

## The comparative responsiveness of the EQ-5D and SF-6D to change in patients with inflammatory arthritis

M. J. Harrison · L. M. Davies · N. J. Bansback · M. J. McCoy · S. M. M. Verstappen · K. Watson · D. P. M. Symmons · The British Society for Rheumatology Biologics Register Control Centre Consortium, on behalf of the BSRBR

Accepted: 8 September 2009 / Published online: 24 September 2009  
© The Author(s) 2009. This article is published with open access at Springerlink.com

### Abstract

**Purpose** Comparative evidence regarding the responsiveness of the EQ-5D and SF-6D in arthritis patients is conflicting and insufficient across the range of disease severity. We examined the comparative responsiveness of the EQ-5D and SF-6D in cohorts of patients with early inflammatory disease through to severe rheumatoid arthritis (RA).

**Methods** Responsiveness was tested using the effect size (ES) and standardised response mean (SRM). Correlation of change in EQ-5D and SF-6D with disease specific measures was tested using Pearson correlations and the Steiger's Z test. Treatment response and self-reported change were used as anchors of important change.

**Results** The EQ-5D was more responsive to deterioration (ES ratio (EQ-5D/SF-6D): 1.6–3.0) and the SF-6D more responsive to improvement (ES ratio (SF-6D/EQ-5D): 1.1–1.8) in health. The SF-6D did not respond well to deterioration in patients with established severe RA (ES and SRM 0.08). The EQ-5D provided larger absolute mean change estimates but with greater variance compared to the SF-6D.

**Conclusions** The comparative responsiveness of the EQ-5D and SF-6D differs according to the direction of change. The level of mean change of the EQ-5D relative to the SF-6D has implications for cost-effectiveness analysis. Use of the SF-6D in patients with severe progressive disease may be inappropriate.

The members of the BSRBR Control Centre Consortium are detailed in the Appendix.

M. J. Harrison · M. J. McCoy · S. M. M. Verstappen · K. Watson · D. P. M. Symmons (✉)  
ARC Epidemiology Unit, Stopford Building, The University of Manchester, Oxford Road, Manchester M13 9PT, UK  
e-mail: [deborah.symmons@manchester.ac.uk](mailto:deborah.symmons@manchester.ac.uk)

M. J. Harrison  
e-mail: [mark.harrison@manchester.ac.uk](mailto:mark.harrison@manchester.ac.uk)

L. M. Davies  
Health Economics Research at Manchester, Health Methodology Research Group, School of Community Based Medicine, University of Manchester, 1st Floor, University Place Oxford Road, Manchester M13 9PL, UK

N. J. Bansback  
Centre for Health Evaluation and Outcome Sciences, St Paul's Hospital, 620B - 1081, Burrard Street, Vancouver, BC V6Z 1Y6, Canada

**Present Address:**  
M. J. McCoy  
The University of Western Australia, Perth, Australia

**Keywords** Quality of life · Outcome assessment (health care) · Reproducibility of results · Arthritis rheumatoid

### Abbreviations

EQ-5D	EuroQol-5D
SF-6D	Short form-6D
RA	Rheumatoid arthritis
ES	Effect size
SRM	Standardised response mean
QALYs	Quality-adjusted life years
NICE	National Institute for Health and Clinical Excellence
SF-36	Short form 36-item health survey
STIVEA	Steroids in very early arthritis
BROSG	British Rheumatoid Outcome Study Group
RCT	Randomised controlled trial
BSRBR	British Society for Rheumatology Biologics Register
HAQ	Health Assessment Questionnaire

DAS28	Disease activity score based on 28 swollen joint counts
ESR	Erythrocyte sedimentation rate
HRQoL	Health-related quality of life
TNF $\alpha$	Tumour necrosis factor alpha
EULAR	European League Against Rheumatism
DMARD	Disease modifying anti-rheumatic drug
VAS	Visual analogue scales
EQ-VAS	EuroQol visual analogue scale
ICER	Incremental cost effectiveness ratio
s.d.	Standard deviation

## Background

There is an increasing demand for economic evaluations of health care which compare the costs and benefits of interventions in order to identify which provide the greatest health gain per unit of investment. Assessments based on quality-adjusted life years (QALYs) gained are recommended for economic evaluation of new interventions [1] and have been adopted by decision-making organisations in many countries including the United Kingdom [2], Canada [3] and the USA [4]. QALYs are the product of the time spent in a health state multiplied by a utility value, representing quality of life, for that particular health state. Utility is the preference for a health state (rated in the presence of choice) relative to full health (scored 1) and death (scored 0).

Decision makers such as the National Institute for Health and Clinical Excellence (NICE) in the United Kingdom aim to maximise the health of the population. They, therefore, require preference data from a representative sample of the public based on ratings of health states described using standardised validated generic instruments using a choice-based method [5]. This allows interventions for a range of different diseases and specialties to be assessed on a standard scale. Generic instruments developed for this purpose include the EQ-5D and SF-6D. The EQ-5D has a number of country-specific choice-based preference weights including the United Kingdom and the USA, whilst the SF-6D to date has only UK preference weights.

In the United Kingdom, NICE currently suggests that the most appropriate measure is the EQ-5D but recognises that the EQ-5D may not be appropriate in all circumstances. The SF-6D has considerable potential as it can be calculated from both SF-36 [6] and SF-12 [7], which have been routinely collected in numerous studies. The choice of preference-based measure, however, depends on the validity of the measure in that setting. The EQ-5D is one the most extensively validated measures for use in patients with

rheumatoid arthritis (RA) [8]. The SF-6D has been less extensively studied in this setting, but evidence to date suggests the measure has potential [8].

One important test of validity is the ability of a measure to reflect the change in patients over time. The EQ-5D and SF-6D have been shown to be capable of detecting some degree of change in RA patients [9–12]. The responsiveness of the EQ-5D and SF-6D have been compared head-to-head in North American populations, but not to date in UK or European populations. In two studies of North American populations, the SF-6D appeared more responsive than the EQ-5D to improvement in patients health [10, 11]. However, other results have been conflicting. In patients with one of a number of rheumatological conditions (51% RA), the EQ-5D was more responsive than the SF-6D to improvement [12]. A recent review of the use of generic utility measures in RA recommended more head-to-head comparisons of the measures in longitudinal studies across the spectrum of RA disease severity [8].

We aimed to compare the responsiveness to change of the EQ-5D and SF-6D in UK patients from a range of studies covering early inflammatory arthritis through to severe RA.

## Methods

Data were taken from four cohorts of patients:

1. The Steroids in Very Early Arthritis (STIVEA) randomised controlled trial (RCT) of intramuscular steroid treatment versus placebo in patients with very early inflammatory arthritis (4–11 weeks duration). The trial follow-up finished in late 2007 [13]. At the time of this analysis, the STIVEA trial remained blinded. Therefore, the patients studied comprised patients receiving either active or placebo treatment, but the proportion receiving each allocation is unknown.
2. British Rheumatoid Outcome Study Group (BROSG) RCT of aggressive versus symptomatic control of inflammation in patients with established (>5 years duration) stable, symptomatic rheumatoid arthritis (RA) followed for 3 years. The BROSG trial was conducted between 1998 and 2001 [14].
3. A subsample from the British Society for Rheumatology Biologics Register (BSRBR) of RA patients treated with anti-TNF therapy and followed for 6 months. The BSRBR was established in October 2001, and the methods of this study have been described in detail previously [15]. As part of the current study, from 1st August 2006 to 31st December 2007, newly enrolled patients were also asked to complete the EQ-5D at baseline and the 6-month assessment.

