

An assessment of the discriminative ability of the EQ-5D_{index}, SF-6D, and EQ VAS, using sociodemographic factors and clinical conditions

Garry R. Barton · Tracey H. Sach · Michael Doherty ·
Anthony J. Avery · Claire Jenkinson · Kenneth R. Muir

Received: 22 July 2006 / Accepted: 1 June 2007 / Published online: 29 June 2007
© Springer-Verlag 2007

Abstract

Objective To assess whether three health-related quality-of-life (HRQL) measures (the EQ-5D_{index}, SF-6D, and EQ VAS) can discriminate between the HRQL of different groups of individuals.

Methods In one UK general practice a cross-sectional survey requested information on six sociodemographic factors, 10 clinical conditions, and the three HRQL measures. Regression analyses were used to assess whether there was a significant difference in HRQL between groups with different sociodemographic factors and those with and without clinical conditions.

Results One thousand eight hundred and sixty-five questionnaires were returned. There was a significant difference between the HRQL of the majority of different groups according to each HRQL measure. However, not all of the measures could discriminate between groups of different ethnicity, gender, or smoking status, or those with and without asthma, stroke, cancer or diabetes.

Conclusion The HRQL of the majority of different groups could be discriminated between by the EQ-5D_{index}, SF-6D, and EQ VAS.

Keywords Construct validity · EQ-5D · SF-6D · Health-related quality of life

Introduction

When testing the construct validity of an instrument (often referred to as construct validation by extreme groups, or discriminative validity) it is common to estimate the difference in health-related quality of life (HRQL) between two different groups of individuals, in order to assess whether a measure can consistently discriminate between two patient groups, where one has a certain trait and the other does not [1]. Such an assessment constitutes an important component of the validation of HRQL measures as such empirical evidence is required to demonstrate that a measure is measuring what it is intended to measure [1, 2]. Thus, within this paper, we seek to assess whether different HRQL measures can discriminate between the HRQL of different groups of individuals. It is, however, important to acknowledge that such an assessment constitutes only one property contributing towards construct validity because evidence of construct validity, as Streiner and Norman [1] point out, can itself only be provided by a series of converging results, and not by a single study—no single experiment can unequivocally prove a construct.

In this paper we use three HRQL measures. The first two (the EQ-5D_{index} [3] and SF-6D [4]) use questionnaires to elicit a health state description, which is in turn converted into a utility score (a scale where 0 is equivalent to death, and 1 is equivalent to full health), and the third (EQ VAS [3]) asks people to rate their current health state on a 0 to 100 scale (see the [Methods](#) section for further details). Marra et al. [5] have argued that there remains a gap in the literature regarding the assessment of utility measures with

G. R. Barton (✉)
School of Economics, University of Nottingham,
Nottingham NG7 2UH, UK
e-mail: mcxgrb@nottingham.ac.uk

T. H. Sach · A. J. Avery · C. Jenkinson · K. R. Muir
School of Community Health Sciences,
University of Nottingham, Nottingham, UK

M. Doherty
Academic Rheumatology, University of Nottingham,
Nottingham, UK

regard to criteria such as construct validity. Moreover, most papers which have estimated the difference in HRQL between different groups of individuals have done so without adjusting for differences between these groups [6–11], or only adjusted for differences in age and sex [12, 13]. Recently, however, two papers [14, 15] have extended this previous literature, by assessing whether the EQ-5D_{index} can discriminate between different groups of individuals, after adjusting for differences in both sociodemographic factors and clinical conditions. Such adjustments were made because particular sociodemographic factors (e.g., age and gender) are associated with many clinical conditions. In this paper, we contribute further to the evidence on the construct validity of HRQL measures by similarly assessing whether the EQ-5D_{index}, and in addition the SF-6D (derived from responses to the SF-36 questionnaire [16]) and EQ VAS, can discriminate between the HRQL of different groups of individuals with different sociodemographic factors and clinical conditions. It should also be noted that, in line with previous studies [13, 15], we consider that a measure can discriminate between the HRQL of different groups when groups are found to have a significantly different level of HRQL, where such comparisons are made both whilst adjusting for, and without adjusting for, other sociodemographic factors and clinical conditions.

We hypothesized that HRQL measures would be able to discriminate in accordance with differences previously observed in the literature. Namely that HRQL would vary according to age [6–8, 10–15, 17], gender [6, 11–15, 18], smoking status [6, 12], ethnicity [6, 14, 15], body mass index (BMI) [18–20], education [7, 10–12, 14, 15], social class [6, 12], economic status [12], housing tenure [12], income [11, 14, 15], and the presence of health problems [6–8, 10, 12–15, 17]. With occupational skill level [21] acting as a proxy for five variables (education, social class, economic status, housing tenure, and income), we used cross-sectional data to assess whether the EQ-5D_{index}, SF-6D and EQ VAS could discriminate between the HRQL of people with different sociodemographic factors (age, gender, ethnicity, smoking status, BMI, and occupational skill level), and people with and without different clinical conditions (back pain, hip pain, knee pain, heart disease, stroke, asthma, cancer, diabetes, rheumatoid arthritis, and osteoarthritis).

Minimally important difference

As outlined above, we initially considered that a measure could discriminate between different groups when there was a significant difference between the HRQL of those groups of individuals. Claxton [22] has however argued that decisions should be made on the basis of mean values, irrespective of whether such differences are statistically

significant. Thus, we also focus on the size of the difference in HRQL between different groups of individuals, and in particular we consider whether these differences can be considered to constitute a minimally important difference (MID). Jaeschke et al. [23] defined an MID as “The smallest difference in score in the domain of interest which patients perceive as beneficial and which would mandate, in the absence of troublesome side effects and excessive cost, a change in the patient’s management” (quote taken from Walters and Brazier [24]). In an attempt to determine the MID Walters and Brazier [24] reviewed eight longitudinal studies in 11 patient groups. They estimated that the MID for the EQ-5D_{index} ranged between –0.011 and 0.140 (mean = 0.074), compared to 0.011–0.097 (mean = 0.041) for the SF-6D [24], where the differential valuation range covered by the EQ-5D_{index} and SF-6D scores was offered as a potential reason for these different MID estimates. Other estimates of the MID, for the EQ-5D_{index}, include 0.033 (based on an effect size of 0.2) [14] and 0.036 (the smallest change associated with a change in levels on the EQ-5D_{index}) [25], and 0.03 for the SF-6D (the mean change in the SF-6D associated with a reported change in general health in nine reviewed studies) [26]. Within this paper we took a conservative approach and inferred that a difference in HRQL of >0.03, between different groups of individuals constituted an MID on the EQ-5D_{index} and SF-6D. As we are unaware of any studies which have estimated the MID for the EQ VAS we similarly assumed, for the purposes of our analysis, that a difference of >3.0 on the EQ VAS constituted an MID.

