD of the group at baseline.[26] SRM is calculated as the ratio of the mean change between baseline and follow-up assessment to the SD of the mean change.[27] By convention, an ES of <0.2 is usually considered trivial, an ES of 0.2–0.5 is considered small, an ES of 0.5–0.8 is considered moderate, an ES of 0.8–1.2 is considered important and an ES of >1.2 is considered very important.[28] SRM values of >0.8 are considered large. Sensitivity to change in the HAQ was calculated as a benchmark. Because the standard errors of distribution-based approaches are not defined, we used bootstrap methods to estimate the 95% confidence intervals for the SRM and ES values to allow a comparison between the SF-6D and the EQ-5D.[29]

In assessing the robustness of the main conclusions, sensitivity analyses involved assessing responsiveness (1) in patients with active disease and those with a high probability of RA—candidates for new treatments for which cost-effectiveness studies are necessary (ie, patients fulfilling the new American College of Rheumatology–European League Against Rheumatism (ACR-EULAR) criteria for RA[30] and with at least moderate disease activity (DAS28 >3.2)), and (2) in patients at 6 and 24 months to confirm sensitivity to change over time (using ES and SRM between baseline and month 6, and between baseline and month 24, respectively).

All analyses involved the use of SAS v9.1 (SAS Institute, Cary, North Carolina, USA). p<0.05 was considered statistically significant.

RESULTS

Characteristics of the population

Table 2 shows the demographic and clinical characteristics of the 813 patients in the ESPOIR cohort at inclusion. In total, 641 (78.8%) patients fulfilled the new ACR-EULAR criteria for RA. The results are for the patients for whom data were available. At 2 years, 692 patients were still being followed up, and all demographic characteristics were similar to those of the initial population.

Utility scores

The distribution of utility scores was bimodal for the EQ-5D and near normal for the SF-6D (online supplementary figure S1). At baseline, the mean utility score for the EQ-5D was 0.518±0.306 (median 0.656 (IQR 0.255–0.725)) and for the SF-6D was 0.582±0.114 (median 0.580 (IQR 0.519–0.646)). The EQ-5D generated a minimum value of −0.594 and a maximum value of 1.0, with 11.8% of patients in health states considered worse than dead (<0) and with 1.5% of patients with a corresponding utility score of 1.0. In contrast, the SF-6D generated a minimum value of 0.301 and a maximum value of 0.923. Thus, no significant floor or ceiling effect was found at baseline. However, at 6 months, 6% of patients had an EQ-5D utility score of 1.0; this proportion increased at 1 year then remained stable over time at ~12%. The proportion of patients with an SF-6D utility score of 1.0 remained low at between 0.5% and 0.7%.

The distribution of utility changes was near normal for the EQ-5D and the SF-6D (figure 1). The correlation of change in the EQ-5D with change in the SF-6D at 12 months was moderate (r=−0.57) (table 3). The correlations of utility change, using the two instruments, with changes in DAS28 and HAQ scores were moderate, similar for the SF-6D and the EQ-5D, and stable over 2 years. The correlations of utility change, using the two instruments, with change in pain at rest were weak. The results were similar at 6 and 24 months (table 3).

Eleven patients experienced a change in the EQ-5D of more than one unit at 12 months. As expected, these patients had a very severe and active disease at baseline (mean DAS28 6.67±1.25;

mean HAQ 2.10±0.53; mean EQ-5D −0.134±0.128; mean SF-6D 0.458±0.088) and showed great improvement at 12 months (change in DAS28 −4.37±0.96; change in HAQ −1.95±0.58). It should be noted that all of these patients had baseline EQ-5D utilities valued at less than 0 or as worse than dead.

Responsiveness

For the entire sample at 12 months, the SF-6D was more sensitive to change than was the EQ-5D: SRM 0.74 (95% CI 0.73 to 0.75) versus 0.49 (95% CI 0.48 to 0.50). For subgroups categorised by disease activity (DAS28 ≤3.2, 3.2–5.1 and >5.1) or by functional ability (HAQ ≤1, 1–2 and >2) at baseline, the SF-6D was always more responsive than the EQ-5D (table 4).

More patients had improved condition than deteriorated condition (table 5).

The EQ-5D produced larger absolute mean change estimates with greater variance than did the SF-6D, whatever the direction of change (table 5).

At 12 months, the SF-6D showed greater sensitivity to change than the EQ-5D with improved condition: SRM 0.83 (0.82 to 0.84) versus 0.57 (0.56 to 0.58). The EQ-5D was more sensitive to change than was the SF-6D for deteriorated condition: SRM −0.20 (−0.23 to −0.18) versus −0.11 (−0.14 to −0.08) (table 5).

