

# Importance of sociodemographic and morbidity aspects in measuring health-related quality of life: performances of three tools

## Comparison of three questionnaire scores

Cecilia Quercioli · Gabriele Messina ·  
Emanuela Barbini · Giovanni Carriero ·  
Mara Fanì · Nicola Nante

Received: 15 April 2008 / Accepted: 14 November 2008  
© The Author(s) 2008. This article is published with open access at Springerlink.com

### Abstract

**Background** Since health-related quality of life (HRQL) measures are numerous, comparisons have been suggested. **Aim** To compare three HRQL measures: SF6D, HUI3 and EQ5D.

**Methods** Three questionnaires (SF36, HUI3, EQ5D) were administered to 1,011 patients attending 16 general practices in two Italian cities. Information about patients' gender, age, education, marital status, smoking, body mass index (BMI) and chronic diseases (hypertension, diabetes, cardiovascular and musculoskeletal diseases) were also collected. Questionnaires scores were calculated using the appropriate algorithms; in particular SF6D scores were obtained from SF36 items. Agreement and correlation between questionnaires scores were investigated using Bland and Altman method and Spearman coefficient. The influence of socio-demographic and morbidity indicators on scores was analysed using the nonparametric quantile regression.

**Results** The Spearman coefficient was about 0.6 for all questionnaires. The 95% limits of agreement of the scores were approximately from -0.5 to 0.3 except for SF6D and EQ5D when they were from -0.4 to 0.2. The measures were influenced by socio-demographic and clinical variables in a similar way, especially SF6D (the index obtained

from SF36) and EQ5D, which appeared to be influenced by the same pattern of factors, including gender, chronic diseases, smoking and BMI.

**Conclusions** Overall, the agreement between questionnaires scores was quite low, whilst the correlation level was good. Questionnaire scores were influenced by socio-demographic and clinical variables in a similar way, especially SF6D and EQ5D. Therefore, the descriptive capacity of SF6D and EQ5D was found to be similar.

**Keywords** Comparison · Euroqol · Health-related quality of life · Health Utility Index · SF-36

### Abbreviations

BMI	Body mass index
CI	Confidence interval
EQ5D	Euroqol 5 Dimensions
GP	General practitioner
HRQL	Health-related quality of life
HUI3	Health Utility Index 3
QALY	Quality adjusted life year
SD	Standard deviation
SF36	Short Form 36
SF6D	Short Form 6 Dimensions

C. Quercioli · G. Messina · E. Barbini · N. Nante (✉)  
Department of Public Health, University of Siena,  
Via Aldo Moro, 53100 Siena, Italy  
e-mail: nante@unisi.it

G. Carriero  
General Practice, Local Health Authority 7, Siena, Italy

M. Fanì  
Primary Health Care Unit, Local Health Authority 2, ASL TO 1,  
Turin, Italy

### Background

Health-related quality of life (HRQL) has become important in fields such as health policy, clinical practice and health outcome evaluation [1–6]. The use of outcome models combining HRQL with morbidity or mortality is increasing. This has led to the development of a series of

measures combining HRQL and traditional outcome indicators, such as QALY [7–9]. HRQL preference/utility-based measures are also increasingly used to obtain QALYs [10, 11].

Since there are many preference/utility-based instruments, there is a general demand for head-to-head comparison studies between different scores [12]. Many of the studies carried out have focussed on the SF6D, a new tool representing the link between psychometric and preference/utility-based measures. Most investigations comparing SF6D with other widely used measures (e.g., EQ5D or HUI) have obtained contradictory findings. In fact, it has been argued that, since these measures have different scoring functions and define a different number of health levels, it is not clear if their results can be compared. SF6D is more sensitive than EQ5D and HUI3 in healthy people, and in detecting small health changes, especially at the top of the scale [11–14]. SF6D also shows a more continuous and normal value distribution than EQ5D and HUI3, which often have very skewed distributions or distributions clustered around a few values. Indeed, EQ5D tends to suffer from clustering and ceiling effect in patients with middle-severity health status [15]. On the other hand, SF6D produces higher values than EQ5D or HUI3 at the lower end of the scale. Thus, SF6D is likely to overestimate the values of very impaired health status [16–22]. Since a recent article [23] reported evidence that SF6D can describe poor health status, it is not yet clear whether SF6D has greater sensitivity and accuracy in the description of health and advantages compared with other widely used, shorter measures, such as EQ5D [20, 24].

EQ5D and HUI3 have only been compared in a limited number of studies. Findings suggest that they give similar results, though HUI3 discriminates better between lower levels of impairment [11, 20, 25, 26].

Although a certain degree of correlation has been found between these measures, which seem to assess a similar essential construct, it is still not clear whether different measures give similar results for HRQL. These previous studies underlined that it is difficult to establish whether one tool is better than another because all have strengths and weaknesses. However, the need for a standard measure is increasingly felt because it has been seen that QALYs calculated with different tools give different results [11, 16, 19]. Further research is also necessary because many of the studies carried out so far suffered from limitations, such as small sample sizes [16, 17, 21], or poor generalisation because of focussing on a single population of patients with a very specific disease [18–20, 22, 26]. Many of these studies [13, 17, 18] only compared two tools, while a broader examination of several different measures could be useful. Finally, few studies have investigated the influence of socio-demographic or morbidity factors on the

performances of different measures [27]. This aspect could be important for two reasons: it could affect the comparability of tools, and it could highlight which tools are most sensitive for distinguishing health differences between populations or groups within populations.

The aim of this study was to compare three distinct HRQL measures, SF6D, HUI3 and EQ5D, in a population not affected by specific diseases, specifically:

- (1) formal agreement and correlation between measures;
- (2) similarities and differences in the way measures are independently related to a range of self-reported socio-demographic and morbidity measures.