Methods

Participants and procedures

As part of the recruitment for a study designed to assess the cost-effectiveness of different lifestyle interventions for knee pain (LIKIP) an ascertainment questionnaire was posted to all people who were registered in one United Kingdom (UK) general practice and aged ≥45 years, with the following exceptions—those who were deemed (by their general practitioner) to be unable to complete information requested in a questionnaire, or had had either a total knee replacement, an amputation of a lower limb, a permanent cardiac pacemaker, or an intra-articular steroid injection within the last 3 months. The ascertainment questionnaire requested information on sociodemographic factors (including age, gender, ethnicity, smoking status, height, weight, and current/previous job title), HRQL (according to the EQ-5D and SF-36 questionnaires), a person’s level of knee pain, and details of other clinical conditions. Additionally, respondents were asked to

provide their name and address if they were willing to take further part in the LIKP study. People were asked to complete the ascertainment questionnaire, and post it to the LIKP study team at the University of Nottingham. The implication of such methods was that no checks were made on the self-reported data. Ethical approval for this study was granted by the Nottingham research ethics committee.

Data from all returned questionnaires were used in this study, even if not all sections of the questionnaire were completed; this also applied to questionnaires on which the person's name and address were not reported. For the purposes of this paper data was extracted in order to enable the information on six sociodemographic factors to be calculated for each person: age, gender, ethnicity, smoking status, BMI (kg/m²) and occupational skill level. In addition, information about 10 common clinical conditions (whether they had back pain, hip pain, and/or knee pain on most days of the last month, and whether they had been diagnosed with heart disease, stroke, asthma, cancer, diabetes, rheumatoid arthritis, and osteoarthritis), and HRQL (according to the EQ-5D and SF-36 questionnaires) was also extracted. People were asked to report which ethnic group best described them. Due to the small number of non-white respondents we categorized each person as either white or non-white. Similarly, people were categorised according to whether they had, or had not, ever smoked regularly for a period of at least 3 months. BMI was categorised in accordance with World Health Organisation (WHO) recommendations [27], and in line with US National Institutes of Health (NIH) clinical guidelines [28], a BMI of <18.5 was considered underweight, of 18.5 to <25 as normal, of 25 to <30 as overweight, and of ≥30 as obese. The occupational skill level of each person was estimated by assigning the standard occupational classification (SOC2000) skill level [21] to their reported job title. The SOC2000 categorizes each job title into one of four skill levels, where 1 is the lowest and 4 the highest, dependent upon the length of time deemed necessary for a person to become fully competent in the performance of tasks associated with such a job title [21]. Respondents who reported they had never worked, or were a housewife, were assigned to the lowest occupational skill level.

Health-related quality of life (HRQL)

HRQL was assessed using the EQ-5D [3] and SF-36 questionnaires [16]. The EQ-5D questionnaire was developed by the EuroQol group, and has two components—five questions concerned with particular health dimensions and a visual analogue scale/thermometer. In the former respondents are asked to report the level of problems they have (no problems, some/moderate problems, and severe/extreme problems) with regard to mobility, self-care, usual

activities, pain/discomfort, and anxiety/depression, the responses to which can then be mapped to one of 243 health states. This provides a measure of HRQL, as an individual utility score (referred to as the EQ-5D_{index}) can be attached to each of these 243 health states. In the UK, based on preferences elicited from a survey of 3395 residents using the time trade-off (TTO) technique [29], regression analysis was used to estimate the utility scores for the EQ-5D_{index} which ranged between −0.594 (33333—the lowest level on all five dimensions) and 1 (11111—the highest level on all five dimensions) [30]. Within the second (EQ VAS) component of the EQ-5D questionnaire people are asked to indicate how good or bad their health state is (on the day they complete the questionnaire), on a visual analogue scale where 0 corresponds to worst imaginable health state and 100 to best imaginable health state [3].

Responses to 11 of the questions on the SF-36 questionnaire [16] can be used to estimate a score on the SF-6D [4], though it is not always necessary for all 11 questions to be fully completed in order for an SF-6D score to be calculated (see Gerard et al. [31] for further discussion). The SF-6D is composed of six dimensions (physical functioning, role limitations, social functioning, pain, mental health, and vitality), each of which having between four and six levels. Based on the preferences elicited from 611 UK residents, using the standard gamble (SG) technique [32], regression analysis was used to estimate utility scores for each of the SF-6D health states [4]. We estimated SF-6D scores using the consistent [33] version of the SF-6D algorithm (see Brazier et al. [34] for further details), upon which utility scores are estimated to range between 0.296 (645655) and 1 (111111).

By completing the EQ-5D and SF-36 questionnaires three measures of HRQL can thereby be estimated—the EQ-5D_{index}, the EQ VAS, and SF-6D. Utility scores are provided by the EQ-5D and SF-6D, whereas the EQ VAS summarizes HRQL on a 0–100 scale.

Analysis

Descriptive statistics

In order to compare the HRQL of different groups of individuals continuous variables were converted into categorical variables, e.g., people were categorized into one of four age groups: (1) 45 to <55 years (2) 55 to <65 years, (3) 65 to <75 years, and (4) ≥75 years. Descriptive statistics concerning the six sociodemographic factors (age, gender, ethnicity, smoking status, BMI, and occupational skill level), and 10 clinical conditions (back pain, hip pain, knee pain heart disease, stroke, asthma, cancer, diabetes, rheumatoid arthritis, and osteoarthritis) were then computed for all respondents in each patient group. Based on

the responses from all individuals, for each of the EQ-5D_{index}, SF-6D, and EQ VAS the range of scores, the median, 25th and 75th percentiles, the mean score, and associated standard deviation, were also calculated. The latter two summary measures were also estimated for each group of individuals with different sociodemographic factors, and those with and without each of the 10 clinical conditions, where the mean scores were concentrated on as these are used in analyses which seek to identify interventions which provide the maximum health gain that can be obtained from a given level of resource use [22, 35].

Discriminative ability

To assess whether each of the three HRQL measures could discriminate between groups of individuals with different sociodemographic factors, and those with and without clinical conditions, the *t*-test was conducted. We first assessed whether different groups had statistically significant ($p < 0.05$) different mean HRQL scores (we refer to these as unadjusted comparisons). Moreover, the independent sample *t*-test was conducted both when there were only two groups of individuals (e.g., male versus female gender), and when comparing the mean HRQL scores for more than two different groups (e.g., BMI underweight, normal BMI, overweight, and obese), where in the latter case pairwise multiple comparisons were conducted between all other groups and the reference group (the group that was expected to have the highest HRQL, e.g., normal BMI). Use of the *t*-test requires the data to be approximately normally distributed. To test whether this was the case the Kolmogorov–Smirnov *Z* test was conducted for each of the three HRQL measures, where a significant *p*-value would lead one to reject the null hypothesis that the data were normally distributed. As it has been shown that responses to the EQ-5D_{index} do not tend to be normally distributed [31], for the EQ-5D_{index} and other measures that failed the normality test, the Mann–Whitney *U* test was also conducted, where again pairwise multiple comparisons were conducted between the HRQL of all other groups and the reference group. Both the *t*-test and Mann–Whitney *U* test were conducted as it has been suggested that further research should be carried out in order to assess the generalizability of the conclusion by Walters and Campbell [36] that nonparametric methods produce similar results to those of conventional statistical methods (i.e., the *t*-test and linear regression), and that the latter are thereby robust to the violation of assumptions that HRQL data are likely to cause (i.e., non-normality).

In order to assess whether the differences in HRQL between different groups of individuals was due to a particular sociodemographic factor, or clinical condition, rather than the effect of confounding factors, multiple linear

regression analysis was also conducted. Three separate regression analyses were conducted where each of the EQ-5D_{index}, SF-6D and EQ VAS acted as the dependent variable, in each regression analysis the six sociodemographic factors and 10 clinical conditions acted as explanatory variables. Again evidence of the ability to discriminate between different groups was considered to be provided when there was a significant difference between the HRQL of different groups of individuals, where this would be the case for regression coefficients which were significantly different from zero.