4. A subsample of patients in the control arm of BSRBR, who also received the EQ-5D at baseline and 6-month assessment in the same time period as the anti-TNF treatment cohort. These patients were biologic-naive with active RA (guideline DAS28 >4.2) currently treated with Disease Modifying Anti-Rheumatic Drugs (DMARDs) and were recruited in parallel within the BSRBR and followed up with identical methodology.

Baseline data for all cohorts included age, sex and disease duration. All patients completed the EQ-5D [16] and the SF-36 [17] (used to calculate the SF-6D utility measure [6]) and the Health Assessment Questionnaire (HAQ), a measure of functional disability. A patient global assessment, the 28 tender and swollen joint counts and the erythrocyte sedimentation rate (ESR) were also collected, which enabled the Disease Activity Score (DAS28) [18] to be calculated. In the BSRBR control arm, the composite DAS28 score was frequently reported in isolation without separate 28 tender and swollen joint counts and the ESR. Higher HAQ (range 0–3), DAS28 (range 0–10), tender and swollen joint counts (range 0–28) and ESR denote more severe disease (Table 1). Lower EQ-5D and SF-6D scores denote poorer HRQoL.

#### Expectations of improvement/deterioration

These four cohorts reflect a range of arthritis states/severity found in routine practice from first presentation with undifferentiated inflammatory arthritis through long-standing established RA to patients with severe, active disease. The health-related quality of life (HRQoL) of patients with early disease (STIVEA trial) and active disease patients BSRBR anti-TNF treatment arm was expected to improve.

Patients from STIVEA may improve in response to steroid treatment, by natural remission which may be expected to occur in up to 25% of patients [19, 20], or by adaptation—improvements in functional disability are often seen in the early stages of RA [21, 22]. Patients in the BSRBR receiving treatments which inhibit the action of TNF $\alpha$  were expected to have dramatically improved outcome [23–25]. Improvement in response to treatment was assessed according to the European League Against Rheumatism (EULAR) response criteria definition based on the DAS28 [26]. Responders were patients achieving a good or moderate EULAR response. Good responders improve by >1.2 units on the DAS28 score and achieve an absolute DAS28 score <3.2 at 6 months. Non-responders improve <0.6 and have a 6-month DAS28 score >5.1. Moderate responders fall in between these definitions.

Patients with long-standing established (BROSG) and severe disease (BSRBR control) receiving DMARD treatment were expected to experience disease progression which would be reflected by deteriorating HRQoL. Successful DMARD treatment is expected to slow disease progression [27], but patients with long disease duration are less likely to respond to DMARD treatment [28, 29].

Change in EQ-5D and SF-6D over the first year of the BROSG trial was assessed in relation to the EuroQol ‘feelings thermometer’ visual analogue scale (EQ-VAS). The EQ-VAS asks the respondent to indicate “*how good or bad is your health today, in your opinion*” on a vertical 0–100 scale. Change between baseline and 1-year follow-up was calculated as a percentage change using the formula  $((EQ-VAS_2 - EQ-VAS_1)/EQ-VAS_1) \times 100$ . The percentage change was then defined as small if it was between 20 and 50% similar to the methods of Marra et al. [10].

**Table 1** Summary of outcome measures used in this study

	Type of measure	Range of scores		Minimum important difference
		Worst	Best	
EQ-5D	Preference-based utility measure/HRQoL	–0.59	1.00	0.05–0.13 [8]
SF-6D	Preference-based utility measure/HRQoL	0.30	1.00	0.03–0.04 [8]
EQ-VAS	Global self-rated health assessment	0	100	9–29 for global health VAS scales [48]
HAQ	Functional disability	3	0	Generally 0.20–0.25 [48, 49]; –0.09 to 0.21 for improvement, 0.15–0.48 for deterioration [42, 48]
DAS28	Disease activity	9.1 <sup>a</sup>	0	0.6 based on EULAR response criteria [26]
28 Tender joint count	Physician assessment of tenderness in 28 joints	28	0	N/k
28 Swollen joint count	Physician assessment of swelling in 28 joints	28	0	N/k
ESR (mm/h)	Laboratory test of inflammatory marker/acute phase reactant	<sup>b</sup>	0	N/k

HRQoL health-related quality of life, ESR erythrocyte sedimentation rate, N/k not known

<sup>a</sup> Using an ESR value of 100; <sup>b</sup> higher values indicate inflammation

Hypotheses were formulated on the basis of expected improvement or deterioration, and the magnitude of expected change within these groups was estimated using ‘benchmark’ criteria for effect sizes of small ( $ES = 0.2$ ), moderate ( $ES = 0.5$ ) and large ( $ES = 0.8$ ) [30].

We expected

1. moderate EQ-5D and SF-6D improvements ( $ES \sim 0.5$ ) in STIVEA patients based on the known improvement in symptoms of patients in early arthritis (for example improvements in HAQ, pain and SF-36 HRQoL), steroid treatment which reduces inflammation and may slow the progression of disease, and the possibility of disease remission [22, 31, 32]. Only half of the STIVEA patients will have been randomised to steroid treatment as part of the trial.
2. moderate to large improvements in EQ-5D and SF-6D ( $ES > 0.5$ ) in BRSBR patients. In a US study of arthritis patients receiving infliximab, an anti-TNF $\alpha$  treatment, patients were shown to have moderate (EQ-5D 0.6) to large (SF-6D 1.4) ES [11].
3. small ( $ES \sim 0.2$ ) improvement or deterioration in patients reporting changes in health over 12 months of the BROSG trial. In a study using similar VAS defined and self-reported improvement and deterioration, ES for deterioration ranged from  $-0.24$  (EQ-5D) to  $-0.55$  (SF-6D) and improvement ranged from  $0.36$  (EQ-5D) to  $0.54$  (SF-6D) [10]. However, these estimates were based on groups with no limit for deterioration or improvement. Our definition of 20–50% improvement/deterioration was restrictive; therefore, it is likely the ES will be lower.
4. Small deterioration ( $ES \sim 0.2$ ) in the BROSG and BSRBR control groups. The progression and duration of arthritis is associated with small gradual increases in functional disability (approximately 0.033 per annum) and accumulated joint destruction [31]. The reduction in EQ-5D and SF-6D scores would be expected to mirror the gradual increase in burden of disease.