## Methods

### Study sample

Cross-sectional data were collected on a sample of patients attending general practices in two Italian cities (Turin and Siena) from May 2003 to April 2004. Turin is a big industrial city (population 1,000,000), whereas Siena is a town (population 50,000) with little industry. They were chosen because a further aim of our study, not described in this paper, was to describe the HRQL of GP patients in different environments for local health policy purposes. All patients attending the general practices during the study period were eligible. The GPs were recruited with the collaboration of the Primary Health Care Unit of the Local Health Authority 2, ASL TO 1 of Turin and the Siena division of the Italian Federation of General Practitioners (Federazione Italiana Medici di Medicina Generale). Seven GPs in Turin and nine in Siena agreed to participate. In the end, 467 patients were enrolled in Turin and 544 in Siena, making a total of 1,011 patients. There were no stated patient exclusion criteria, but children (<16 years old) and severely ill (hospitalised or bed ridden) patients were not recruited due to the type of patients attending the GPs.

### Study procedures

The GPs were recruited, and the licenses to use the questionnaires and the algorithms for calculating the scores were acquired. HRQL data were collected using the Italian version of the three questionnaires: SF36, HUI3 and EQ5D. They were administered sequentially to the patients at the general practices. The patients were informed about the study by the GPs and were asked to participate. Written informed consent was obtained from all who accepted. Refusals were not recorded. The following information was also obtained from each patient: date of birth, education, marital status (married, not married, divorced/widowed),

smoking status (non-smoker, current smoker, ex-smoker), height and weight to calculate BMI, and any history of hypertension, cardiovascular diseases, diabetes or musculoskeletal disorders (morbidity indicators were chosen according to the advice of the GPs who indicated their patients' main complaints).

SF6D score was estimated from the SF36 questionnaire using the established algorithm. EQ5D score was calculated using United Kingdom (UK) preference weights [28] because Italian ones were not available. Patients were divided into 10-year age groups, except for the youngest and oldest (<30, 30–39, 40–49, 50–59, 60–69, 70+). Years of education were classified according to the Italian school system. BMI was divided according to US Centers for Disease Control & Prevention guidelines (BMI  $\geq$  25 = overweight) [29].

### HUI3

HUI3 is a preference/utility-based measure that describes eight attributes, “vision”, “hearing”, “speech”, “ambulation”, “dexterity”, “emotion”, “cognition” and “pain”, and is able to identify 972,000 health states. The HUI system is based on a “within the skin” approach to health status assessment that concentrates on physical and emotional aspects, ignoring social ones. The score ranges from 0 (death) to 1 (full health), but can also take negative values that indicate states worse than death [30–34].

### EQ5D

Euroqol5D is a preference/utility-based measure. The score ranges from 0 (death) to 1 (full health), but may take negative values. It describes five dimensions: “mobility”, “self-care”, “usual activities”, “pain/discomfort” and “anxiety/depression”, and identifies 243 health states [28, 35–37].

### SF36

SF36 is a psychometric measure that produces a profile with eight dimensions: “physical functioning”, “role limitations due to physical problems” and “role limitations due to emotional problems”, “pain”, “general health”, “vitality”, “social functioning” and “mental health” [38]. The eight dimensions can take values from 0 (worst health) to 100 (best health) [39].

### SF6D

SF6D is a preference/utility-based measure created to obtain a score from the SF36. The eight dimensions of SF36 were reduced to six SF6D dimensions: “physical

functioning”, “role limitation”, “social functioning”, “pain”, “mental health” and “vitality”. The SF6D identifies 18,000 health states. The score, which ranges from 0 (death) to 1 (full health), can be calculated from the SF36 if the ten items used to identify the six dimensions of SF6D are completed [16, 40, 41].

### Data analysis

The analysis was carried out using Stata 8.1. The questionnaire score distributions were analysed. Correlation was investigated using the Spearman coefficient. Agreement between scores was investigated using the Bland and Altman [42] method.

Finally, how the questionnaire scores were influenced by socio-demographic and morbidity indicators was studied by univariate and multivariate analysis, using the non-parametric quantile (Least Absolute Value) regression. Missing values were excluded because they represented only a small percentage of the data (6% for BMI and 3% for chronic diseases) and because the groups of respondents and non-respondents were similar in socio-demographic and morbidity characteristics. Significant ( $p < 0.05$ ) differences were only found for marital status (the respondent group contained a higher proportion of married persons).

Multivariate analysis was carried out fitting a forward step-up model. Only variables found to be associated with questionnaire scores by the univariate analysis were entered in the model. It was built up starting from the socio-demographic indicators age and sex, which were considered important factors. Then education and marital status were entered followed by indicators related to lifestyle, smoking status and BMI. Finally, the role of chronic diseases was considered. Wald tests were carried out to compare the new models with the previous ones.