Within each regression analysis the constant term in each of the three regressions should be interpreted as the estimated HRQL of people in the reference group of each of the sociodemographic factors and clinical conditions, i.e., those who are aged 45–55 years, male, white, have never smoked regularly, have a normal BMI, the highest occupational skill level, and none of the 10 clinical conditions. Additionally, the use of alternative forms of each of the continuous explanatory variables was explored (e.g., age was included as a transformed continuous variable, rather than a categorical variable) in an attempt to identify the model which was best able to explain variation in the dependent variable, whilst (where possible) still trying to satisfy the standard assumptions of linear regression analysis (see Wooldridge [37] for further details of these). Where necessary, transformations of the dependent variables were also explored in an attempt to meet these standard assumptions.

Minimally important difference

As was pointed out in the Introduction it can be argued that the size of the mean difference between the HRQL of different groups of individuals is more important than whether such differences are actually statistically significant. Thus, we also assessed whether the size of the mean differences in each of the regression analyses could be considered to constitute an MID, i.e., whether the mean differences were greater than our previously inferred MID of >0.03 for the EQ-5D_{index} and SF-6D, and >3.0 for the EQ VAS. This, for instance, enables one to assess whether there is a MID between the HRQL of those with and without knee pain, after adjusting for factors that may differ between these groups of individuals.

Finally, in order to estimate how the differences in HRQL differ according to the two measures of utility (i.e., how the differences vary across different sociodemographic factors and clinical conditions) we conducted a further linear regression. As in the previous regression analyses the sociodemographic factors and clinical conditions acted as explanatory variables, but here the dependent variable was the difference between the EQ-5D_{index} and

SF-6D score (EQ-5D_{index} minus SF-6D). Within this analysis a significant difference (i.e., an estimated regression coefficient significantly different from zero) would mean that the difference in HRQL between two particular groups was estimated to be significantly higher for either the EQ-5D_{index} or the SF-6D.

Results

Participants

There were 6,765 people registered with the general practice on 1 July 2004; 3,122 were aged ≥ 45 years, and 2,770 were sent an ascertainment questionnaire. Of these, 1,865 (67.3% of those who were sent one) returned the ascertainment questionnaire to the LIKP study team at the University of Nottingham. However, not all of the 1,865 respondents answered all questions—for the three HRQL measures the level of missing data ranged between 6.9% for the EQ-5D_{index} and 13.6% for the SF-6D (see Table 1), compared to between 4.3% (gender) and 7.2% (BMI) for the sociodemographic factors, and up to 6.1% (hip pain) for the clinical conditions (Table 2). All subsequently reported estimates, including percentage calculations, are based on those who answered the relevant question.

Analysis

Descriptive statistics

The mean age of respondents was 64.7 years (range 45–99 years), and the mean BMI was 26.0 kg/m² (range 14.3–51.6 kg/m²); 55.2% were female, 97.6% were of white ethnicity, 46.5% had smoked regularly for a period of at least 3 months, 32.5% were classified as having the highest occupational skill level, and 10.1% (2.6% of whom reported they had never worked or were a housewife) were classified as having the lowest occupational skill level (Table 2). With regard to the 10 clinical conditions 23.9% were classified as having knee pain, 18.5% back pain,

11.8% osteoarthritis, 10.9% heart disease, 10.8% hip pain, 6.7% asthma, 6.0% diabetes, 5.4% cancer, 5.1% rheumatoid arthritis, and 3.3% stroke (Table 2).

Summary scores for the three HRQL measures are shown in Table 1. It can be seen that the mean score was slightly higher on the EQ-5D_{index}, compared to the SF-6D, whereas the median score was slightly higher on the SF-6D. This result can be explained by the fact that the EQ-5D_{index} scores were more negatively skewed (see Table 1), and by the fact that 31.4% of respondents had a utility score of 1 according to the EQ-5D_{index}, compared to 4.3% on the SF-6D. With regard to the EQ VAS the mean score was 75.54, and 1.2% reported a score of 100.

Discriminative ability

In Table 2 the mean HRQL scores, for each of the EQ-5D_{index}, SF-6D and EQ VAS, are presented for groups of individuals with different sociodemographic factors, and those with and without the clinical conditions. Across all three HRQL measures, for each sociodemographic factor and each clinical condition it can be seen that, as hypothesized, those people in the each of the reference groups (i.e., younger age, male, white ethnicity, never smoked regularly, normal BMI, highest occupational skill level, and no clinical conditions) had a higher mean HRQL. However, we judged discriminative ability to mean that at least one group of individuals had a significantly ($p < 0.05$) lower mean HRQL than those in the reference group. On each of the three HRQL measures, this criteria was met for groups who were older (compared to younger respondents), obese (compared to normal BMI), had a lower occupational skill level (compared to those with the highest occupational skill level), and for all groups who had a certain clinical condition (compared to those who did not), with the exception of cancer. The criteria was also met for females, compared to males (on the EQ-5D_{index} and SF-6D), non-white ethnicity, compared to white ethnicity (EQ VAS), those who have smoked regularly, compared to those that have not (EQ-5D_{index} and EQ VAS), underweight BMI, compared to normal BMI (SF-6D), and for cancer (EQ VAS).

Table 1 Summary scores for each of the EQ-5D_{index}, SF-6D, and EQ VAS

	EQ-5D _{index}	SF-6D	EQ VAS
Number of respondents (% missing)	1,737 (6.9%)	1,612 (13.6%)	1,717 (7.9%)
Mean score (standard deviation)	0.778 (0.239)	0.765 (0.148)	75.54 (17.08)
95% confidence interval	0.772–0.795	0.761–0.775	74.90–76.58
Median score	0.796	0.806	80.00
25–75 percentile	0.691–1.000	0.651–0.886	50.00–90.00
Range	–0.355 to 1.000	0.296 to 1.000	0.00 to 100.00
Kolmogorov–Smirnov Z test	7.347‡	4.558‡	5.970‡
Skewness	–1.738	–0.501	–1.076

* $p < 0.05$, † $p < 0.01$,

‡ $p < 0.001$

Table 2 Mean HRQL estimates, and standard deviations (SD), for the EQ-5D_{index}, SF-6D, and EQ VAS for groups of individuals with different sociodemographic factors and those with and without clinical conditions