### Statistical analysis

We treated responsiveness as a part of the validation process of an outcome measure which requires longitudinal data and methods distinct from other techniques used to assess other types of validity [33]. We defined responsiveness using the effect size (ES) and standardised response mean (SRM) [34]. Both provide a ratio of signal (mean change) to noise (standard deviation). ES for this study was calculated using a formula based on Cohen’s  $d$ ,  $d = (\bar{x}_1 - \bar{x}_2/s)$ . The source of the standard deviation ( $s$ ) in

Cohen’s formula is not specified, as the true standard deviation is assumed to remain the same regardless of the mean of the population. The ES in this study used mean change between baseline and a follow-up assessment and the standard deviation of the group at baseline:  $ES = \mu_1 - \mu_2/\sigma\chi_1$ , where  $\mu_1$  = mean at follow-up,  $\mu_2$  = mean at baseline and  $\sigma\chi_1$  = standard deviation of the group mean at baseline. The SRM is calculated in the same way as the ES, although the standard deviation ( $s$ ) is the standard deviation of the mean change ( $\bar{x}_1 - \bar{x}_2$ ) instead of the baseline standard deviation:  $SRM = \mu_1 - \mu_2/\sigma(\chi_1 - \chi_2)$ , where  $\mu_1$  = mean at follow-up,  $\mu_2$  = mean at baseline, and  $\sigma(\chi_1 - \chi_2)$  = standard deviation of the group change between baseline and follow-up.

The Pearson product-moment correlation was used to calculate the correlation of change between measures. The comparative strength of correlations between a disease-specific outcome measure and the EQ-5D and SF-6D utility measures was compared using Steiger’s  $Z$  test for two correlated correlation coefficients [35, 36].

Floor and ceiling effects (the percentage of patients occupying the worst/best health states) were calculated and considered small if  $\leq 15\%$  of patients occupy the worst and best health states, respectively, and serious if  $> 15\%$  of patients occupy these states. These criteria have been used previously in reviews of outcome measures in musculoskeletal disease [8, 37].

## Results

### Baseline characteristics

The study population consisted of 466 patients from the BROSG trial, 182 patients from STIVEA, 223 patients from the BSRBR register and 188 from the BSRBR comparison cohort. One hundred and eighty-eight (84%) of the BSRBR patients received adalimumab and 35 (16%) received infliximab. The disease duration of patients ranged from 7.8 weeks (s.d. 2.6) in the STIVEA trial to 13.4 years (s.d. 11.5) in the BSRBR. There were differences in demographic and clinical characteristics between the four groups of patients (Table 2).

Patients in the BROSG study had the highest mean EQ-5D scores (mean 0.59, s.d. 0.22) followed by the BSRBR control arm (mean 0.55, s.d. 0.27), STIVEA (mean 0.46, s.d. 0.31) and the BSRBR group (mean 0.34, s.d. 0.33). The pattern was the same for SF-6D scores, but these scores were consistently higher than EQ-5D scores; mean scores ranged from 0.64 (s.d. 0.13) in BROSG to 0.50 (s.d. 0.09) in the BSRBR treatment arm.

**Table 2** Baseline characteristics of patients from the four cohorts, ordered by mean utility score

	BROSG <i>n</i> = 466	BSRBR control <i>n</i> = 188	STIVEA <i>n</i> = 182	BSRBR <i>n</i> = 223
Age (years)	60.6 (11.2)	60.2 (11.7)	55.5 (15.0)	57.9 (12.2)
Sex, female, <i>n</i> (%)	317 (68%)	142 (76%)	182 (69%)	168 (75%)
Disease duration (years)	12.5 (6.7)	11.8 (10.9)	0.15 (0.05)	13.4 (11.5)
EQ-5D	0.59 (0.22)	0.55 (0.27)	0.46 (0.31)	0.34 (0.33)
SF-6D	0.64 (0.13)	0.59 (0.12)	0.56 (0.13)	0.50 (0.09)
EQ-VAS	64.8 (18.1)	58.3 (18.7)	58.8 (21.3)	48.1 (21.1)
HAQ <sup>a</sup>	1.28 (0.70)	1.38 (0.75)	1.23 (0.69)	1.92 (0.60)
DAS28 <sup>a</sup>	4.04 (1.26)	4.99 (1.09)	5.39 (1.10)	6.41 (1.03)
28 Tender joint count <sup>a</sup>	5.1 (5.9)	7.6 (6.0)	10.6 (6.6)	15.3 (7.2)
28 Swollen joint count <sup>a</sup>	4.2 (4.2)	5.0 (4.5)	8.9 (5.1)	10.3 (6.4)
ESR (mm/h) <sup>a</sup>	23.8 (22.9)	35.1 (24.1)	36.6 (26.5)	42.7 (27.8)

Numbers are mean (s.d.) unless otherwise stated

EQ-5D EuroQol-5D, SF-6D short form-6D, EQ-VAS EuroQol visual analogue scale, STIVEA steroids in very early arthritis, BROSG British Rheumatoid Outcome Study Group, BSRBR British Society for Rheumatology Biologics Register, HAQ Health Assessment Questionnaire, DAS28 Disease Activity Score based on 28 swollen joint counts, ESR = erythrocyte sedimentation rate, s.d., standard deviation

<sup>a</sup> Higher values denote more severe disease

### Change over time

A EULAR response could be calculated for 161 BSRBR patients (72%) and 171 STIVEA patients (94%). One hundred and thirty-two (82%) of the BSRBR patients and 135 (79%) of the STIVEA patients were responders. Over half of the STIVEA responders (55%) and two-thirds of BSRBR (35%) were good responders.

A total of 436 out of 466 patients in the BROSG trial attended a 12-month follow-up and completed the EQ-VAS. Eighty-one patients (19%) had an EQ-VAS score worse (>20 & <50%) than baseline at the 1-year follow-up, and 62 patients (14%) had an EQ-VAS score better (>20 & <50%) than at baseline. The improvers were called BROSG(I) and the deteriorators BROSG(D). The proportion of patients from each treatment arm of the BROSG trial was similar in each group. Four hundred and six patients completed 3 years of follow-up in the BROSG trial.

Patients in the BROSG trial and BROSG(D) and the BSRBR control arm deteriorated over the period of follow-up; mean change in EQ-5D ranged from −0.05 to −0.13, and SF-6D ranged from −0.01 to −0.04 (Table 3). Deterioration in SF-6D in the BSRBR control arm was minimal (mean −0.01, s.d. 0.09). The HAQ scores for all these groups deteriorated (mean 0.09–0.16). The DAS28 scores did not show consistent direction of change in these groups, worsening only in the BROSG(D) reported group.

In BROSG(I) patients and those in the BSRBR and STIVEA trial, the EQ-5D (mean 0.06–0.20) and SF-6D (mean 0.03–0.13) indicated improvement. All other outcome measures reflected this improvement over the

follow-up period apart from HAQ, which deteriorated in the BROSG(I) patients (0.07). The improvement in patients in the BSRBR and STIVEA studies was considerable for all outcome measures.