## Results

### Basic characteristics of population surveyed

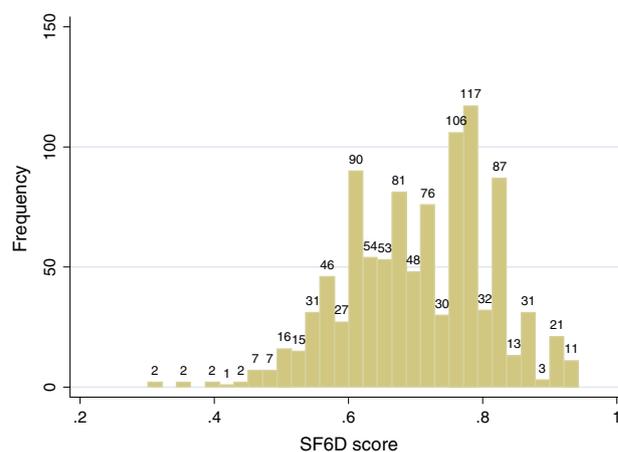
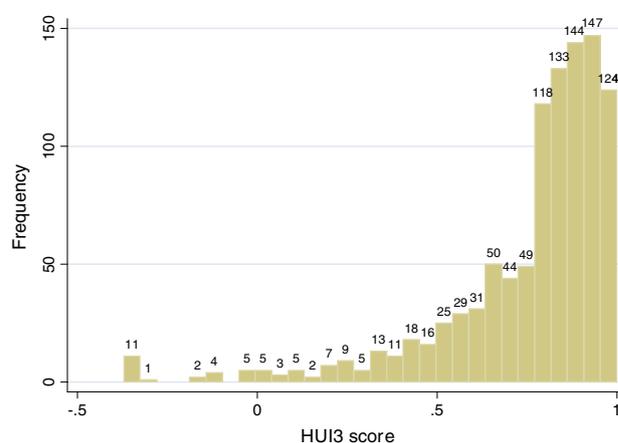
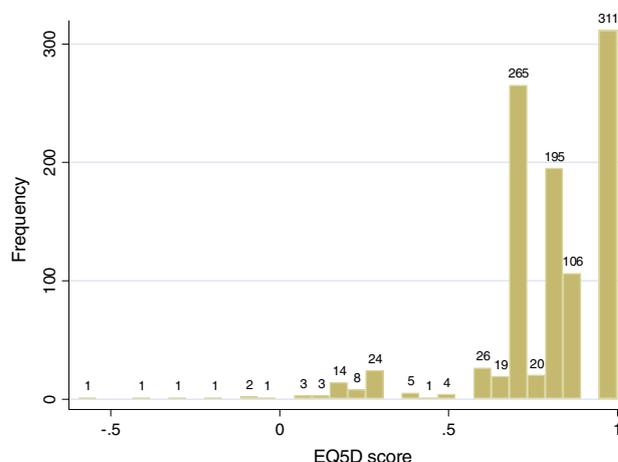
The mean age of patients was about 49 years (SD = 18). There was a higher percentage of females (59%) than males (41%). Sixty-one percent of the patients were married. About half of the sample had never smoked. Current and ex-smokers were more numerous among males (25 and 33%, respectively) than females (22 and 14%, respectively). Mean BMI was 24 (SD = 3.9). Thirty-three percent of the sample reported being overweight. Males were more often overweight (44%) than females (25%). Twenty-two percent of patients suffered from hypertension, 28% from musculoskeletal diseases, 8% from cardiovascular diseases and 5% from diabetes (Table 1).

**Table 1** Characteristics of the study population ( $N = 1011$ )

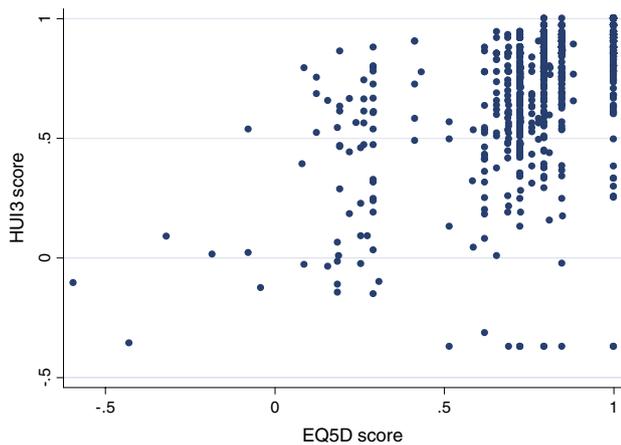
	%Males (N)	%Females (N)	%Overall (N)
<b>Age (years)</b>			
<30	16.3 (68)	15.9 (94)	16.1 (162)
30–39	20.8 (87)	20.6 (122)	20.7 (209)
40–49	15.5 (65)	14.3 (85)	14.8 (150)
50–59	15.8 (66)	18.2 (108)	17.2 (174)
60–69	16.3 (68)	14.8 (88)	15.4 (156)
70+	15.3 (64)	16.2 (96)	15.8 (160)
<b>Years of education</b>			
0–5	12.4 (52)	20.2 (120)	17.0 (172)
8	24.2 (101)	21.4 (127)	22.6 (228)
13	41.4 (173)	42.0 (249)	41.7 (422)
13+	22.0 (92)	16.4 (97)	18.7 (189)
<b>Marital status</b>			
Married	61.2 (256)	61.4 (364)	61.3 (620)
Not married	34.5 (144)	27.8 (165)	30.6 (309)
Divorced/widowed	4.3 (18)	10.8 (64)	8.1 (82)
<b>Smoking status</b>			
Never smoked	41.6 (174)	63.4 (376)	54.4 (550)
Current smoker	24.9 (104)	22.4 (133)	23.4 (237)
Ex-smoker	33.5 (140)	14.2 (84)	22.2 (224)
<b>BMI</b>			
Not-overweight	48.1 (201)	69.7 (413)	60.7 (614)
Overweight	44.0 (184)	24.9 (148)	32.9 (332)
Not specified	7.9 (33)	5.4 (32)	6.4 (65)
<b>Hypertension</b>			
Affected	21.5 (90)	21.6 (128)	21.6 (218)
Not specified	4.5 (19)	2.7 (16)	3.5 (35)
<b>Cardiovascular diseases</b>			
Affected	8.4 (35)	7.2 (43)	7.7 (78)
Not specified	4.5 (19)	2.7 (16)	3.5 (35)
<b>Diabetes</b>			
Affected	5.5 (23)	4.7 (28)	5.0 (51)
Not specified	4.5 (19)	2.7 (16)	3.5 (35)
<b>Musculo-skeletal</b>			
Affected	18.9 (79)	33.9 (201)	27.7 (280)
Not specified	4.5 (19)	2.7 (16)	3.5 (35)