	<i>N</i>	EQ-5D _{index} mean (SD)	SF-6D mean (SD)	EQ VAS mean (SD)
Age				
≥75 yrs	392	0.696 (0.275)‡	0.700 (0.157)‡	68.59 (18.72)‡
≥65 to < 75 yrs	435	0.773 (0.224)‡	0.767 (0.144)	76.05 (16.13)†
≥55 to <65 yrs	511	0.800 (0.234)	0.787 (0.143)	77.53 (15.97)
≥45 to <55 yrs ^R	430	0.829 (0.206)	0.788 (0.137)	78.45 (16.25)
Gender				
Female	985	0.754 (0.249)‡	0.747 (0.149)‡	75.15 (17.80)
Male ^R	800	0.806 (0.224)	0.785 (0.145)	75.92 (16.17)
Ethnicity				
Non-white	42	0.716 (0.226)†	0.732 (0.149)	67.95 (18.70)†
White ^R	1,726	0.779 (0.240)	0.765 (0.148)	75.65 (17.02)
Smoked regularly				
Yes	829	0.762 (0.244)†	0.760 (0.147)	74.65 (17.14)*
No ^R	954	0.792 (0.234)	0.770 (0.148)	76.36 (16.96)
BMI				
Obese	277	0.695 (0.277)‡	0.708 (0.153)‡	68.84 (18.05)‡
Overweight	625	0.780 (0.236)	0.772 (0.147)	75.43 (17.14)†
Underweight	24	0.760 (0.167)	0.713 (0.142)*	75.13 (17.78)
Normal ^R	804	0.803 (0.223)	0.781 (0.142)	77.80 (15.99)
Occupational skill level				
1 (Lowest)	176	0.727 (0.281)‡	0.731 (0.158)‡	71.52 (18.29)‡
2	589	0.755 (0.240)‡	0.753 (0.148)‡	74.98 (16.99)‡
3	410	0.762 (0.248)‡	0.760 (0.155)†	74.48 (17.67)‡
4 (Highest) ^R	567	0.834 (0.197)	0.794 (0.134)	78.84 (15.15)
Back pain				
Yes	325	0.577 (0.295)‡	0.656 (0.153)‡	64.96 (19.55)‡
No ^R	1428	0.826 (0.195)	0.791 (0.135)	78.06 (15.39)
Hip pain				
Yes	189	0.538 (0.309)‡	0.666 (0.149)‡	67.14 (20.66)‡
No ^R	1562	0.810 (0.210)	0.796 (0.141)	78.19 (15.92)
Knee pain				
Yes	424	0.610 (0.270)‡	0.666 (0.147)‡	67.14 (19.69)‡
No ^R	1351	0.831 (0.202)	0.797 (0.134)	78.19 (15.28)
Heart disease				
Yes	203	0.678 (0.283)‡	0.706 (0.152)‡	65.58 (19.28)‡
No ^R	1662	0.791 (0.229)	0.773 (0.146)	76.79 (16.37)
Stroke				
Yes	62	0.612 (0.318)‡	0.664 (0.162)‡	65.83 (20.01)‡
No ^R	1803	0.784 (0.234)	0.769 (0.146)	75.88 (16.87)
Asthma				
Yes	125	0.724 (0.271)*	0.715 (0.143)‡	69.95 (19.02)‡
No ^R	1740	0.782 (0.236)	0.769 (0.148)	75.97 (16.85)
Cancer				
Yes	100	0.744 (0.280)	0.739 (0.162)	70.82 (21.61)*
No ^R	1765	0.780 (0.236)	0.767 (0.147)	75.82 (16.74)
Diabetes				
Yes	111	0.673 (0.283)‡	0.715 (0.171)†	64.56 (19.01)‡
No ^R	1754	0.785 (0.234)	0.769 (0.146)	76.27 (16.70)
Rheumatoid arthritis				
Yes	95	0.568 (0.283)‡	0.636 (0.155)‡	61.26 (20.08)‡
No ^R	1770	0.790 (0.231)	0.773 (0.144)	76.33 (16.55)
Osteoarthritis				
Yes	220	0.607 (0.269)‡	0.675 (0.150)‡	65.87 (17.83)‡
No ^R	1645	0.802 (0.224)	0.778 (0.143)	76.92 (16.52)

N refers to the number of responding individuals within each group, for those who completed the respective part of the questionnaire. Results of the *t*-test show which groups had statistically significant different mean HRQL scores, where^R denotes the reference group (the group to which others were compared) ($p \leq 0.05$, $p < 0.01$, $p \leq 0.001$). The *p*-values for the Mann–Whitney *U* test are also reported (* $p < 0.05$, † $p < 0.01$, ‡ $p < 0.001$)

The significance values for the Mann–Whitney U test are shown in Table 2. It can be seen that of the 66 pairwise comparisons there was only one comparison which was found to be significant ($p < 0.05$) according to the t -test, but not the Mann–Whitney U test (the t -test estimated that, according to the SF-6D, the HRQL of those aged ≥ 65 to < 75 years was significantly different to those aged ≥ 45 to < 55 years), and two comparisons which were found to be significant according to the Mann–Whitney U test, but not the t -test (the Mann–Whitney U test estimated that, according to the EQ-5D_{index}, those who were aged ≥ 55 to < 65 yrs had a significantly different HRQL to those who were aged ≥ 45 to < 55 years, and that those of non-white ethnicity had a significantly different HRQL to those of white ethnicity).

Results of the linear regression analyses, conducted to assess whether each of the EQ-5D_{index}, SF-6D and EQ VAS could discriminate between respondents with different sociodemographic factors and those with and without clinical conditions, after adjusting for other factors, are presented in Table 3. Overall the configuration of six sociodemographic factors, and 10 clinical conditions, were able to explain between 25.1% (EQ VAS) and 34.7% (EQ-5D_{index}) of the variation in HRQL (Table 3). Additionally, it was estimated that people with the most favorable levels on each of the sociodemographic factors, and none of the clinical conditions, would have a mean score of 0.957 on the EQ-5D_{index}, 0.859 on the SF-6D, and 84.86 on the EQ VAS. Relative to these scores, each of the three HRQL measures, estimated that there was a significant loss in HRQL for groups who were older (e.g., on the EQ-5D_{index} the average score for those who were aged > 75 was 0.052 lower than those aged ≥ 45 to < 55 years), obese, had a lower occupational skill level, back pain, hip pain, knee pain, heart disease, rheumatoid arthritis or osteoarthritis. There was also a significant loss in HRQL for groups who were female (according to the EQ-5D_{index} and SF-6D), smoked regularly (EQ-5D_{index}), or had had a stroke (EQ-5D_{index} and SF-6D), asthma (SF-6D and EQ VAS), cancer (EQ VAS), or diabetes (EQ-5D_{index} and EQ VAS).

These regression analyses did however show, particularly for the EQ-5D_{index}, that some of the standard assumptions of linear regression analysis were not met. For example, we found that rather than the residuals having a normal distribution they had a negative skewness, and that the Breusch–Pagan test (F statistic) suggested heteroskedasticity to be present (see Table 3). That said, the values of the variance inflation factor (VIF) suggested that none of the explanatory variables were perfectly related to one another. Due to the negative skewness of the residuals we transformed the EQ-5D_{index} score by adding a value of 0.59 to it (to ensure it was a positive value), and then raising it to the power of three: $(EQ-5D_{index} + 0.59)^3$. When this new

variable was used as the dependent variable in the regression the adjusted R^2 improved, the new residuals had a lower level of skewness, and there was no evidence of heteroskedasticity (see square brackets in Table 3). Additionally, with one exception, the explanatory variables which were statistically significant within this regression analysis were the same explanatory variables as those which were statistically significant within the regression when the untransformed EQ-5D_{index} variable acted as the dependent variable. The exception was the variable of occupation skill level, where when the dependent variable was $(EQ-5D_{index} + 0.59)^3$, those with the lowest skill level (1) were not estimated to have a significantly lower HRQL than those with the highest skill (4) after adjusting for other factors (see Table 3). Conversely, those with a skill level of 2 were estimated to have a significantly lower HRQL than those with the highest skill when the dependent variable was $(EQ-5D_{index} + 0.59)^3$, whereas this had not been the case when the untransformed EQ-5D_{index} variable acted as the dependent variable. This robustness therefore increases the level of confidence that one can have in the aforementioned results where the EQ-5D_{index} acted as the dependent variable.