### ES and SRM

The hypothesised magnitude of change, based on the ES, for each of the 6 groups of patients defined by direction of expected change, was equaled or exceeded on 5 occasions by the EQ-5D, and on 4 occasions by the SF-6D (Table 4). The ES for the EQ-5D in patients in the BSRBR group (ES = 0.46) was slightly smaller than the hypothesised moderate response (ES ~ 0.5). The ES for the SF-6D was smaller than expected for patients in the BROSG group (ES = 0.15) and the BSRBR Control (ES = 0.08) group where a small effect size was expected (ES ~ 0.20); the latter group had a very small effect size. The ES for patients from patients reporting a deterioration over 1 year of follow-up (BROSG(D) was larger than anticipated for both the EQ-5D (ES = 0.62) and SF-6D (ES = 0.35). The ES for improvement in patients from STIVEA (EQ-5D ES = 0.64, SF-6D ES = 0.97) unexpectedly exceeded those for BSRBR (EQ-5D ES = 0.46, SF-6D ES = 0.82).

Responsiveness, whether assessed by the ES or SRM, yielded largely similar results (Table 4). However, the comparative responsiveness of the EQ-5D and SF-6D differed according to the direction of change. When health deteriorated over the follow-up period, the EQ-5D was consistently more responsive than the SF-6D. The EQ-5D was most notably more responsive than the SF-6D in the



**Table 3** Mean change over time (s.d.) in each of the groups of patients

	BROSG ( <i>n</i> = 406)	BROSG (D) ( <i>n</i> = 81)	BROSG (I) ( <i>n</i> = 62)	BSRBR control ( <i>n</i> = 188)	BSRBR ( <i>n</i> = 223)	STIVEA ( <i>n</i> = 182)
Duration (months)	36	12	12	6	6	12
EQ-5D	−0.05 (0.24)	−0.13 (0.24)	0.06 (0.27)	−0.07 (0.28)	0.15 (0.34)	0.20 (0.31)
SF-6D	−0.02 (0.11)	−0.04 (0.10)	0.03 (0.09)	−0.01 (0.09)	0.08 (0.12)	0.13 (0.16)
EQ-VAS	−11.8 (28.8)	N/a	N/a	−1.08 (17.8)	10.9 (24.8)	11.2 (25.9)
HAQ <sup>a</sup>	0.16 (0.47)	0.13 (0.48)	0.07 (0.33)	0.09 (0.92)	−0.33 (0.61)	−0.41 (0.68)
DAS28 <sup>a</sup>	−0.09 (1.22)	0.53 (0.99)	−0.03 (0.92)	−0.87 (1.36)	−2.23 (1.47)	−1.85 (1.41)
28 Tender joint count <sup>a</sup>	−0.2 (5.6)	2.4 (5.7)	−0.6 (6.9)	−	−9.6 (7.9)	−6.0 (7.6)
28 Swollen joint count <sup>a</sup>	−1.3 (4.5)	1.3 (4.5)	0.2 (5.0)	−	−6.6 (6.2)	−6.3 (5.4)
ESR (mm/hr) <sup>a</sup>	1.0 (20.2)	−0.4 (30.8)	−2.2 (17.2)	−	−13.6 (22.4)	−15.1 (27.5)

BROSG(D) = EQ-VAS score worse (>20 & <50%) than baseline; BROSG(I) = EQ-VAS score better (>20 & <50%) than at baseline

Follow-up periods differ due to the follow-up assessments available in each study

EQ-5D EuroQol-5D, SF-6D Short Form-6D, EQ-VAS EuroQol visual analogue scale, STIVEA steroids in very early arthritis, BROSG British Rheumatoid Outcome Study Group, BSRBR British Society for Rheumatology Biologics Register, HAQ Health Assessment Questionnaire, DAS28 Disease Activity Score based on 28 swollen joint counts, ESR erythrocyte sedimentation rate, s.d. standard deviation

<sup>a</sup> Higher values denote more severe disease

**Table 4** Responsiveness of the EQ-5D and SF-6D to change in each of the groups, ordered by increasing magnitude of change (EQ-5D)

	Mean (s.d.) change EQ-5D	Effect size (ES)			ES ratio <sup>a</sup>	Standardised response mean (SRM)		SRM ratio <sup>a</sup>
		$H_{\Delta}$	EQ-5D	SF-6D		EQ-5D	SF-6D	
<b>Deterioration</b>								
BROSG	−0.05 (0.24)	~0.20	0.24	0.15	1.6	0.22	0.17	1.3
BSRBR control	−0.07 (0.28)	~0.20	0.24	0.08	3.0	0.23	0.08	2.9
BROSG(D) (1 year)	−0.13 (0.24)	~0.20	0.62	0.35	1.8	0.51	0.37	1.4
<b>Improvement</b>								
BROSG(I) (1 year)	0.06 (0.27)	~0.20	0.22	0.25	1.1	0.21	0.29	1.4
BSRBR	0.15 (0.34)	>0.50	0.46	0.82	1.8	0.44	0.64	1.5
STIVEA	0.20 (0.31)	~0.50	0.64	0.97	1.5	0.64	0.83	1.3

$H_{\Delta}$  hypothesised change (effect size), BROSG (D) = EQ-VAS score worse (>20% & <50%) than baseline; BROSG (I) = EQ-VAS score better (>20% & <50%) than at baseline

EQ-5D EuroQol-5D, SF-6D Short Form-6D, STIVEA steroids in very early arthritis, BROSG British Rheumatoid Outcome Study Group, BSRBR British Society for Rheumatology Biologics Register, s.d. standard deviation

<sup>a</sup> Calculated as the larger of the effect size for the EQ-5D or SF-6D divided by the smaller effect size

BSRBR control group (ES ratio 3.0); the SF-6D failed to respond to deterioration [mean change −0.01 (0.09)] in this group. All ES ratios indicated that the EQ-5D was more than 1.5 times more responsive to deterioration. In contrast, when patients improved over follow-up, the SF-6D was more responsive than the EQ-5D, particularly in the STIVEA (ES ratio 1.5) and BSRBR groups (ES ratio 1.8) where a large improvement was detected.

#### Correlation

The correlation of change in EQ-5D and SF-6D ranged from 0.25 in the BROSG 12-month deterioration group to

0.48 in the STIVEA patients (Table 5). The change in SF-6D correlated more strongly than the change in EQ-5D with change in patient EQ-VAS rated health in all of the cohorts apart from the BSRBR control cohort group, where correlations were equal. Change in DAS28 and its components (tender and swollen joint counts, ESR) was generally more strongly correlated with change in SF-6D score than change in EQ-5D score. Similarly, change in HAQ in STIVEA and BSRBR was significantly more strongly correlated with the change in SF-6D score than change in EQ-5D score. The SF-6D was significantly more strongly correlated than the EQ-5D with the DAS28 in the BROSG 12-month improvement group.