### Questionnaire scores

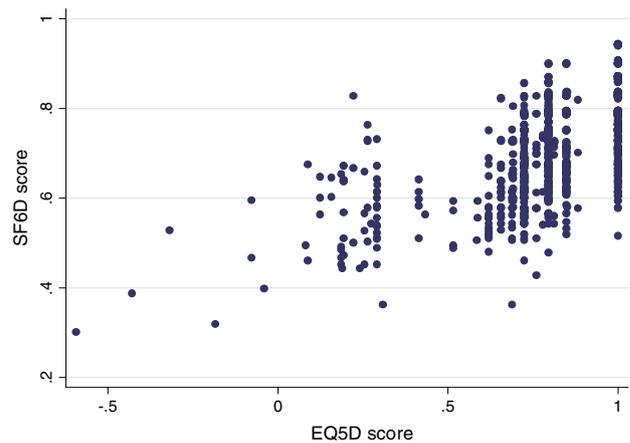
SF6D scores showed an almost normal distribution (Fig. 1), whereas HUI3 and EQ5D scores were skewed to the left (Figs. 2, 3). SF6D produced a narrower range of values: while EQ5D and HUI3 showed negative scores, the minimum SF6D score was 0.301 (SF6D mean 0.70, SD 0.11, median 0.71, 25th percentile 0.62, 75th percentile 0.79, maximum 0.94; EQ5D mean 0.80, SD 0.20, median 0.80, 25th percentile 0.72, 75th percentile 1, minimum  $-0.594$ , maximum 1; HUI3 mean 0.76, SD 0.24, median 0.85, 25th percentile 0.70, 75th percentile 0.91, minimum  $-0.371$ ,

**Fig. 1** SF6D frequency distribution**Fig. 2** HUI3 frequency distribution**Fig. 3** EQ5D frequency distribution

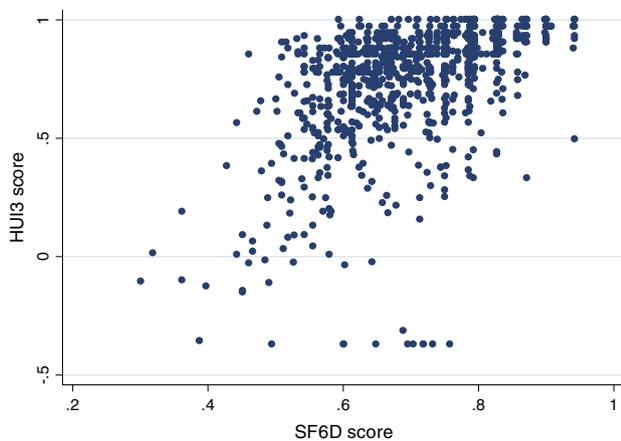
maximum 1). EQ5D showed a ceiling effect (31% of people scored the highest value, compared to 6.5% for HUI3 and 0% for SF6D). The distribution of EQ5D clustered around a few values with a big gap between 0.883 and 1.



**Fig. 4** Relationship between EQ5D and HUI3



**Fig. 6** Relationship between EQ5D and SF6D



**Fig. 5** Relationship between SF6D and HUI3

### Correlation and agreement

The relationship between the questionnaire scores is described in Figs. 4–6. SF6D and EQ5D showed the highest level of association. This was confirmed by Spearman coefficient, which was 0.59 for the association between SF6D and EQ5D, 0.58 for SF6D and HUI3 and 0.57 for EQ5D and HUI3.

The best agreement was found between SF6D and EQ5D (95% limits of agreement from  $-0.414$  to  $0.230$ ). The worst agreement was achieved by HUI3 and EQ5D (95% limits of agreement from  $-0.463$  to  $0.387$ ).

### Multivariate analysis

Seven models were constructed for each questionnaire. Table 2 shows the results of the first model, which included age and sex, and the final model, where all the other variables were entered. In the case of SF6D, after adjusting

for all the variables, females showed an average score 0.068 points lower than males (regression coefficient =  $-0.068$ ,  $p < 0.001$ ). Ex-smokers showed a slightly lower score than no-smokers (regression coefficient =  $-0.023$ ,  $p = 0.027$ ). Overweight people had an average score 0.018 points lower than people of normal weight (regression coefficient =  $-0.018$ ,  $p = 0.020$ ). Hypertension, cardiovascular diseases and musculoskeletal diseases show an effect on the scores (respectively: regression coefficient =  $-0.036$ ,  $p = 0.001$ ; regression coefficient =  $-0.035$ ,  $p = 0.029$ ; regression coefficient =  $-0.054$ ,  $p < 0.001$ ). Age, education and marital status did not influence scores.

In the case of EQ5D, females showed an average score 0.085 points lower than males (regression coefficient =  $-0.085$ ,  $p < 0.001$ ). People in the age group 60–69 showed an average score 0.058 points ( $p = 0.028$ ) lower than people in the youngest group. Married people had an average score higher than not married people (regression coefficient =  $0.032$ ,  $p = 0.056$ ). Overweight people had an average score 0.027 points lower than people of normal weight (regression coefficient =  $-0.027$ ,  $p = 0.028$ ). Hypertension and musculoskeletal diseases show an effect on the scores (respectively: regression coefficient =  $-0.027$ ,  $p = 0.065$ ; regression coefficient =  $-0.071$ ,  $p < 0.001$ ). EQ5D scores were also influenced by smoking status: current smokers had a score 0.035 points lower than no smokers (regression coefficient =  $-0.035$ ,  $p = 0.010$ ).

In the case of HUI3, there were no significant gender differences in score values (females, regression coefficient =  $-0.012$ ,  $p = 0.303$ ). The score increased with higher education levels. Hypertension, cardiovascular diseases and musculoskeletal diseases showed an effect on the scores (respectively: regression coefficient =  $-0.037$ ,  $p = 0.014$ ; regression coefficient =  $-0.055$ ,  $p = 0.013$ ; regression coefficient =  $-0.081$ ,  $p < 0.001$ ).