For patients with improved condition, the sensitivity to change in the SF-6D was similar to that in the HAQ, whereas for patients with deteriorated condition, the sensitivity to change in the EQ-5D was close to that in the HAQ (table 5).

The results were similar at 6 and 24 months, except for patients with deteriorated condition at 6 months: the EQ-5D and the SF-6D did not detect deteriorated disease activity (DAS28 mean change 0.48±0.44) and showed a very small gain in utility. The results were similar for patients fulfilling the new ACR-EULAR criteria for RA and patients with at least moderate disease activity at baseline (data not shown).

DISCUSSION

This study is the first to compare the responsiveness of the EQ-5D and the SF-6D in a large cohort of patients with EA. The responsiveness of the EQ-5D and the SF-6D differed according to the direction of change in EA condition. The SF-6D was more responsive to improved condition, and the EQ-5D was more responsive to deteriorated condition. The high mean change in the EQ-5D relative to the SF-6D has consequences for cost-effectiveness analyses.

One of the strengths of our study is its investigation of a large diverse group of patients with EA. The observational ESPOIR cohort aimed to include all patients with EA regardless of disease level, age and sex, and our study implied the performance of the instruments in a real-life setting. The study also included a large number of patients with longitudinal assessment and individual data allowing for the differentiation of patients with improved and deteriorated conditions. Moreover, we could investigate patients with active disease and those with a high probability of RA—candidates for new treatments for which cost-effectiveness studies are necessary.

Our study has some limitations. We did not compare the test–retest reliability of these two instruments. The scoring algorithms used for both instruments were developed from the general population in the UK because no such algorithm was available in France at the time of the study.

However, the use of an algorithm from the same population for the EQ-5D and the SF-6D might result in a more valid comparison. Valuations of profiles (USA, Denmark, Spain and Japan)

Table 5: Mean change and responsiveness of the EQ-5D and the SF-6D according to the evolution of patients' condition at 6, 12 and 24 months

Disease activity improvement	Month 6 n= 622				Month 12 n=646				Month 24 n= 612			
	Mean change	SD of change	SRM* [95% CI]	ES* [95% CI]	Mean change	SD of change	SRM* [95% CI]	ES* [95% CI]	Mean change	SD of change	SRM* [95% CI]	ES* [95% CI]
DAS28 #	-2.07	1.29			-2.20	1.32			-2.47	1.44		
HAQ	-0.494	0.636	-0.77 [-0.78;-0.76]	-0.72 [-0.73;-0.71]	-0.508	0.652	-0.79 [-0.80;-0.78]	-0.75 [-0.76;-0.74]	-0.561	0.680	-0.82 [-0.83;-0.81]	-0.81 [-0.82;-0.80]
SF-6D	0.098	0.121	0.81 [0.81;0.82]	0.86 [0.85;0.87]	0.105	0.127	0.83 [0.82;0.84]	0.92 [0.91;0.93]	0.121	0.136	0.89 [0.88;0.90]	1.06 [1.05;1.07]
EQ-5D	0.154	0.301	0.50 [0.50;0.51]	0.50 [0.49;0.51]	0.173	0.313	0.57 [0.56;0.58]	0.58 [0.57;0.59]	0.189	0.313	0.60 [0.59;0.61]	0.61 [0.60;0.62]
Disease activity deterioration	**Month 6 n=97**				**Month 12 n=61**				**Month 24 n=53**			
	Mean change	SD of change	SRM* [95% CI]	ES* [95% CI]	Mean change	SD of change	SRM* [95% CI]	ES* [95% CI]	Mean change	SD of change	SRM* [95% CI]	ES* [95% CI]
DAS28#	0.48	0.44			0.68	0.55			0.95	0.86		
HAQ	0.08	0.42	0.21 [0.19;0.23]	0.15 [0.14;0.16]	0.129	0.518	0.24 [0.22;0.26]	0.23 [0.21;0.26]	0.316	0.587	0.53 [0.51;0.55]	0.63 [0.59;0.67]
SF-6D	0.015	0.098	0.18 [0.16;0.20]	0.19 [0.17;0.21]	-0.008	0.097	-0.11 [-0.14;-0.08]	-0.10 [-0.13;-0.08]	-0.005	0.109	-0.01 [-0.04;0.02]	-0.05 [-0.08;-0.01]
EQ-5D	0.018	0.264	0.06 [0.03;0.08]	0.07 [0.05;0.09]	-0.053	0.259	-0.20 [-0.23;-0.18]	-0.20 [-0.23;-0.17]	-0.034	0.289	-0.11 [-0.14;-0.07]	-0.13 [-0.16;-0.10]