**Table 5** Correlation of change between outcome measures

		EQ-5D	SF-6D	DAS28	HAQ	EQ-VAS	28 Tender	28 Swollen	ESR
STIVEA	EQ-5D	–	0.48	–0.40	–0.53	–0.52	–0.27	–0.04	–0.35
	SF-6D	0.48	–	–0.50	–0.67 <sup>†</sup>	–0.59	–0.34	–0.16	–0.31
BROSG (3-year)	EQ-5D	–	0.38	–0.16	–0.28	0.26	–0.11	0.08	–0.18
	SF-6D	0.38	–	–0.17	–0.36	0.32	–0.10	0.01	–0.23
BROSG(D) (1 year)	EQ-5D	–	0.25	–0.16	–0.14	0.16	–0.03	–0.13	–0.11
	SF-6D	0.25	–	–0.29	–0.24	0.31	–0.04	–0.11	–0.12
BROSG(I) (1 year)	EQ-5D	–	0.37	–0.17	–0.31	0.19	–0.16	–0.03	0.01
	SF-6D	0.37	–	–0.37 <sup>†</sup>	–0.20	0.32	–0.18	–0.25 <sup>†</sup>	–0.16
BSRBR	EQ-5D	–	0.44	–0.16	–0.37	0.41	–0.14	–0.07	0.01
	SF-6D	0.44	–	–0.34	–0.64 <sup>††</sup>	0.48	–0.28	–0.23 <sup>†</sup>	0.04
BSRBR control	EQ-5D	–	0.38	–0.16	–0.22	0.26	–	–	–
	SF-6D	0.38	–	–0.23	–0.21	0.26	–	–	–

BROSG(D) = EQ-VAS score worse (>20 & <50%) than baseline; BROSG(I) = EQ-VAS score better (>20 & <50%) than at baseline

EQ-5D EuroQol-5D, SF-6D Short Form-6D, EQ-VAS EuroQol Visual Analogue Scale, STIVEA steroids in very early arthritis, BROSG British Rheumatoid Outcome Study Group, BSRBR British Society for Rheumatology Biologics Register, HAQ Health Assessment Questionnaire, DAS28 Disease Activity Score based on 28 swollen joint counts, ESR erythrocyte sedimentation rate, *s.d.* standard deviation

Steiger's Z-test: <sup>†</sup>  $P < 0.05$ ; <sup>††</sup>  $P < 0.01$

### Floor and ceiling effects

Overall floor and ceiling effects of the EQ-5D and SF-6D were small in this study (Table 6). No patient scored at the floor of the EQ-5D, and 2% or fewer scored at the floor of the SF-6D. However, floor effects existed for individual domains. Floor effects for EQ-5D pain/discomfort were small in the BROSG (7%) and BSRBR control groups (9%), but serious (26% in STIVEA and 39% in BSRBR). Serious floor effects were also evident in the role limitation subscale of the SF-6D (32–65%), the vitality subscale (19–36%) and in the physical functioning scale (18–24%) in the BSRBR, BSRBR control and STIVEA groups.

There were no serious ceiling effects for the EQ-5D (<1–8%) and no patient scored at the ceiling of the SF-6D. However, serious ceiling effects existed in the self-care (22–49%) and anxiety/depression (44–66%) EQ-5D domains for all groups, in the usual activities domain for STIVEA (15%), BROSG (22%) and the BSRBR control (17%) groups, and mobility in BROSG (22%), BSRBR control (23%) and STIVEA (36%). Ceiling effects were serious in all groups for the social functioning subscale (18–45%) and all groups but STIVEA (12%) for the mental health subscale (16–27%). In addition, there were small to serious (7–27%) ceiling effects in the mobility subscale.

### Discussion

This study is the first to compare the responsiveness of the EQ-5D and SF-6D to longitudinal changes in UK RA patients with different expected disease trajectories. These

ranged from patients with early disease expected to improve through to patients with severe long-standing disease expected to deteriorate. Our results have highlighted key differences in the ability of the EQ-5D and SF-6D to measure change. Of note, the EQ-5D was more responsive to deterioration in health than the SF-6D, whereas the SF-6D was more responsive to improvement. The SF-6D was unable to detect further deterioration in a group of patients with already severe disease.

The finding that the EQ-5D is more responsive to deterioration than the SF-6D is in keeping with all previous reports in the literature [10, 12]. Similarly, the greater responsiveness of the SF-6D to improvement supports the majority of previous findings [10, 11]. All previous studies have suggested that both measures are generally responsive to change in the RA patient [8]. However, the SF-6D has a clear limitation in severe RA patients, which has not previously been demonstrated.

The ability of the SF-6D to detect change is thought to be inhibited by the high floor of the measure. However, in this study, few patients scored at the floor of the SF-6D in any of the cohorts, although within-domain floor effects were considerable for the role limitation, vitality and physical functioning domains. These domains relate to key aspects of the limitation caused by RA, and the floor effects may explain the lack of response when patients deteriorated further in the BSRBR cohort. These floor effects were severe but smaller in the groups where the SF-6D was responsive to deterioration. However, the correlations of change in the SF-6D with change in the HAQ and DAS28 scores were stronger than corresponding correlations with the EQ-5D. This suggests that the superior responsiveness

**Table 6** Floor/ceiling effects for the EQ-5D and SF-6D

	BROSG (%) n = 406	BROSG (D) (%) n = 81	BROSG (I) (%) n = 62	BSRBR control (%) n = 188	STIVEA (%) n = 182	BSRBR (%) n = 223
EQ-5D	0/1	0/<1	0/8	0/6	0/< 1	0/1
Mobility	0/22	0/10	0/25	0/23	0/36	0/10
Self-care	<1/47	<1/32	0/49	2/43	1/42	2/22
Usual activities	4/22	9/8	2/23	4/17	14/15	15/4
Pain/discomfort	8/7	15/2	2/6	13/9	26/1	39/3
Anxiety/depression	1/66	3/62	0/65	3/56	4/54	5/44
SF-6D	2/0	<1/0	0/0	2/0	2/0	2/0
Physical functioning [6]	12/1	18/2	12/1	18/2	18/1	24/1
Role limitation [4]	39/24	52/15	32/27	64/7	48/14	65/7
Social functioning [5]	4/39	8/22	1/45	8/25	9/27	9/18
Pain [6]	2/2	3/3	0/2	8/2	22/1	11/2
Mental health [5]	3/22	7/16	2/25	5/28	6/12	3/27
Vital [5]	19/1	28/1	11/2	36/2	30/4	36/3

BROSG(D) = EQ-VAS score worse (>20 & <50%) than baseline; BROSG(I) = EQ-VAS score better (>20% & <50%) than at baseline

Numbers in brackets for SF-6D are the number of levels within domains

EQ-5D EuroQol-5D, SF-6D Short Form-6D, STIVEA steroids in very early arthritis, BROSG British Rheumatoid Outcome Study Group, BSRBR British Society for Rheumatology Biologics Register

of the EQ-5D to deterioration in health in this group is related to some aspect other than functional disability or disease activity. The EQ-5D had no floor effects in the overall utility score or any domains in the BSRBR control group, providing scope for detecting extra deterioration in all aspects of disease measured by this instrument in these patients. The domains of the SF-6D with floor effects, role limitation, physical functioning and vitality are likely to be captured by the self-care and usual activities domains of the EQ-5D, which have no floor effects.