Minimally important difference

Within each of the regression analyses differences in HRQL which we consider to constitute an MID are shown by a tick (\surd) in Table 4. It can be seen that the deficit in HRQL associated with older age, obese BMI, lower occupational skill level, back pain, hip pain, knee pain, heart disease, stroke, asthma, rheumatoid arthritis, and osteoarthritis were deemed to be a MID on each of the three HRQL measures as the differences in HRQL (between these groups and the respective reference groups) were > 0.03 on both the EQ-5D_{index} and SF-6D, and > 3.0 on the EQ VAS. The deficit in HRQL was also considered to be an MID for those of non-white ethnicity (according to the EQ VAS), underweight BMI (EQ-5D_{index} and SF-6D), cancer (SF-6D and EQ VAS), and diabetes (EQ-5D_{index} and EQ VAS). Thus, after adjusting for other factors, none of the three measures estimated that the deficit in HRQL associated with gender or smoking status constituted an MID.

Finally, we used linear regression analysis to assess whether there was a significant difference between (1) the loss in HRQL, for a particular group, according to the EQ-5D_{index}, and (2) the loss in HRQL for the same group according to the SF-6D. The results in Table 5 show that when the EQ-5D_{index} score minus the SF-6D score acted as the dependent variable most of the explanatory variables had a negative value. Such a negative value would indicate that the loss in HRQL was estimated to be higher according

Table 3 Parameter estimates (95% confidence intervals) of the explanatory variables used to estimate variation in the EQ-5D_{index}, SF-6D, and EQ VAS (* $p < 0.05$, † $p < 0.01$, ‡ $p < 0.001$)

Explanatory variable	EQ-5D _{index} [(EQ-5D _{index} +0.59) ³ Estimate (95% CI)]	SF-6D Estimate (95% CI)	EQ VAS Estimate (95% CI)
Constant	0.957 (0.929 to 0.985)‡	0.859 (0.841 to 0.878)‡	84.86 (82.74 to 86.98)‡
Age			
>75 versus ≥45 to <55 years	-0.052 (-0.082 to -0.021)‡	-0.042 (-0.063 to -0.022)‡	-4.10 (-6.41 to -1.79)‡
≥65 to <75 versus ≥45 to <55 years	-0.011 (-0.039 to 0.017)	0.008 (-0.011 to 0.026)	1.24 (-0.87 to 3.35)
≥55 to <65 versus ≥45 to <55 years	-0.013 (-0.039 to 0.012)	0.008 (-0.009 to 0.025)	0.59 (-1.37 to 2.54)
Gender			
Female versus male	-0.029 (-0.050 to -0.009)†	-0.021 (-0.035 to -0.007)†	0.92 (-0.67 to 2.51)
Ethnicity			
Non-white versus white	-0.014 (-0.084 to 0.057)	-0.006 (-0.054 to 0.041)	-5.30 (-10.64 to 0.04)
Smoked regularly			
Yes versus no	-0.028 (-0.047 to -0.009)†	-0.007 (-0.020 to 0.005)	-0.59 (-2.05 to 0.87)
BMI			
Obese versus normal	-0.038 (-0.066 to -0.009)*[†]	-0.038 (-0.057 to -0.020)‡	-4.70 (-6.90 to -2.50)‡
Overweight versus normal	-0.006 (-0.027 to 0.015)	0.000 (-0.014 to 0.014)	-1.16 (-2.78 to 0.46)
Underweight versus normal	-0.031 (-0.114 to 0.051)	-0.065 (-0.120 to -0.010)*	-2.55 (-8.69 to 3.59)
Occupational skill level			
Lowest (1) versus highest (4)	-0.039 (-0.074 to -0.004)*[ns]	-0.031 (-0.055 to -0.007)*	-3.87 (-6.55 to -1.19)‡
Skill 2 versus highest	-0.022 (-0.046 to 0.003)*[†]	-0.007 (-0.024 to 0.009)	-1.64 (-3.50 to 0.23)
Skill 3 versus highest	-0.046 (-0.071 to -0.020)‡	-0.020 (-0.037 to -0.003)*	-2.49 (-4.42 to -0.55)*
Back pain			
Yes versus no	-0.155 (-0.182 to -0.128)‡	-0.081 (-0.099 to -0.063)‡	-7.69 (-9.71 to -5.66)‡
Hip pain			
Yes versus no	-0.110 (-0.144 to -0.076)‡	-0.046 (-0.069 to -0.023)‡	-5.62 (-8.21 to -3.03)‡
Knee pain			
Yes versus no	-0.111 (-0.136 to -0.087)‡	-0.070 (-0.086 to -0.053)‡	-4.12 (-5.97 to -2.28)‡
Heart disease			
Yes versus no	-0.052 (-0.083 to -0.021)‡[†]	-0.031 (-0.052 to -0.011)†	-6.19 (-8.55 to -3.84)‡
Stroke			
Yes versus no	-0.074 (-0.129 to -0.019)†[*]	-0.065 (-0.103 to -0.028)‡	-3.41 (-7.59 to 0.78)
Asthma			
Yes versus no	-0.032 (-0.069 to 0.006)	-0.045 (-0.070 to -0.021)‡	-4.64 (-7.45 to -1.84)†
Cancer			
Yes versus no	-0.024 (-0.065 to 0.018)	-0.036 (-0.045 to 0.013)	-5.29 (-8.47 to -2.11)†
Diabetes			
Yes versus no	-0.058 (-0.098 to -0.018)†	-0.019 (-0.046 to 0.008)	-7.19 (-10.29 to -4.10)‡
Rheumatoid arthritis			
Yes versus no	-0.102 (-0.144 to -0.059)‡	-0.067 (-0.097 to -0.038)‡	-7.97 (-11.27 to -4.68)‡
Osteoarthritis			
Yes versus no	-0.079 (-0.109 to -0.049)†[‡]	-0.041 (-0.061 to -0.021)‡	-5.45 (-7.73 to -3.17)†
Adjusted R ² squared (N)	0.347 (1,568) [0.363]	0.298 (1,471)	0.251 (1,556)
Skewness of residuals	-1.622 [-0.622]	-0.329	-0.861
Breusch-Pagan test (F statistic)	2.006‡ [0.986]	1.282‡	1.561‡
Variance inflation factor (range)	1.017–1.677	1.018–1.618	1.017–1.649

Signs indicating the level of significance are also shown for the dependent variable (EQ-5D_{index} + 0.59)³ when the significance level differed from that for the EQ-5D_{index} [see square brackets]

NS Nonsignificant

Table 4 Parameter estimates which were considered to constitute a minimally important difference (MID) i.e. >0.03 for the EQ-5D_{index} and SF-6D, and >3.0 for the EQ VAS