* SRM, ES and 95% CI were obtained by bootstrap methods.
# Mean change and SD of change of DAS28 were indicated but SRM and ES were not calculated, as DAS28 was the determinant of disease activity improvement and deterioration.
SRM, standardized response mean; ES, effect size; DAS28, Disease Activity Score for 28 joints; HAQ, Health Assessment Questionnaire.

may vary according to how sets of preference weights for health profiles were developed. We chose to use British estimates since a UK study—the PRINCEPS study, which had the largest sample size overall and in the European population—provided the most robust estimates and is widely used in Europeans countries where national references are missing/lacking. Our results were similar to those found for more heterogeneous RA.[14 18] All previous studies have suggested that both measures are generally responsive to changed condition in patients with RA, consistent with our results.[13–15 18] Concerning EA, Harrison *et al*[16] showed that the SF-6D was more responsive than the EQ-5D for 181 patients included in the Steroids in Very Early Arthritis RCT of intramuscular steroid treatment versus placebo over 1 year: SRM 0.83 versus 0.64. However, the authors did not differentiate patients by disease activity, functional ability or disease evolution. Moreover, they did not provide 95% CI to allow comparisons between the different SRMs.

Several reasons might explain this difference between utility scores, and concerns have been voiced about the ability of the EQ-5D to measure change owing to its bimodal distribution, its three answer modalities and possible ceiling effects within domains.[4 14 15 18] Five-level answer modalities for the EQ-5D have been developed but are not yet widely used.[31] One explanation for the difference in utility scores is that the health descriptive system of the SF-6D does not allow for negative values and thus assigns a 0.296 value to the most severe health state produced by the descriptive system, whereas the EQ-5D score allows for negative scores.[32 33] Second, EQ-5D utility scores are TTO based, tending to result in high values for mild states, whereas SF-6D scores are standard gamble based, tending to result in high values for severe states.[34 35] For the TTO valuation procedure, each state is considered to last 10 years without change,[1] which may explain in part the low valuations given to the most severe health states. The low EQ-5D utility scores for patients with severe disabilities may also be explained by the content of the EQ-5D. Of the five dimensions, four (mobility, self-care, usual activity and pain/discomfort) are likely to be particularly affected in patients with EA. Finally, the scoring algorithm subtracts a value of 0.269 for a score of 3 (ie, severe problems) in at least one dimension.

The SF-6D may be more sensitive (because of its larger descriptive system) than the EQ-5D for patients experiencing mild to moderate health problems.[2] With only three response levels for each dimension, the EQ-5D does not allow for gradations between no problem, moderate problem, and severe problem.

The correlations of utility change, using the two instruments, with change in the DAS28 and HAQ scores were moderate, similar for the SF-6D and the EQ-5D, and stable over 2 years. Moderate correlations suggest that both tools measure a similar underlying construct and show similar longitudinal validity. Utilities measure patients' perspective on their health state using population-assigned weights, while the HAQ assesses patients' report of their functioning and DAS28 only reflects the activity. Although these two concepts are related, they are not identical.

Responsiveness indices give an indication of whether a measure can detect a true change over noise but does not indicate whether the detected change is meaningful or useful.[36]

As an outcome measure in an epidemiological setting, the SF-6D requires a smaller sample size than the EQ-5D to detect improvements in patients whose health is improving. The opposite is true for patients with worsening condition. However, in economic analyses, the approach is different. The larger mean change in patients with improving and deteriorating conditions suggests that an intervention will more likely be considered cost-effective if assessed by the EQ-5D rather than by the SF-6D. In cost–utility analysis, the incremental cost of an intervention in a group is divided by its incremental effectiveness in this group, measured by an instrument such as the EQ-5D or the SF-6D. The gain in the EQ-5D will be larger and will provide more favourable incremental cost–utility ratios than the SF-6D. A recent study of RA reported that the change estimated with the EQ-5D resulted in a cost per QALY that was more than 50% lower than the cost per QALY calculated with the SF-6D.[37] However, because the mean effect is only an estimate, the uncertainty around the estimate must be presented. The smaller variance of the SF-6D should result in less uncertainty in estimating the relative cost-effectiveness of two treatments.