The mean change of the EQ-5D exceeded the mean change of the SF-6D in all of the cohorts used in this study. This has implications in the use of responsiveness statistics and in using the measure for cost-effectiveness analyses. The EQ-5D was less responsive than the SF-6D to improvement despite the larger mean change using the EQ-5D highlighting the variance around the measure. The SF-6D has smaller increments between scoring levels than the EQ-5D which allows patients to report smaller improvements. This may explain why the SF-6D is more responsive to small but important improvements. The SF-6D shows relatively small absolute change but has a small standard deviation [11, 12], which leads to a good responsiveness statistics. Changes in a single domain of the EQ-5D can result in changes of 0.036–0.655 in the overall utility score. Therefore, change in a small number of patients can lead to a large group mean change effect. The impact on the overall EQ-5D scores is largest when a domain is scored at the most severe level for the first time. This attracts both reductions in utility associated with the change in domain (range 0.094–0.386) and reduction of

0.269 for the first domain scored as severe, known as the N3 term.

The larger mean change in improving and deteriorating patients suggests that an intervention will be more likely to be seen as cost-effective if assessed using the EQ-5D rather than the SF-6D. In cost-effectiveness analysis, the incremental cost of an intervention is divided by its incremental effectiveness, measured using a measure such as the EQ-5D or SF-6D. The larger the effect estimate, the lower the cost per unit of effect. A recent study in RA reported that change estimated using the EQ-5D resulted in a cost per QALY over 50% lower than the cost per QALY calculated using the SF-6D [38]. However, as 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 a concerning the relative cost-effectiveness of two treatments.

There are weaknesses in the use of responsiveness statistics. There is an array of such statistics, and different methods may lead to different conclusions [33, 39, 40]. To date, no measure has been proven conclusively to be superior to another. Furthermore, responsiveness statistics is limited in the information they convey. They give an indication of whether a measure can detect a statistically significant difference between two groups. However, statistical significance is dependent on factors external to the measure under study such as sample size and does not indicate whether the change detected is meaningful or useful [33, 41]. The ES and SRM express change in terms of the standard deviation and provide a useful indication of the relative sample sizes required to detect statistically significant difference between



groups [39]. The basis of ES and SRM is that relevant change should exceed random noise or the variability in unchanged patients. These measures used without an anchor of important change give no information about the ability of the instrument to measure change in the underlying construct, [33] and essentially are measures of sensitivity. We used the responsiveness measures alongside some external reference of change, for example the change in EQ-VAS score, or response to treatment. This was not possible for the overall BROSG data or for the BSRBR control cohort, and we, therefore, cannot assume that all patients in these cohorts deteriorated; the HAQ score for these cohorts suggested deterioration of functional disability approaching the minimally important difference for clinical practice [42]; however, the DAS28 scores in the BSRBR control cohort suggested some improvement in disease activity from the high baseline level.

Comparison of change in different cohorts was limited by the different follow-up periods used. The analysis relied on data collected concurrently within each study and was therefore limited by the design of each cohort. Patients in the BSRBR treatment and control studies were only followed for 6 months. This may be sufficient to capture the large expected improvements in patients treated with anti-TNF therapy, but may be insufficient to capture clinically meaningful deterioration in patients in the control arm continuing with traditional treatment of RA. However, the change in EQ-5D for this latter group of patients was in excess of estimates of the minimum important difference for this measure [8].

A further limitation of the ES is that in highly selected groups of patients, the ES may be artificially inflated. The BROSG and STIVEA trials by definition were selected groups of patients. It was therefore important to use the SRM to verify the results based on the ES. In only one instance were the conclusions based on the ES and SRM conflicting, and where this occurred, the difference between responsiveness of the EQ-5D and SF-6D was marginal. Finally, the methods of the ES and SRM assume that all patients change in the same direction [39]. It is likely that there is some misclassification of change between the anchors and some of the outcome measures. This may be particularly true for the EQ-VAS scale which frames a patient's health on the day in question; therefore, change in EQ-VAS assesses the difference between a person's health on 2 days, 1 year apart. The framing of the question gives considerable potential for transient and possibly trivial factors to influence the rating of change in health. However, the design of this study aimed to classify patients on the basis of important change and only compared responsiveness of the EQ-5D and SF-6D in patients changing in the same direction by more than a certain amount.

The ability to measure change in the RA patient is indicative of longitudinal construct validity [43]. Concerns

have been voiced about the ability of the EQ-5D to measure change due to its bimodal distribution, crude scoring and possible ceiling effects within domains [11, 12, 44, 45]. These issues were evident in the data used in this study, but the EQ-5D appeared to respond to both improvement and deterioration. The EQ-5D was more responsive to deterioration, and the SF-6D more responsive to improvement in patients with inflammatory arthritis. The SF-6D does not appear appropriate for use in patients with established severe RA, who are expected to experience disease progression. Responsiveness of a measure affects the power of a given sample size to detect a statistically significant difference. As an outcome measure in an epidemiological setting, the SF-6D requires a smaller sample size than the EQ-5D to detect improvement in patients whose health is getting better. The opposite is true in worsening patients. In economic analysis, however, the approach is different. The incremental cost-effectiveness ratio (ICER) is the primary estimate of cost-effectiveness of an intervention, and if this value is less than a decision-makers willingness to pay, then the intervention should be adopted [46, 47]. The EQ-5D consistently provided larger mean change estimates than the SF-6D, even when less responsive than the SF-6D (due to the greater variance around the EQ-5D), which would result in a more optimistic incremental cost-effectiveness ratio.

## Conclusion

The results from the four cohorts of patients used in this study demonstrate that the comparative responsiveness of the EQ-5D and SF-6D differs according to the direction of change; the EQ-5D was more responsive to deterioration in health than the SF-6D, whereas the SF-6D was more responsive to improvement. The level of mean change of the EQ-5D, which is consistently larger than that of the SF-6D, has potentially serious implications for decision-making on the basis of cost-effectiveness analysis; the EQ-5D is likely to provide more optimistic cost-effectiveness ratios than the SF-6D. Our results support the responsiveness of the EQ-5D to improvement and deterioration across a range of arthritis states/severity. The SF-6D was responsive to improvement in cohorts of patients with range of arthritis severity and to deterioration in patients with established stable disease; however, use of the SF-6D in patients with severe progressive disease may be inappropriate.

**Acknowledgments** This STIVEA study was funded by the Arthritis Research Campaign UK. The authors would like to thank all the rheumatologists and research nurses of the participating hospitals and all GPs who referred patients to the rheumatology departments. We also would like to thank all members of the Trial Steering Committee of this study. The BROSG project was funded by the NHS Executive, UK

(NHS HTA project number 94/45/02). The views and opinions expressed within do not necessarily reflect those of the NHS Executive. The NHS Executive commissioned this work, but played no part in the design, data collection, analysis, interpretation, report writing or decision to publish this paper. The BROSG Study Group: Dr D Mulherin (Cannock), Dr S Knight (Macclesfield), Prof D Scott (King's College, London), Dr P Dawes (Stoke-on-Trent), Dr M Davis (Truro). The British Society for Rheumatology Biologics Register is supported by a research grant from the British Society for Rheumatology to the University of Manchester, which is indirectly funded by Schering-Plough, Wyeth Laboratories, Abbott Laboratories, Amgen and Roche. We thank all rheumatologists and patients who provided data used in this analysis.