Explanatory variable	EQ-5D _{index}	SF-6D	EQ VAS
Age			
>75 versus ≥45 to <55 years	√	√	√
≥65 to <75 versus ≥45 to <55 years			
≥55 to <65 versus ≥45 to <55 years			
Gender			
Female versus male			
Ethnicity			
Non-white versus white			√
Smoked regularly			
Yes versus No			
BMI			
Obese versus normal	√	√	√
Overweight versus normal			
Underweight versus normal	√	√	
Occupational skill level			
Lowest (1) versus highest (4)	√	√	√
Skill 2 versus highest			
Skill 3 versus highest	√		
Back pain			
Yes versus no	√	√	√
Hip pain			
Yes versus no	√	√	√
Knee pain			
Yes versus no	√	√	√
Heart disease			
Yes versus no	√	√	√
Stroke			
Yes versus no	√	√	√
Asthma			
Yes versus no	√	√	√
Cancer			
Yes versus no			√
Diabetes			
Yes versus no	√		√
Rheumatoid arthritis			
Yes versus no	√	√	√
Osteoarthritis			
Yes versus no	√	√	√

A tick (√) denotes that there was an MID in HRQL between the two groups of individuals

to the EQ-5D_{index} than the SF-6D. Additionally, the significant values indicate that the loss in HRQL associated with certain age and occupational skill level groups, and back pain, hip pain, knee pain, heart disease, diabetes, and osteoarthritis was estimated to be significantly greater according to the EQ-5D_{index}. That said it should be noted

that only 13.6% of the variation in the difference in HRQL between the two instruments can be explained by the six sociodemographic factors and 10 clinical conditions (Table 5).

Discussion

When undertaking both adjusted and unadjusted comparisons we have demonstrated that each of the EQ-5D_{index}, SF-6D, and EQ VAS estimated that there was a significant difference in HRQL between most groups with different sociodemographic factors and those with and without clinical conditions (see Tables 2 and 3). Indeed, after adjusting for other factors, each of the three measures was able to discriminate between the HRQL of those groups who had different ages, BMI, occupational skill level and those with and without back pain, hip pain, knee pain, heart disease, rheumatoid arthritis, and osteoarthritis. However, not all of the three measures could discriminate between the HRQL of groups of individuals with different genders, ethnicity, smoking status, underweight BMI, or those with and without stroke, asthma, cancer, or diabetes. By comparing the size of the mean HRQL differences between groups to the inferred MID it can also be seen that most of the differences associated with the clinical conditions would be deemed important, but that, after adjusting for other factors, this would not be the case for many of the sociodemographic factors (see Table 4). Finally, there was also a tendency for the mean deficit in HRQL associated with different clinical conditions to be estimated to be higher according to the EQ-5D_{index} than the SF-6D (see Table 5).

Comparisons with other studies

Our results are in agreement with other studies, which have shown that HRQL measures are generally able to discriminate between the HRQL of patient groups that differ in terms of sociodemographic factors, or the presence of clinical conditions [6–8, 10, 12–15, 17–20]. For example, for the EQ-5D_{index}, our results (shown in brackets), taken from Table 3, are highly comparable to two recent studies [14, 15]. After adjusting for age, gender, ethnicity, and clinical conditions, Lubetkin et al. [15] estimated that the mean difference in HRQL to be –0.060 for those aged ≥70 years, compared to those aged 18–39 years (>75 compared to ≥45 to <55 years = –0.052), –0.023 for females, compared to males (–0.029), up to –0.032 for those of non-white ethnicity (–0.014), –0.047 for diabetes (–0.058), –0.040 for asthma (–0.032), –0.064 for heart disease (–0.052) and –0.063 for stroke (–0.074). Similarly, after adjusting for age, gender, ethnicity, income, and

Table 5 Parameter estimates of the explanatory variables used to estimate variation in the difference between the EQ-5D_{index} and SF-6D score (* $p < 0.05$, † $p < 0.01$, ‡ $p < 0.001$)

Explanatory variable	Estimate (95% confidence interval)
Constant	0.096 (0.073 to 0.118)‡
Age	
>75 versus ≥45 to <55 years	-0.008 (-0.033 to -0.017)
≥65 to <75 versus ≥45 to <55 years	-0.026 (-0.049 to 0.003)*
≥55 to <65 versus ≥45 to <55 years	-0.018 (-0.039 to -0.003)
Gender	
Female versus male	-0.003 (-0.020 to 0.014)
Ethnicity	
Non-white versus white	-0.010 (-0.067 to 0.048)
Smoked regularly	
Yes versus no	-0.015 (-0.031 to -0.001)
BMI	
Obese versus normal	-0.002 (-0.025 to 0.022)
Overweight versus normal	-0.002 (-0.020 to 0.015)
Underweight versus normal	0.031 (-0.038 to 1.000)
Occupational skill level	
Lowest (1) versus highest (4)	-0.019 (-0.049 to 0.010)
Skill 2 versus highest	-0.018 (-0.039 to 0.002)
Skill 3 versus highest	-0.026 (-0.047 to -0.005)*
Back pain	
Yes versus no	-0.079 (-0.101 to -0.057)‡
Hip pain	
Yes versus no	-0.056 (-0.084 to -0.028)‡
Knee pain	
Yes versus no	-0.039 (-0.0059 to -0.019)‡
Heart disease	
Yes versus no	-0.026 (-0.052 to -0.001)*
Stroke	
Yes versus no	-0.022 (-0.068 to 0.025)
Asthma	
Yes versus no	0.014 (-0.016 to 0.044)
Cancer	
Yes versus no	0.006 (-0.029 to 0.041)
Diabetes	
Yes versus no	-0.037 (-0.070 to -0.004)*
Rheumatoid arthritis	
Yes versus no	-0.032 (-0.068 to 0.004)
Osteoarthritis	
Yes versus no	-0.042 (-0.067 to -0.017)‡
Adjusted R^2 squared (N)	0.136 (1,328)

education, Sullivan et al. [14] estimated that the mean difference to be -0.055 for congestive heart failure, -0.034 for acute myocardial infarction, -0.048 for other ill-defined heart disease (heart disease = -0.052), -0.0207 for asthma (-0.032), -0.035 for diabetes mellitus without complication

(diabetes = -0.058), -0.0610 for osteoarthritis (-0.079), and -0.085 for rheumatoid arthritis and related disease (rheumatoid arthritis = -0.102).

Implications

Amongst other things, we have shown that there are small, but systematic, differences in the estimated deficit in HRQL, associated with certain sociodemographic factors and clinical conditions, as estimated by the EQ-5D_{index} and SF-6D (see Tables 3 and 5). One implication of this is that there is the potential for the estimated improvement in HRQL, associated with an intervention, to vary according to the HRQL instrument with which it is measured, as has been shown previously [9, 38–41]. Thus, there is also the potential for the estimated cost-effectiveness of an intervention to vary according to the HRQL measure that is used, and, in light of this, others have argued that estimates of cost-effectiveness which are based on different HRQL measures should not be compared [42], or that they should only be compared with extreme caution [38].

By comparing the estimated deficit in HRQL between patient groups to the MID we have additionally shown that the deficit associated with all clinical conditions, with the exception of cancer (see next section for further discussion), can be considered important. Conversely, in line with the results of Lubetkin et al. [15], this was not the case for differences between different genders. That said, it has been pointed out that preference-based measures were primarily designed to inform resource allocation decisions, and that therefore changes in costs need to be considered, as well as changes in HRQL [43].