**Table 2** Results of the multivariate analysis investigating the influence of socio-demographic and morbidity indicators on questionnaire scores

	SF6D			HUI3			EQ5D					
	Model 1			Model 1			Model 1					
	Coeff. (SE)	<i>p</i>		Coeff. (SE)	<i>p</i>		Coeff. (SE)	<i>p</i>				
Constant	0.84 (0.019)	-	0.83 (0.022)	-	0.90 (0.015)	-	0.85 (0.031)	-	0.98 (0.020)	-	0.99 (0.031)	-
Sex												
Females	-0.072 (0.009)	<0.001	-0.068 (0.008)	<0.001	-0.025 (0.007)	0.001	-0.012 (0.012)	0.303	-0.071 (0.009)	<0.001	-0.085 (0.012)	<0.001
Age (years)												
<30												
30-39	0.003 (0.016)	0.850	-0.001 (0.015)	0.918	0.001 (0.012)	0.999	-0.004 (0.020)	0.822	-0.052 (0.017)	0.002	-0.016 (0.020)	0.414
40-49	-0.017 (0.017)	0.321	-0.012 (0.017)	0.491	0.001 (0.013)	0.999	-0.017 (0.023)	0.463	-0.052 (0.018)	0.004	-0.035 (0.023)	0.133
50-59	-0.017 (0.016)	0.301	-0.007 (0.018)	0.700	-0.025 (0.013)	0.049	-0.017 (0.024)	0.482	-0.123 (0.016)	<0.001	-0.048 (0.024)	0.051
60-69	-0.021 (0.017)	0.215	0.013 (0.019)	0.515	-0.061 (0.013)	<0.001	-0.004 (0.026)	0.866	-0.123 (0.017)	<0.001	-0.058 (0.026)	0.028
70+	-0.074 (0.017)	<0.001	-0.011 (0.020)	0.554	-0.128 (0.013)	<0.001	-0.028 (0.027)	0.313	-0.123 (0.017)	<0.001	-0.052 (0.027)	0.060
Years of education												
0-5												
8			0.010 (0.014)	0.452			0.025 (0.019)	0.180			-0.026 (0.019)	0.151
13			0.016 (0.013)	0.208			0.038 (0.018)	0.037			0.004 (0.018)	0.839
>13			0.027 (0.015)	0.080			0.038 (0.020)	0.069			0.007 (0.020)	0.732
Marital status												
Not married												
Married			0.018 (0.012)	0.138			0.025 (0.016)	0.129			0.032 (0.016)	0.056
Divorced/widow			0.007 (0.018)	0.687			-0.032 (0.022)	0.155			0.017 (0.025)	0.484
Smoking status												
Never smoked												
Current smoker			-0.010 (0.010)	0.293			-0.012 (0.013)	0.360			-0.035 (0.013)	0.010
Ex-smoker			-0.023 (0.010)	0.027			0.001 (0.014)	0.999			-0.026 (0.014)	0.074
BMI												
Overweight			-0.018 (0.009)	0.020			-0.012 (0.012)	0.308			-0.027 (0.012)	0.028
Hypertension												
Affected			-0.036 (0.011)	0.001			-0.037 (0.015)	0.014			-0.028 (0.015)	0.065
Cardiovascular diseases												
Affected			-0.035 (0.016)	0.029			-0.055 (0.022)	0.013			-0.006 (0.022)	0.777
Diabetes												
Affected			-0.005 (0.019)	0.771			-0.007 (0.026)	0.777			-0.011 (0.026)	0.652
Musculo- skeletal diseases												
Affected			-0.054 (0.010)	<0.001			-0.081 (0.014)	<0.001			-0.071 (0.014)	<0.001

## Discussion and conclusions

This study compared three widely used HRQL measures and tried to clarify if they measure HRQL in a similar way. The three measures showed different score ranges and distributions, but EQ5D and HUI3 were more similar to each other in distribution, mean, median, maximum and minimum score than to SF6D. SF6D scores appeared to have a more normal distribution and covered a narrower range of values. EQ5D showed a ceiling effect with 31% of people scoring the highest value. As highlighted by studies with similar results, EQ5D seems to fail to describe mild-severity health levels [16, 27].

In fact, with only three levels for each dimension, EQ5D does not grade between fair and good health. However, SF6D produced higher values for health conditions at the lower end of the scale. The lowest SF6D score obtained in this study was 0.301, while EQ5D and HUI3 produced negative values. Other studies with similar results [11, 13, 21] interpreted this performance as a poor ability of SF6D to distinguish severely impaired status. Overall, these findings seem to confirm those of other studies: SF6D produced higher values at the lower end of the scale and EQ5D at the upper end.

Differences between the questionnaires were also outlined by agreement between scores in the low range. The 95% limits of agreement were quite large, ranging approximately from  $-0.5$  to  $0.3$ , which, on a score scale from 0 to 1, is an important discrepancy. The best agreement was achieved by SF6D and EQ5D and the worst by EQ5D and HUI3. These results were probably due both to construct and statistical issues. Regarding the construct issue, HUI3 describes eight dimensions of health, six of which are related to particular aspects (such as vision or hearing), all focussed on the physical area of health. Only two scales are related to emotional and mental health, and none to social aspects of health. In fact, HUI3 is based on a “within the skin” approach to health status assessment that concentrates on physical and emotional areas and eliminates the social one because it is “outside the skin” [31–34]. On the contrary, SF6D and EQ5D not only describe physical and emotional dimensions of health, but also the social one. These different constructs could explain the worse agreement between HUI3 and the other measures. Regarding the statistical issue, the particularly poor agreement between HUI3 and EQ5D could be due to the skewed distributions of their scores. They both showed a ceiling effect, but, in the case of EQ5D, there is also major clustering of scores around a few values, which decreases the heterogeneity of the sample and hence the level of agreement whose statistics rely upon variance. Moreover, the extreme score values can lead to occurrence of outliers in the differences distribution and therefore widen the limits of agreement.