**Open Access** This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

## Appendix

The members of The British Society for Rheumatology Biologics Register (BSRBR) Control Centre Consortium

Musgrave Park Hospital, Belfast (Dr Allister Taggart); Cannock Chase Hospital, Cannock Chase (Dr Tom Price); Christchurch Hospital, Christchurch (Dr Neil Hopkinson); Derbyshire Royal Infirmary, Derby (Dr Sheila O'Reilly); Russells Hall Hospital, Dudley (Dr George Kitas); Gartnavel General Hospital, Glasgow (Dr Duncan Porter); Glasgow Royal Infirmary, Glasgow (Dr Hilary Capell); Leeds General Infirmary, Leeds (Prof Paul Emery); King's College Hospital, London (Dr Ernest Choy); Macclesfield District General Hospital, Macclesfield (Prof Deborah Symmons); Manchester Royal Infirmary, Manchester (Dr Ian Bruce); Freeman Hospital, Newcastle-upon-Tyne (Dr Ian Griffiths); Norfolk and Norwich University Hospital, Norwich (Prof David Scott); Poole General Hospital, Poole (Dr Paul Thompson); Queen Alexandra Hospital, Portsmouth (Dr Fiona McCrae); Hope Hospital, Salford (Dr Romela Benitha); Selly Oak Hospital, Selly Oak (Dr Ronald Jubb); St Helens Hospital, St Helens (Dr Rikki Abernethy); Haywood Hospital, Stoke-on-Trent (Dr Andy Hassell); Kings Mill Centre, Sutton-In Ashfield (Dr David Walsh).

## References

- Drummond, M. F., Sculpher, M. J., Torrance, G. W., O'Brien, B. J., & Stoddart, G. L. (2005). *Methods for the economic evaluation of health care programmes* (3rd ed.). Oxford: Oxford University Press.
- National Institute for Health and Clinical Excellence. (2005). *A guide to NICE*. London: National Institute for Clinical Excellence.
- Ontario Ministry of Health and Long-Term Care. (2007). *Ontario guidelines for economic analysis of pharmaceutical products*. Internet 2004 August [cited 2007 Jan 31]; Available from: URL: [http://www.health.gov.on.ca/english/providers/pub/drugs/economic/economic\\_mn.html](http://www.health.gov.on.ca/english/providers/pub/drugs/economic/economic_mn.html).
- Sullivan, S. D., Lyles, A., Luce, B., & Grigar, J. (2001). AMCP guidance for submission of clinical and economic evaluation data to support formulary listing in U.S. health plans and pharmacy benefits management organizations. *Journal of Managed Care Pharmacy*, 7(4), 272–282.
- NICE. (2004). *Guide to the methods of technology appraisal*. London: National Institute for Clinical Excellence. Ref Type: Pamphlet.
- Brazier, J., Roberts, J., & Deverill, M. (2002). The estimation of a preference-based measure of health from the SF-36. *Journal of Health Economics*, 21(2), 271–292.
- Brazier, J. E., & Roberts, J. (2004). The estimation of a preference-based measure of health from the SF-12. *Medical Care*, 42(9), 851–859.
- Harrison, M. J., Davies, L. M., Bansback, N. J., Ingram, M., Anis, A. H., & Symmons, D. P. (2008). The validity and responsiveness of generic utility measures in rheumatoid arthritis: A review. *The Journal of Rheumatology*, 35(4), 592–602.
- Hurst, N. P., Kind, P., Ruta, D., Hunter, M., & Stubbings, A. (1997). Measuring health-related quality of life in rheumatoid arthritis: Validity, responsiveness and reliability of EuroQol (EQ-5D). *British Journal of Rheumatology*, 36(5), 551–559.
- Marra, C. A., Rashidi, A. A., Guh, D., Kopec, J. A., Abrahamowicz, M., Esdaile, J. M., et al. (2005). Are indirect utility measures reliable and responsive in rheumatoid arthritis patients? *Quality of Life Research*, 14, 1333–1344.
- Russell, A. S., Conner-Spady, B., Mintz, A., Mallon, C., & Maksymowych, W. P. (2003). The responsiveness of generic health status measures as assessed in patients with rheumatoid arthritis receiving infliximab. *The Journal of Rheumatology*, 30(5), 941–947.
- Conner-Spady, B., & Suarez-Almazor, M. E. (2003). Variation in the estimation of quality-adjusted life-years by different preference-based instruments. *Medical Care*, 41(7), 791–801.
- Verstappen, S. M. M., McCoy, M. J., Roberts, C., Hassell, A. B., Dale, N. E., & Symmons, D. P. M. (2008). A 3 week course of IM steroid injection may prevent the progression of very early inflammatory polyarthritis: Results of the STIVEA trial. *Annals of the Rheumatic Diseases*, 67(Suppl II), 61.
- Symmons, D., Tricker, K., Harrison, M., Roberts, C., Davis, M., Dawes, P., et al. (2006). Patients with stable long-standing rheumatoid arthritis continue to deteriorate despite intensified treatment with traditional disease modifying anti-rheumatic drugs—results of the British Rheumatoid Outcome Study Group randomized controlled clinical trial. *Rheumatology (Oxford)*, 45, 558–565.
- Silman, A., Symmons, D., Scott, D. G., & Griffiths, I. (2003). British society for rheumatology biologics register. *Annals of the Rheumatic Diseases*, 62(Suppl 2), ii28–ii29.
- The EuroQol Group. (1990). EuroQol—a new facility for the measurement of health-related quality of life. *The EuroQol Group. Health Policy*, 16(3), 199–208.
- Ware, J. E., Jr., & Sherbourne, C. D. (1992). The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. *Medical Care*, 30(6), 473–483.
- Prevoo, M. L. L., Vanthof, M. A., Kuper, H. H., Vanleeuwen, M. A., Vandeputte, L. B. A., & Vanriel, P. L. C. M. (1995). Modified disease-activity scores that include 28-joint counts—development and validation in a prospective longitudinal-study of patients with rheumatoid-arthritis. *Arthritis and Rheumatism*, 38(1), 44–48.
- Bansback, N., Harrison, M., Brazier, J., Davies, L., Kopec, J., Marra, C., et al. (2008). Health state utility values: A description