Limitations

There were two potential deficiencies in the wording of the ascertainment questionnaire which may have affected the reliability of the self-reported data, though we believe that only the second potential deficiency may have affected the results of this study. The first potential deficiency was that people were simply asked to tick the respective box if they had been diagnosed with heart disease, stroke, asthma, cancer, diabetes, rheumatoid arthritis or osteoarthritis. Thus, in all circumstances, we infer a non-response to mean that they had not been diagnosed with the respective condition (i.e., the level of missing data was recorded as zero for these clinical conditions in Table 2). Such an approach does however leave open the possibility that some of these non-responses should have been categorized as missing (as the respondent did not consider the question). Thus, in a small number of cases, some respondents may have been categorized as not having the clinical condition, when in fact they had been diagnosed with a clinical

condition but did not answer this particular question. The impact of this miscategorization would have been to understate the differences between groups of individuals, but as the HRQL measures were still generally able to discriminate between the different groups of individuals we do not believe this impinged upon the results of this study. Moreover, the impact of this oversight is minimized when we seek to identify an MID as only those who completed data on all other clinical conditions and sociodemographic factors were included in the linear regression analyses. The second similar oversight is that we asked people to report if they had been diagnosed by their doctor as having cancer. At least two respondents (we only began to monitor this part way through the data entry procedure) ticked the box to say they had been diagnosed with cancer, but in addition noted that it had now been ‘cured’. Thus, as well as the fact that there will have been a large variety of cancers for those included in the cancer category, some may also not have been affected by the condition when they completed the questionnaire. One potential implication of this is that, as those in the cancer and non-cancer categories may be more similar than intended, this may explain why neither the EQ-5D_{index} nor SF-6D could discriminate between the HRQL of these patient groups.

Though the *t*-test and regression analysis are commonly used to analyze responses to the EQ-5D_{index} [13–15, 44, 45] there may be some limitations with regard to the statistical analysis performed within this paper. These largely arise because the data for each of the HRQL measures were not normally distributed (a requirement for the *t*-test) and some of the standard assumptions of regression analysis were not met. That said, we demonstrated that the results of the *t*-test were robust as the qualitative interpretation of these results (in relation to the $p < 0.05$ cut-off) was virtually identical to those obtained using the Mann–Whitney *U* test. Similarly, when a transformation of the EQ-5D_{index} was used as the dependent variable, instead of the EQ-5D_{index} itself, more of the assumptions of regression analysis were met, whilst the qualitative interpretation of all but one explanatory variable remained unchanged. Thus our results support the conclusion, of Walters and Campbell [36], that the *t*-test and linear regression are robust to the violation of assumptions that HRQL data are likely to cause.

Further potential limitations include those that are common to cross-sectional subpopulation studies of this nature. These include the fact that causality is generally easier to infer in a longitudinal study [46], that the use of self-reported data may result in biased estimates of the prevalence of clinical conditions [47], and that participants in medical research tend to be of a higher social class than nonparticipants [48, 49]. Indeed the latter of these may explain why the number of non-white people may have

been underrepresented in our study (see Table 2 for further details). Some protection against reporting spurious effects is however provided by the large sample size of our study, and by using regression analysis to control for a number of potentially confounding variables. Finally, it should be noted that, in this paper, we have only assessed the discriminative ability of the three HRQL measures with regard to specific sociodemographic factors and clinical condition. Thus, as well as providing only information on a property which contributes towards construct validity we are also unable to comment on the performance of these measures with regard to other important validity and reliability criteria (see [2], [50] for a discussion of these).

Conclusion

We have demonstrated that each of the EQ-5D_{index}, SF-6D, and EQ VAS was able to discriminate between the HRQL of the majority of groups of individuals with different sociodemographic factors, and those with and without clinical conditions. However, for groups of different gender, ethnicity, smoking status, or BMI or those with and without stroke, asthma, cancer, or diabetes not all of the three HRQL measures were able to discriminate between the HRQL of these groups of individuals. Finally, it was shown that the mean deficit in HRQL associated with many clinical conditions was estimated to be higher according to the EQ-5D_{index} than the SF-6D. This, in turn, implies that the benefits of alleviating such conditions, and thereby the cost-effectiveness of different interventions, will also tend to be more favorable according to the EQ-5D_{index} compared to the SF-6D. Further research is needed to confirm whether this is the case, which conditions it applies to, and whether cost-effectiveness estimates which are based on different instruments need to be adjusted accordingly.

Acknowledgments We thank all those who completed the Lifestyle Interventions for Knee Pain (LIKIP) study questionnaire. The LIKP study was funded by the UK Arthritis Research Campaign (ARC) (grant number 13550). PhD funding for Garry Barton was provided by the UK Economic & Social Research Council (ESRC) (PTA-037-2004-00051).

References

1. Streiner, D.L., Norman, G.R.: Health Measurement Scales: A Practical Guide to Their Development and Use (3rd edn). Oxford University Press, New York (2003)
2. Brazier, J.E., Deverill, M., Green, C., Harper, R., Booth, A.: A review of the use of health status measures in economic evaluation. *Health Technol. Assess.* **3**, 1–164 (1999)
3. Brooks, R.: EuroQol: the current state of play. *Health Policy* **37**, 53–72 (1996)