Our study cannot answer the question of which index score is more valid as a measure of preference-based HRQoL. However, we can provide tentative suggestions, with the caveat that the choice of instrument should be guided by the appropriateness of the tool for its intended use. We have demonstrated three important elements. First, the SF-6D was more responsive than the EQ-5D for improved EA condition. Second, the smaller variance of the SF-6D should result in less uncertainty in estimating the relative cost-effectiveness of two treatments. Third, the SF-6D was more conservative than the EQ-5D, providing more optimistic cost-effectiveness ratios than the SF-6D. Therefore, the SF-6D may be more appropriate for use in RCTs of biological treatments for patients with EA.

Contributors CGV, RMF, PF and PB collected and analysed the data. CGV, ACR, BF and FG conceived the study and participated in its design and coordination. CGV, ACR and BF drafted the manuscript. All authors helped to revise the manuscript. All authors read and approved the final manuscript.

Acknowledgements The authors thank Nathalie Rincheval for expert monitoring and data management, and all the investigators who recruited and followed patients (F Berenbaum, Paris-Saint Antoine; MC Boissier, Paris-Bobigny; A Cantagrel, Toulouse; B Combe, Montpellier; M Dougados, Paris-Cochin; P Fardellone, Amiens; B Fautrel and P Bourgeois, Paris-La Pitié; RM Flipo, Lille; Ph Goupille, Tours; F Liote, Paris-Lariboisière; X Le Loet, Rouen; X Mariette, Paris-Bicêtre; O Meyer, Paris-Bichat; A Saraux, Brest; Th Schaeverbeke, Bordeaux; J Sibilia, Strasbourg).

Competing interests None.

Ethics approval The protocol for the ESPOIR cohort study was approved by the ethics committee of Montpellier, France. All patients gave their signed informed consent before inclusion in the study.

Provenance and peer review Not commissioned; externally peer reviewed.

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. **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.
4. **Marra CA**, Esdaile JM, Guh D, *et al.* A comparison of four indirect methods of assessing utility values in rheumatoid arthritis. *Med Care* 2004;**42**:1125–31.
5. **Marra CA**, Marion SA, Guh DP, *et al.* Not all "quality-adjusted life years" are equal. *J Clin Epidemiol* 2007;**60**:616–24.
6. **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.
7. **Lillegraven S**, Kristiansen IS, Kvien TK. Comparison of utility measures and their relationship with other health status measures in 1041 patients with rheumatoid arthritis. *Ann Rheum Dis* 2010;**69**:1762–7.
8. **Longworth L**, Bryan S. An empirical comparison of EQ-5D and SF-6D in liver transplant patients. *Health Econ* 2003;**12**:1061–7.
9. **Xie F**, Li SC, Luo N, *et al.* Comparison of the EuroQol and short form 6D in Singapore multiethnic Asian knee osteoarthritis patients scheduled for total knee replacement. *Arthritis Rheum* 2007;**57**:1043–9.
10. **Barton GR**, Sach TH, Avery AJ, *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.