- of their development and application for rheumatic diseases. *Arthritis and Rheumatism*, 59(7), 1018–1026.
20. Symmons, D. P., Barrett, E. M., Bankhead, C. R., Scott, D. G., & Silman, A. J. (1994). The incidence of rheumatoid arthritis in the United Kingdom: Results from the Norfolk Arthritis Register. *British Journal of Rheumatology*, 33(8), 735–739.
  21. Wiles, N. J., Lunt, M., Barrett, E. M., Bukhari, M., Silman, A. J., Symmons, D. P., et al. (2001). Reduced disability at five years with early treatment of inflammatory polyarthritis: Results from a large observational cohort, using propensity models to adjust for disease severity. *Arthritis and Rheumatism*, 44(5), 1033–1042.
  22. Verstappen, S. M., Jacobs, J. W., Bijlsma, J. W., Heurkens, A. H., van Booma-Frankfort, C., Borg, E. J., et al. (2003). Five-year followup of rheumatoid arthritis patients after early treatment with disease-modifying antirheumatic drugs versus treatment according to the pyramid approach in the first year. *Arthritis and Rheumatism*, 48(7), 1797–1807.
  23. Moreland, L. W., Schiff, M. H., Baumgartner, S. W., Tindall, E. A., Fleischmann, R. M., Bulpitt, K. J., et al. (1999). Etanercept therapy in rheumatoid arthritis. A randomized, controlled trial. *Annals of Internal Medicine*, 130(6), 478–486.
  24. Maini, R., St Clair, E. W., Breedveld, F., Furst, D., Kalden, J., Weisman, M., et al. (1999). Infliximab (chimeric anti-tumour necrosis factor alpha monoclonal antibody) versus placebo in rheumatoid arthritis patients receiving concomitant methotrexate: A randomised phase III trial. ATTRACT Study Group. *Lancet*, 354(9194), 1932–1939.
  25. Weinblatt, M. E., Keystone, E. C., Furst, D. E., Moreland, L. W., Weisman, M. H., Birbara, C. A., et al. (2003). Adalimumab, a fully human anti-tumor necrosis factor alpha monoclonal antibody, for the treatment of rheumatoid arthritis in patients taking concomitant methotrexate: The ARMADA trial. *Arthritis and Rheumatism*, 48(1), 35–45.
  26. van Gestel, A. M., Haagsma, C. J., & van Riel, P. L. (1998). Validation of rheumatoid arthritis improvement criteria that include simplified joint counts. *Arthritis and Rheumatism*, 41, 1845–1850.
  27. Grigor, C., Capell, H., Stirling, A., McMahon, A. D., Lock, P., Vallance, R., et al. (2004). Effect of a treatment strategy of tight control for rheumatoid arthritis (the TICORA study): A single-blind randomised controlled trial. *Lancet*, 364(9430), 263–269.
  28. Anderson, J. J., Wells, G., Verhoeven, A. C., & Felson, D. T. (2000). Factors predicting response to treatment in rheumatoid arthritis: The importance of disease duration. *Arthritis and Rheumatism*, 43(1), 22–29.
  29. Hider, S. L., Buckley, C., Silman, A., Symmons, D. P. M., & Bruce, I. N. (2004). Factors influencing response to disease modifying anti-rheumatic drugs in patients with rheumatoid arthritis. *The Journal of Rheumatology* (in press).
  30. Cohen, J. (1977). *Statistical power analysis for the behavioural sciences*. New Jersey: Lawrence Erlbaum.
  31. Scott, D. L., Smith, C., & Kingsley, G. (2003). Joint damage and disability in rheumatoid arthritis: An updated systematic review. *Clinical and Experimental Rheumatology*, 21(5 Suppl 31), S20–S27.
  32. Scott, D. L. (2007). Early rheumatoid arthritis. *British Medical Bulletin*, 81–82, 97–114.
  33. Terwee, C. B., Dekker, F. W., Wiersinga, W. M., Prummel, M. F., & Bossuyt, P. M. M. (2003). On assessing responsiveness of health-related quality of life instruments: Guidelines for instrument evaluation. *Quality of Life Research*, 12(4), 349–362.
  34. Norman, G. R., Wyrwich, K. W., & Patrick, D. L. (2007). The mathematical relationship among different forms of responsiveness coefficients. *Quality of Life Research*, 16(5), 815–822.
  35. Steiger, J. H. (1980). Tests for comparing elements of a correlation matrix. *Psychological Bulletin*, 87(2), 245–251.
  36. Meng, X. L., Rosenthal, R., & Rubin, D. B. (1992). Comparing correlated correlation-coefficients. *Psychological Bulletin*, 111(1), 172–175.
  37. Veenhof, C., Bijlsma, J. W. J., van den Ende, C. H. M., van Dijk, G. M., Pisters, M. F., & Dekker, J. (2006). Psychometric evaluation of osteoarthritis questionnaires: A systematic review of the literature. *Arthritis Care and Research*, 55(3), 480–492.
  38. Marra, C. A., Marion, S. A., Guh, D. P., Najafzadeh, M., Wolfe, F., Esdaile, J. M., et al. (2007). Not all “quality-adjusted life years” are equal. *Journal of Clinical Epidemiology*, 60(6), 616–624.
  39. Liang, M. H. (2000). Longitudinal construct validity: Establishment of clinical meaning in patient evaluative instruments. *Medical Care*, 38(9 suppl), I184–I190.
  40. Wright, J. G., & Young, N. L. (1997). A comparison of different indices of responsiveness. *Journal of Clinical Epidemiology*, 50(3), 239–246.
  41. Husted, J. A., Cook, R. J., Farewell, V. T., & Gladman, D. D. (2000). Methods for assessing responsiveness: A critical review and recommendations. *Journal of Clinical Epidemiology*, 53(5), 459–468.
  42. Pope, J. E., Khanna, D., Norrie, D., & Ouimet, J. M. (2009). The minimally important difference for the health assessment questionnaire in rheumatoid arthritis clinical practice is smaller than in randomized controlled trials. *The Journal of Rheumatology*, 36(2), 254–259.
  43. Hays, R. D., & Hadorn, D. (1992). Responsiveness to change: An aspect of validity, not a separate dimension. *Quality of Life Research*, 1(1), 73–75.
  44. Wolfe, F., & Hawley, D. J. (1997). Measurement of the quality of life in rheumatic disorders using the EuroQol. *British Journal of Rheumatology*, 36(7), 786–793.
  45. Marra, C. A., Esdaile, J. M., Guh, D., Kopec, J. A., Brazier, J. E., Koehler, B. E., et al. (2004). A comparison of four indirect methods of assessing utility values in rheumatoid arthritis. *Medical Care*, 42(11), 1125–1131.
  46. Briggs, A. H., O’Brien, B. J., & Blackhouse, G. (2002). Thinking outside the box: Recent advances in the analysis and presentation of uncertainty in cost-effectiveness studies. *Annual Review of Public Health*, 23, 377–401.
  47. Claxton, K. (1999). The irrelevance of inference: A decision-making approach to the stochastic evaluation of health care technologies. *Journal of Health Economics*, 18(3), 341–364.
  48. Wells, G. A., Tugwell, P., Kraag, G. R., Baker, P. R., Groh, J., & Redelmeier, D. A. (1993). Minimum important difference between patients with rheumatoid arthritis: The patient’s perspective. *The Journal of Rheumatology*, 20(3), 557–560.
  49. Kosinski, M., Zhao, S. Z., Dedhiya, S., Osterhaus, J. T., & Ware, J. E., Jr. (2000). Determining minimally important changes in generic and disease-specific health-related quality of life questionnaires in clinical trials of rheumatoid arthritis. *Arthritis and Rheumatism*, 43(7), 1478–1487.