However, the scores showed a good correlation. A Spearman coefficient of 0.6 is described as very high [43]. In our study, the Spearman coefficient was around 0.6 for all questionnaires, indicating a good level of correlation, especially between SF6D and EQ5D. This level of correlation was similar, but to some extent inferior to those found in other studies [18, 20].

These findings highlighted that the results of HRQL measures may be influenced by their frameworks and the different methods used to calculate the scores. For example, SF6D could overestimate the health status of persons with severe illness and could be more suitable for surveys on the general population or people with a fair to good health status. However, EQ5D seems to overestimate middle-severity health status and could therefore be less suitable for describing the health status of the general population and more useful for patients with invalidating disease. Moreover, the three questionnaires are not interchangeable, and their results cannot be compared because their results show poor agreement, especially between HUI3 and EQ5D. This aspect could be a major issue in comparisons among populations because health status measured with different questionnaires is unlikely to be comparable.

Considering the above, it comes as a surprise to discover that the three measures had similar performance, especially SF6D and EQ5D, in relation to socio-demographic and clinical variables. In fact, this study highlighted that SF6D and EQ5D scores are influenced by the same pattern of factors. Multivariate analysis showed scores of the two questionnaires were influenced, in particular, by gender, with females showing poorer health than males, and by chronic diseases, especially musculoskeletal. HRQL seems to be influenced by the impact that diseases have on daily life rather than by the severity or possible complications of a disease. In fact, musculoskeletal diseases, usually painful and debilitating, influence HRQL more than hypertension or cardiovascular diseases. The two questionnaires show a gender difference in health, though both are also influenced by factors related to lifestyle, such as smoking and BMI.

HUI3, on the contrary, had slightly different performance. It seemed to be influenced by educational level and especially hypertension, cardiovascular and musculoskeletal diseases. However, HUI3 did not reveal health differences between males and females and did not seem to be influenced by factors related to lifestyle. This different performance could be related, as mentioned above, to the approach of the questionnaire, which focusses on physical and emotional aspects and excludes social ones.

These results give rise to some considerations. First, the scores of all questionnaires were influenced by musculoskeletal diseases, which are conditions characterised by physical pain, difficulty of movements, immobility and,

often, partly because of the associated pain, deterioration in daily activities, and which, therefore, could have an effect also on vitality and social life. Therefore, HRQL seems to be influenced more by painful, and consequently daily-activities-limiting, conditions than by diseases that may be more serious, but are often asymptomatic. This could be a limit as well as a way to highlight different aspects of health. In fact, HRQL measures may underestimate health status for painful, but not fatal diseases, while overestimating health status in the case of serious, but asymptomatic diseases. HRQL measures may therefore help detect health needs that would otherwise remain concealed, but should probably be used to integrate other health measures, such as mortality, which are more objective, but more crude, or they could be “adjusted” for morbidity conditions assessed by more objective methods (such as morbidity indexes, which describe the severity of a disease).

Secondly, none of the three questionnaires, with the exception of EQ5D, seemed to be influenced by age, after adjusting for the other variables. This suggests that most of the decrease in HRQL in old age is due to factors other than age itself, such as diseases or other conditions like loneliness, which are more frequent in the elderly.

Thirdly, SF6D and EQ5D show a similar capacity of discrimination, while HUI3 seems to be less able to distinguish different categories of people. In particular, EQ5D seems to be the only one to detect some health differences between age groups and among smokers and non-smokers, whilst SF6D identifies differences between non-smokers and ex-smokers. These results should be considered when choosing a measure. Although the questionnaires have diverse frameworks and their crude scores may be different and difficult to compare, they appeared to be influenced by socio-demographic and morbidity variables in a similar way, especially EQ5D and SF6D. This shifts emphasis from the structural and construct similarities of different instruments to the behaviour that they reveal when applied in the field. The study endeavoured to examine the performance of the three measures when used to describe patients’ condition and their determinants instead of merely comparing ranges or distributions of scores. The results obtained could help in the choice of instrument, also considering that this study did not focus on a group of patients with a specific disease in order that the results are more generalisable.

The present study shows some limitations: (1) all the information about morbidity indicators is self-reported by patients so they could be misclassified; (2) participation was voluntary so there could be selection bias; (3) refusals were not recorded so it was impossible to assess whether people who refused to answer the questionnaires differed from people who agreed. However, for the aim of the

study, these possible sources of error should not be of great concern, because the biases would involve all three instruments in the same way, and comparison would not be altered. However, these possible sources of error could affect the ability of the study to generalise the findings. Another problem could be the order in which the questionnaires were administered. Since they were always administered in the same order, the last one could have suffered from loss of accuracy. However, EQ5D, the last one allocated to patients, is the shortest and easiest, so its implemation was as good as for the other two.

In conclusion, our results show that EQ5D and HUI3 were closer to one another in many ways (score distribution, mean, median, minimum and maximum), but SF6D and EQ5D scores were more similar in the way they were influenced by socio-demographic and morbidity indicators. In some cases, such as for smoking and age, EQ5D had better discrimination capacity. It is difficult to determine which is the best instrument, but, apart from a descriptive capacity similar or better than the other instruments, EQ5D seems to have the advantage of being easier to answer.

**Acknowledgements** The authors acknowledge the following colleagues for help with data collection: A. Bardi, M. Borghi, S. Ciampolini, G. D’Eredità, P. Galloni, P. Liberatori, M. Monciatti, S. Orecchia, M. Picchi, L. Ranzani, R. Ravazzani, G. Righi, C. Rossi, R. Rutto, C. Satta, M.T. Torino and C. Vita. A special acknowledgement goes to Prof. Franco Sassi, Department of Social Policy, London School of Economics and Political Science (UK), for valuable advice during the preparation and revision of the paper.