11. **Gaujoux-Viala C**, Rat AC, Guillemin F, *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.
12. **Hurst NP**, Kind P, Ruta D, *et al.* Measuring health-related quality of life in rheumatoid arthritis: validity, responsiveness and reliability of EuroQol (EQ-5D). *Br J Rheumatol* 1997;**36**:551–9.
13. **Marra CA**, Rashidi AA, Guh D, *et al.* Are indirect utility measures reliable and responsive in rheumatoid arthritis patients? *Qual Life Res* 2005;**14**:1333–44.
14. **Russell AS**, Conner-Spady B, Mintz A, *et al.* The responsiveness of generic health status measures as assessed in patients with rheumatoid arthritis receiving infliximab. *J Rheumatol* 2003;**30**:941–7.
15. **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.
16. **Harrison MJ**, Davies LM, Bansback NJ, *et al*. The comparative responsiveness of the EQ-5D and SF-6D to change in patients with inflammatory arthritis. *Qual Life Res* 2009;**18**:1195–205.
17. **Verstappen SM**, McCoy MJ, Roberts C, *et al*. Beneficial effects of a 3-week course of intramuscular glucocorticoid injections in patients with very early inflammatory polyarthritis: results of the STIVEA trial. *Ann Rheum Dis* 2010;**69**:503–9.
18. **Harrison MJ**, Davies LM, Bansback NJ, *et al.* The validity and responsiveness of generic utility measures in rheumatoid arthritis: a review. *J Rheumatol* 2008;**35**:592–602.
19. **Combe B**, Benessiano J, Berenbaum F, *et al.* The ESPOIR cohort: a ten-year follow-up of early arthritis in France: methodology and baseline characteristics of the 813 included patients. *Joint Bone Spine* 2007;**74**:440–5.
20. **Prevoo ML**, van't Hof MA, Kuper HH, *et al.* Modified disease activity scores that include twenty-eight-joint counts. Development and validation in a prospective longitudinal study of patients with rheumatoid arthritis. *Arthritis Rheum* 1995;**38**:44–8.
21. **Fries JF**, Spitz P, Kraines RG, *et al.* Measurement of patient outcome in arthritis. *Arthritis Rheum* 1980;**23**:137–45.
22. **Ware JE** Jr, Sherbourne CD. The MOS 36-item Short-Form Health Survey (SF-36): I. Conceptual framework and item selection. *Med Care* 1992;**30**:473–83.
23. **Khanna D**, Tsevat J. Health-related quality of life—an introduction. *Am J Manag Care* 2007;**13** Suppl 9:S218–23.
24. **Brazier JE**, Roberts J. The estimation of a preference-based measure of health from the SF-12. *Med Care* 2004;**42**:851–9.
25. **Veenhof C**, Bijlsma JW, van den Ende CH, *et al.* Psychometric evaluation of osteoarthritis questionnaires: a systematic review of the literature. *Arthritis Rheum* 2006;**55**:480–92.
26. **Lipsey MW**, Wilson DB. Practical meta-analysis. Applied social research methods series. London: Sage, 2001;**49**:34–71.
27. **Norman GR**, Wyrwich KW, Patrick DL. The mathematical relationship among different forms of responsiveness coefficients. *Qual Life Res* 2007;**16**:815–22.
28. **Cohen J**. A power primer. *Psychol Bull* 1992;**112**:155–9.
29. **Chang E**, Abrahamowicz M, Ferland D, *et al*. Comparison of the responsiveness of lupus disease activity measures to changes in systemic lupus erythematosus activity relevant to patients and physicians. *J Clin Epidemiol* 2002;**55**:488–97.
30. **Aletaha D**, Neogi T, Silman AJ, *et al.* 2010 rheumatoid arthritis classification criteria: an American College of Rheumatology/European League Against Rheumatism collaborative initiative. *Ann Rheum Dis* 2010;**69**:1580–8.
31. **Janssen MF**, Birnie E, Haagsma JA, *et al.* Comparing the standard EQ-5D three-level system with a five-level version. *Value Health* 2008;**11**:275–84.
32. **Brazier J**, Roberts J, Tsuchiya A, *et al.* A comparison of the EQ-5D and SF-6D across seven patient groups. *Health Econ* 2004;**13**:873–84.
33. **Adams R**, Walsh C, Veale D, *et al.* Understanding the relationship between the EQ-5D, SF-6D, HAQ and disease activity in inflammatory arthritis. *Pharmacoeconomics* 2010;**28**:477–87.
34. **Dolan P**, Gudex C, Kind P, *et al.* Valuing health states: a comparison of methods. *J Health Econ* 1996;**15**:209–31.
35. **Tsuchiya A**, Brazier J, Roberts J. Comparison of valuation methods used to generate the EQ-5D and the SF-6D value sets. *J Health Econ* 2006;**25**:334–46.
36. **Crosby RD**, Kolotkin RL, Williams GR. Defining clinically meaningful change in health-related quality of life. *J Clin Epidemiol* 2003;**56**:395–407.
37. **Brennan A**, Bansback N, Nixon R, *et al.* Modelling the cost effectiveness of TNF-alpha antagonists in the management of rheumatoid arthritis: results from the British Society for Rheumatology Biologics Registry. *Rheumatology (Oxford)* 2007;**46**:1345–54.

Corrections

Gaujoux-Viala C, Rat A-C, Guillemin F, *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. doi:10.1136/annrheumdis-2011-200891.

The following statement should have been included in this article:

Grant support An unrestricted grant from Merck Sharp and Dohme (MSD) was allocated for the first 5 years of the ESPOIR cohort study. Two additional grants from INSERM supported part of the biological database. The French Society of Rheumatology, Abbott, Amgen and Wyeth also supported the ESPOIR cohort study.

Ann Rheum Dis 2012;**71**:1756. doi:10.1136/annrheumdis-2011-200891corr1

Responsiveness of EQ-5D and SF-6D in patients with early arthritis: results from the ESPOIR cohort

Cécile Gaujoux-Viala, Anne-Christine Rat, Francis Guillemin, et al.

Ann Rheum Dis 2012 71: 1478-1483 originally published online March 14, 2012
doi: 10.1136/annrheumdis-2011-200891