4. Brazier, J.E., Roberts, J., Deverill, M.: The estimation of a preference-based measure of health from the SF-36. *J. Health Econ.* **21**, 271–292 (2002)
5. Marra, C.A., Esdaile, J.M., Guh, D., Kopec, J.A., Brazier, J.E., Koehler, B.E., Chalmers, A., Anis A.H.: A comparison of four indirect methods of assessing utility values in rheumatoid arthritis. *Med. Care* **42**, 1125–1131 (2004)
6. Petrou, S., Hockley, C.: An investigation into the empirical validity of the EQ-5D and SF-6D based on hypothetical preferences in a general population. *Health Econ.* **14**, 1169–1189 (2005)
7. Macran, S., Weatherly, H., Kind, P.: Measuring population health: a comparison of three generic health status measures. *Med. Care* **41**, 218–231 (2003)
8. Johnson, J.A., Coons, S.J.: Comparison of the EQ-5D and SF-12 in an adult US sample. *Qual. Life Res.* **7**, 155–166 (1998)
9. Lamers, L.M., Bouwmans, C.A.M., van Straten, A., Donker, M.C.H., Hakkaart, L.: Comparison of EQ-5D and SF-6D utilities in mental health patients. *Health Econ.* **15**, 1229–1236 (2006)
10. Johnson, J.A., Pickard, A.S.: Comparison of the EQ-5D and SF-12 health surveys in a general population survey in Alberta, Canada. *Med. Care* **38**, 115–121 (2000)
11. Luo, N., Johnson, J.A., Shaw, J.W., Feeny, D., Coons, S.J.: Self-reported health status of the general adult U.S. population as assessed by the EQ-5D and Health Utilities Index. *Med. Care* **43**, 1078–1086 (2005)
12. Kind, P., Dolan, P., Gudex, C., Williams, A.: Variations in population health status: results from a UK national questionnaire survey. *BMJ* **316**, 736–741 (1998)
13. Lubetkin, E.I., Jia, H., Gold, M.R.: Construct validity of the EQ-5D in low-income Chinese American primary care patients. *Qual. Life Res.* **13**, 1459–1468 (2004)
14. Sullivan, P.W., Lawence, W.F., Ghushchyan, V.: A national catalog of preference-based scores for chronic conditions in the United States. *Med. Care* **43**, 736–749 (2005)
15. Lubetkin, E.I., Jia, H., Franks, P., Gold, M.R.: Relationship among sociodemographic factors, clinical conditions, and health-related quality of life: examining the EQ-5D in the U.S. general population. *Qual. Life Res.* **14**, 2187–2196 (2005)
16. Ware, J.E., Sherbourne, C.: The MOS 36 item short-form health survey: conceptual framework and item selection. *Med. Care* **30**, 473–483 (1992)
17. Brazier, J.E.: The SF-36 health survey questionnaire—a tool for economists. *Health Econ.* **2**, 213–215 (1993)
18. Johnson, J.A., Connolly, M., Zuberbuhler, P., Brown, N.E.: Health-related quality of life for adults with cystic fibrosis: a regression approach to assessing the impact of recombinant human DNase. *Pharmacotherapy* **20**, 1167–1174 (2000)
19. Jia, H., Lubetkin, E.I.: The impact of obesity on health-related quality-of-life in the general adult US population. *J Public Health (Oxf)* **27**, 156–164 (2005)
20. Macran, S.: The relationship between body mass index and health-related quality of life (Discussion Paper 190). University of York: Centre for Health Economics (2004)
21. Office for National Statistics: Standard occupational classification 2000 (Volume 2): the coding index. The Stationery Office, London (2000)
22. Claxton, K.: The irrelevance of inference: a decision-making approach to the stochastic evaluation of health care technologies. *J. Health Econ.* **18**, 341–364 (1999)
23. Jaeschke, R., Singer, J., Guyatt, G.H.: Measurement of health status: ascertaining the minimal clinical important difference. *Control Clin. Trials* **10**, 407–415 (1989)
24. Walters S.J., Brazier J.E.: Comparison of the minimally important difference for two health state utility measures: EQ-5D, SF-6D. *Qual Life Res* **14**, 1523–1532 (2005)
25. Marra C.A., Woolcott J.C., Kopec J.A., Shojania K., Offer R., Brazier J.E., Esdaile J.M., Anis A.H.: A comparison of generic, indirect utility measures (the HUI2, HUI3, SF-6D, the EQ-5D) and disease-specific instruments (the RAQoL and the HAQ) in rheumatoid arthritis. *Soc. Sci. Med.* **60**, 1571–1582 (2005)
26. Walters, S.J., Brazier, J.E.: What is the relationship between the minimally important difference and health state utility values? The case of the SF-6D. *Health Qual. Life Outcomes* **1**, 4 (2003)
27. WHO: Obesity: preventing and managing the global epidemic. Report of a WHO consultation on Obesity. WHO, Geneva (2001)
28. National Institutes of Health: Clinical guidelines on the identification, evaluation, and treatment of overweight and obesity in Adults—the evidence report. *Obes. Res.* **6**, 51S–209S (1998)
29. Torrance, G.W.: Measurement of health state utilities for economic appraisal. *J. Health Econ.* **5**, 1–30 (1986)
30. Dolan, P., Gudex, C., Kind, P., Williams, A.: A social tariff for the EuroQol: results from a UK general population survey (Discussion Paper 138). University of York: Centre for Health Economics (1995)
31. Gerard, K., Nicholson, T., Mullee, M., Mehta, R., Roderick, P.: EQ-5D versus SF-6D in an older, chronically ill patient group. *Appl. Health Econ. Health Policy* **3**, 91–102 (2004)
32. Dolan, P., Gudex, C., Kind, P., Williams, A.: Valuing health states: a comparison of methods. *J. Health Econ.* **15**, 209–231 (1996)
33. Badia, X., Roset, M., Herdman, M.: Inconsistent responses in three preference-elicitation methods for health states. *Soc. Sci. Med.* **49**, 943–950 (1999)
34. Brazier, J.E., Roberts, J., Tsuchiya, A., Busschbach, J.: A comparison of the EQ-5D and SF-6D across seven patient groups. *Health Econ.* **13**, 873–884 (2004)
35. Fenwick, E., Claxton, K., Sculpher, M.J.: Representing uncertainty: the role of cost-effectiveness acceptability curves. *Health Econ.* **10**, 779–787 (2001)
36. Walters, S.J., Campbell, M.J.: The use of bootstrap methods for analysing health-related quality of life outcomes (particularly the SF-36). *Health Qual. Life Outcomes* **2**, 70 (2004)
37. Wooldridge, J.M.: *Introductory Econometrics: A Modern Approach*. Thomson South-Western, Mason (2006)
38. Conner-Spady, B., Suarez-Almazor, M.E.: Variation in the estimation of quality-adjusted life-years by different preference-based instruments. *Med. Care* **41**, 791–801 (2003)
39. Barton, G.R., Bankart, J., Davis, A.C., Summerfield, A.Q.: Comparing utility scores before and after hearing-aid provision: Results according to the EQ-5D, HUI3 and SF-6D. *Appl. Health Econ. Health Policy* **3**, 103–105 (2004)
40. Longworth, L., Bryan, S.: An empirical comparison of EQ-5D and SF-6D in liver transplant patients. *Health Econ.* **12**, 1061–1067 (2003)
41. Pickard, A.S., Johnson, J.A., Feeny, D.H.: Responsiveness of generic health-related quality of life measures in stroke. *Qual. Life Res.* **14**, 207–219 (2005)
42. Lenert, L., Kaplan, R.M.: Validity and interpretation of preference-based measures of health-related quality of life. *Med. Care* **38**(Supplement 2), 138–150 (2000)
43. Drummond, M.: Introducing economic and quality of life measurements into clinical studies. *Ann Med* **33**, 344–349 (2001)
44. Hurst, N.P., Kind, P., Ruta, D., Hunter, M., Stubbings, A.: Measuring health-related quality of life in rheumatoid arthritis: validity, responsiveness and reliability of EuroQol (EQ-5D). *Rheumatology (Oxford)* **36**, 551–559 (1997)
45. Sach, T.H., Barton, G.R.: Interpreting parental proxy reports of (health-related) quality of life for children with unilateral cochlear implants. *Int. J. Pediatr. Otorhinolaryngol.* **17**, 435–445 (2007)

46. Maddigan, S.L., Feeny, D.H., Johnson, J.A.: Health-related quality of life deficits associated with diabetes and comorbidities in a Canadian National Population Health Survey. *Qual. Life Res.* **14**, 1311–1320 (2005)
47. Wu, S.C., Li, C.Y., Ke, D.S.: The agreement between self-reporting and clinical diagnosis for selected medical conditions among the elderly in Taiwan. *Public Health* **114**, 137–142 (2000)
48. Cartwright, A.: Who responds to postal questionnaires? *J. Epidemiol. Community Health* **40**, 267–273 (1986)
49. McColl, E., Jacoby, A., Thomas, L., Soutter, J., Bamford, C., Steen, N., Thomas, R., Harvey, E., Garratt, A., Bond, J.: Design and use of questionnaires: a review of best practice applicable to surveys of health service staff and patients. *Health Technol. Assess.* **5**, 1–256 (2001)
50. Fitzpatrick, R., Davey, C., Buxton, M.J., Jones, D.R.: Criteria for assessing patient based outcome measures for use in clinical trials. *Health Technol. Assess.* **2**(14), 1–74 (1998)