**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.

## References

1. Guyatt, G.H., Feeny, D.H., Patrick, D.L.: Measuring health-related quality of life. *Ann. Intern. Med.* **118**(8), 622–629 (1993)
2. Patrick, D.L., Bergner, M.: Measurement of health status in the 1990s. *Annu. Rev. Public Health* **11**, 165–183 (1990). doi:[10.1146/annurev.pu.11.050190.001121](https://doi.org/10.1146/annurev.pu.11.050190.001121)
3. Nelson, E.C., Berwick, D.M.: The measurement of health status in clinical practice. *Med. Care* **27**(Suppl. 3), 77–89 (1989). doi:[10.1097/00005650-198903001-00007](https://doi.org/10.1097/00005650-198903001-00007)
4. Higginson, I.J., Carr, A.J.: Using quality of life measures in the clinical setting. *BMJ* **322**, 1297–1300 (2005). doi:[10.1136/bmj.322.7297.1297](https://doi.org/10.1136/bmj.322.7297.1297)
5. Greenfield, S.: The state of outcome research: are we on target? *N. Engl. J. Med.* **320**(17), 1142–1143 (1989)
6. Goldfield, N.: The hubris of health status measurement: a clarification of its role in the assessment of medical care. *Int. J. Qual. Health Care* **8**(2), 115–123 (1996). doi:[10.1093/intqhc/8.2.115](https://doi.org/10.1093/intqhc/8.2.115)
7. Carr-Hill, R.A.: Assumptions of the QALY procedure. *Soc. Sci. Med.* **29**(3), 468–477 (1989). doi:[10.1016/0277-9536\(89\)90296-7](https://doi.org/10.1016/0277-9536(89)90296-7)
8. Johnson, J.A., Conner-Spady, B., Pickard, A.S.: Health-related Quality of Life as a predictor of future physician visits and

- hospitalization. Proceedings of the 17th Plenary Meeting of the Euroqol Group; 2000 Sep 28–29; Pamplona, Spain; pp. 184–191
9. Kaplan, R.M.: The significance of quality of life in health care. *Qual. Life Res.* **12**(Suppl. 1), 3–16 (2003). doi:[10.1023/A:1023547632545](https://doi.org/10.1023/A:1023547632545)
  10. Carr-Hill, R., Morris, J.: Current practice in obtaining the “Q” in QALYs; a cautionary note. *BMJ* **303**, 699–701 (1992)
  11. Kopec, J.A., Willison, K.D.: A comparative review of four preference-weighted measures of health-related quality of life. *J. Clin. Epidemiol.* **56**, 317–325 (2003). doi:[10.1016/S0895-4356\(02\)00609-1](https://doi.org/10.1016/S0895-4356(02)00609-1)
  12. Brazier, J., Deverill, M., Green, C., Harper, R., Booth, A.: A review of the use of health status measures in economic evaluation. *Health Technol. Assess.* **3**(9), i–iv, 1–164 (1999)
  13. Longworth, L., Bryan, S.: An empirical comparison of the EQ-5D and SF-6D in liver transplant patients. *Health Econ.* **12**, 1061–1067 (2003). doi:[10.1002/hec.787](https://doi.org/10.1002/hec.787)
  14. Hawthorne, G., Richardson, J., Day, N.A.: A comparison of the Assessment of Quality of Life (AQoL) with four other generic utility instruments. *Ann. Med.* **33**(5), 358–370 (2001). doi:[10.3109/07853890109002090](https://doi.org/10.3109/07853890109002090)
  15. Johnson, J.A., Pickard, A.S.: Comparison of the EQ-5D and the SF-12 health surveys in a general population survey in Alberta, Canada. *Med. Care* **38**(1), 115–121 (2000). doi:[10.1097/00005650-200001000-00013](https://doi.org/10.1097/00005650-200001000-00013)
  16. Conner-Spady, B., Suarez-Almazor, M.E.: Variation in the Estimation of Quality-adjusted life-years by different preference-based Instruments. *Med. Care* **41**(7), 791–801 (2003). doi:[10.1097/00005650-200307000-00003](https://doi.org/10.1097/00005650-200307000-00003)
  17. O’Brien, B.J., Spath, M., Blackhouse, G., Severens, J.L., Dorian, P., Brazier, J.: A view from the bridge: agreement between the SF-6D utility algorithm and the Health Utilities Index. *Health Econ.* **12**, 975–981 (2003). doi:[10.1002/hec.789](https://doi.org/10.1002/hec.789)
  18. Brazier, J., 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). doi:[10.1002/hec.866](https://doi.org/10.1002/hec.866)
  19. 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**(11), 1125–1131 (2004). doi:[10.1097/00005650-200411000-00012](https://doi.org/10.1097/00005650-200411000-00012)
  20. McDonough, C.M., Grove, M.R., Tosteson, T.D., Lurie, J.D., Hilibrand, A.S., Tosteson, A.N.A.: Hilibrand As, Tosteson ANA, Comparison of EQ-5D, HUI, and SF-36-derived societal health state values among Spine Patient Outcomes Research Trial (SPORT) participants. *Qual. Life Res.* **14**, 1321–1332 (2005). doi:[10.1007/s11136-004-5743-2](https://doi.org/10.1007/s11136-004-5743-2)
  21. 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). doi:[10.1007/s11136-004-3928-3](https://doi.org/10.1007/s11136-004-3928-3)
  22. Kaplan, R.M., Groessl, E.J., Sengupta, N., Sieber, W.J., Ganiats, T.G.: Comparison of measured utility scores and imputed scores from the SF-36 in patients with rheumatoid arthritis. *Med. Care* **43**(1), 79–87 (2005)
  23. Stirling, B., Longworth, L.: Measuring health-related utility: why the disparity between EQ-5D and SF-6D? *Eur. J. Health Econ.* **50**, 253–260 (2005)
  24. Gerard, K., Nicholson, T., Mullee, M., Metha, R., Roderick, P.: EQ-5D versus SF-6D in an older, Chronically Ill Patient Group. *Appl. Health Econ. Health Policy* **3**(2), 91–102 (2004)
  25. Belanger, A., Berthelot, J.M., Guimond, E., Houle, C.: A head-to-head comparison of two generic health status measures in the household population: McMaster Health Utility Index (Mark 3) and the EQ-5D. Ottawa: statistics Canada, Health Analysis and Modelling Group (2000)
  26. Bosch, J.L., Hunink, M.G.M.: Comparison of the Health Utilities Index Mark 3 (HUI3) and the Euroqol Eq-5D in patients for intermittent claudication. *Qual. Life Res.* **9**, 591–601 (2000). doi:[10.1023/A:1008929129537](https://doi.org/10.1023/A:1008929129537)
  27. Macran, S., Weatherly, H., Kind, P.: Measuring population health. A comparison of three generic health status measures. *Med. Care* **41**(2), 218–231 (2003). doi:[10.1097/00005650-200302000-00004](https://doi.org/10.1097/00005650-200302000-00004)
  28. Dolan, P.: Modeling valuations for Euroqol health states. *Med. Care* **35**(11), 1095–1108 (1997). doi:[10.1097/00005650-199711000-00002](https://doi.org/10.1097/00005650-199711000-00002)
  29. BMI—Body Mass Index: BMI for adults, centers for disease control and prevention. Retrieved August 2005 from: <http://www.cdc.gov/nccdphp/dnpa/bmi/bmi-adult.htm>
  30. Torrance, G.W., Boyle, M.H., Horwood, S.P.: Application of multi-attribute utility theory to measure social preferences for health states. *Oper. Res.* **30**(6), 1043–1069 (1982). doi:[10.1287/opre.30.6.1043](https://doi.org/10.1287/opre.30.6.1043)
  31. Horsman, J., Furlong, W., Feeny, D., Torrance, G.: The Health Utility Index (HUI): concepts, measurement properties and applications. *Health Qual. Life Outcomes* **1**(1), 54 (2003). doi:[10.1186/1477-7525-1-54](https://doi.org/10.1186/1477-7525-1-54)
  32. Torrance, G.W., Furlong, W., Feeny, D., Boyle, M.: Multi-attribute preference functions—health utility index. *Pharmacoeconomics* **7**(6), 503–520 (1995). doi:[10.2165/00019053-199507060-00005](https://doi.org/10.2165/00019053-199507060-00005)
  33. Feeny, D., Furlong, W., Boyle, M., Torrance, G.W.: Multi-attribute health status classification systems—health utility index. *Pharmacoeconomics* **7**(6), 490–502 (1995). doi:[10.2165/00019053-199507060-00004](https://doi.org/10.2165/00019053-199507060-00004)
  34. Feeny, D., Furlong, W., Torrance, G.W., Goldsmith, C., Zhu, Z., DePauw, S., Denton, M., Boyle, M.: Multiattribute and single-attribute utility functions for the health utilities index Mark 3 system. *Med. Care* **40**(2), 113–128 (2002). doi:[10.1097/00005650-200202000-00006](https://doi.org/10.1097/00005650-200202000-00006)
  35. The Euroqol Group: Euroqol—a new facility for the measurement of health-related quality of life. *Health Policy* **16**(3), 199–208 (1990). doi:[10.1016/0168-8510\(90\)90421-9](https://doi.org/10.1016/0168-8510(90)90421-9)
  36. Brooks, R.: The Euroqol Group, Euroqol: the current state of play. *Health Policy* **37**(1), 53–72 (1996). doi:[10.1016/0168-8510\(96\)00822-6](https://doi.org/10.1016/0168-8510(96)00822-6)
  37. Greiner, W., Weijnen, T., Nieuwenhuizen, M., Oppe, S., Badia, X., Busschbach, J., Buxton, M., Dolan, P., Kind, P., Krabbe, P., Ohinmaa, A., Parkin, D., Roset, M., Sintonen, H., Tsuchiya, A., de Charro, F.: A single European currency for EQ-5D health states. Results from a six-country study. *Eur. J. Health Econ.* **4**(3), 222–231 (2003). doi:[10.1007/s10198-003-0182-5](https://doi.org/10.1007/s10198-003-0182-5)
  38. Ware, J.E., Sherbourne, C.D.: The M.O.S.: 36-item Short Form Health Survey (SF-36). *Med. Care* **30**, 473–483 (1992). doi:[10.1097/00005650-199206000-00002](https://doi.org/10.1097/00005650-199206000-00002)
  39. Brazier, J.: The Short-Form 36 (SF-36) Health Survey and its use in pharmacoeconomic evaluation. *Pharmacoeconomics* **7**(5), 403–415 (1995)
  40. Brazier, J., Usherwood, T., Harper, R., Thomas, K.: Deriving a preference-based single index from the UK SF-36 health survey. *J. Clin. Epidemiol.* **51**(11), 1115–1128 (1998). doi:[10.1016/S0895-4356\(98\)00103-6](https://doi.org/10.1016/S0895-4356(98)00103-6)
  41. Brazier, J., Roberts, J., Deverill, M.: The estimation of a preference-based measure of health from the SF-36. *J. Health Econ.* **21**(2), 271–292 (2002). doi:[10.1016/S0167-6296\(01\)00130-8](https://doi.org/10.1016/S0167-6296(01)00130-8)
  42. Bland, J.M., Altman, D.G.: Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* **i**, 307–310 (1986)
  43. McDowell, I., Newell, C.: Measuring health: a guide to rating scales and questionnaires. Oxford University Press, New York (1996)