How consistent are health utility values?

**Pedro L. Ferreira · Lara N. Ferreira ·
Luis N. Pereira**

Abstract The use of preference-based generic instruments to measure the health-related quality of life of a general population or of individuals suffering from a specific disease has been increasing. However, there are several discrepancies between instruments in terms of utility results. This study compares SF-6D and EQ-5D when administered to patients with cataracts and aims at explaining the differences. Agreement between EQ-5D and SF-6D health state classifications was assessed by correlation coefficients. Simple correspondence analysis was used to assess the agreement among the instrument's descriptive systems and to investigate similarities between dimensions' levels. Cluster analysis was used to classify SF-6D and EQ-5D levels into homogeneous groups. There was evidence of floor effects in SF-6D and ceiling effects in EQ-5D. Comparisons of means showed that SF-6D values exceeded EQ-5D values. Agreement between both instruments was high, especially between similar dimensions. However, different valuation methods and scoring algorithms contributed to the main differences found. We suggest that one or both instruments should be revised, in terms of their descriptive systems or their scoring algorithms, in order to overcome the weakness found.

P. L. Ferreira (✉)
Faculty of Economics, University of Coimbra, Coimbra,
Portugal
e-mail: pedrof@fe.uc.pt

P. L. Ferreira · L. N. Ferreira · L. N. Pereira
Centre for Health Studies & Research, University of Coimbra,
Coimbra, Portugal

L. N. Ferreira
School of Management, Hospitality and Tourism, University
of the Algarve, Faro, Portugal

Introduction

The use of preference-based generic instruments to measure the health-related quality of life (HRQL) of a general population or of individuals suffering from a specific disease has been increasing. These instruments, based on the multi-attribute utility theory, generate utilities and are essentially generic HRQL instruments with predefined preference weights. The preference weights or utility scores for the different health states are derived through a valuation process, using techniques as the standard gamble (SG), the time trade-off (TTO) or the visual analogue scale (VAS). Some of the most used multi-attribute utility measures are the EuroQoL (EQ-5D) [1–3], the Health Utilities Index (HUI) [4–6], the Short-Form 6D (SF-6D) [7, 8], the Quality of Well-Being Scale (QWB) [9–11], and the Assessment of Quality of Life (AQoL) [12, 13]. Their ease of administration has contributed to their increased use as a source of quality weightings in economic evaluations and in clinical trials.

However, there are several discrepancies in terms of utility results between instruments. In fact, many researchers found significant differences in global utility scores obtained by different multi-attribute utility instruments [14–24]. The objective of the present study is to compare the SF-6D and the EQ-5D and to investigate the differences in agreement between them. The main goal is to understand the possible reasons for the divergences found and to analyze their implications.

Methods

Study sample

Patients with cataracts waiting for a surgery at two hospitals in the Algarve, Portugal, from May to August 2005 were identified for this study. Patients were approached during an outpatient visit and asked to participate. Informed consent was obtained from all study participants, who answered self-administered questionnaires; those with difficulties in seeing their contents or who were illiterate were helped by a nurse. The order of the questionnaires was predefined and was the same throughout the study: SF-6D, EQ-5D, and Catquest. We used the official Portuguese versions of those instruments. One month after surgery, when patients came for a fourth follow-up visit, they were asked to complete the same questionnaires and were again helped by a nurse, if necessary. In this paper we only present the results from SF-6D and EQ-5D baseline assessment.

SF-6D

SF-6D is a new single-index summary preference-based measure of health derived from 11 items of SF-36 by a team at the University of Sheffield [8]. The items of SF-36 are converted into a six-dimensional health state classification system, the SF-6D, with four to six levels, allowing for a total of 18,000 unique health states. Dimensions of SF-6D include physical functioning, role limitations, social functioning, pain, mental health, and vitality. Different health states are assigned to values derived from valuations of a sample of 249 SF-6D health states using SG in a representative sample of the United Kingdom (UK) population [8]. The SF-6D score can be regarded as a continuous value on a 0.30–1.00 scale, where 1.00 indicates full health [8].

EQ-5D

EQ-5D was developed by the EuroQoL Group, a multidisciplinary group of researchers, as a standard generic instrument for describing and valuing quality of life that could be used to generate cross-national comparisons of health state [25]. It is composed of two parts. The first is a descriptive system consisting of five dimensions (mobility, self care, usual activities, pain/discomfort, and anxiety/depression), with three levels each, allowing for a total of 243 health states. States have been valued by a representative sample of the UK general population using the TTO valuation technique [2, 3]. Models were estimated to predict single-index scores for all health states, named the EQ-5D index. This index allows values below zero corresponding to conditions worse than dead. The second is a

VAS, looking like a thermometer, with values corresponding to each respondent's current perception regarding his/her personal HRQL. Respondents are asked to rate their current health on a scale from 0 to 100, semantically anchored by worst and best imaginable health states [1]. It is a self-administered questionnaire, easy to apply, and its brevity has been considered a plus.

Statistical analysis

Only subjects who fully completed the SF-6D, the EQ-5D, and the VAS were considered; no replacement or imputation was performed on missing response items. Frequencies and descriptive statistics were computed to characterize the study sample. Comparisons between utility measures were possible through descriptive statistics, as well as Spearman correlation coefficients. Both parametric [t-tests and analysis of variance (ANOVA)] and nonparametric tests (Kruskal–Wallis tests) were used to look for significant differences in utilities among sociodemographic groups. These differences were considered statistically significant if P-values were less than 0.10. Nonparametric tests were used because of the heterogeneity of variances observed in some cases and the nonnormality of some dimensions. Simple correspondence analysis (SCA) was used to assess the agreement among the instruments' descriptive systems and to look for similarities between dimensions' levels. Cluster analysis was used to classify SF-6D and EQ-5D levels into homogeneous groups. The statistical software used for the analyses were SPSS version 13, and SAS version 8.0.

Agreement among utility measures

Correspondence analysis is a descriptive and exploratory technique designed to analyze simple two-way and multiway tables containing some measure of correspondence between rows and columns. The results provide information, similar in nature to that produced by factor analysis techniques, about the structure of categorical variables included in the table. This technique is used for displaying the associations among a set of categorical variables in a scatterplot or map, allowing a visual examination of any pattern or structure in the data. Correspondence analysis is a technique for displaying multivariate categorical data graphically, by deriving coordinates to represent categories of the variables involved, which may then be plotted to provide an illustration of the data [26]. Displaying the categories of a contingency table in a scatterplot encompasses the concept of distance between the percentage profiles of each variable. When analyzing the scatterplot one should be aware that directly associated variables will have close coordinates and, therefore, will be plotted near to each other.

We used SCA with the purpose of assessing the agreement among the instruments' descriptive systems and of investigating similarities between dimensions' levels.

Clustering SF-6D and EQ-5D levels

Aiming at classifying SF-6D and EQ-5D levels into homogeneous groups, we also carried out a hierarchical agglomerative cluster analysis and a partitional k-means clustering. In the first case, we applied the hierarchical cluster analysis to the six dimensions identified through the SCA, using the Ward, furthest neighbor, and within-groups agglomeration methods and the squared Euclidean distance as a distance measure. In the k-means cluster analysis, used to obtain a better classification, each point was assigned to the centroid by using the furthest neighbor method. The use of all these methods enabled us to validate the cluster analysis. The decision regarding the number of clusters to choose was based on the fusion coefficients, the cut-off of the dendrogram, the elbow criterion, and R^2 measures.

Results

Sample

From the 360 participants who received the baseline questionnaire, 300 completed the VAS as part of the EQ-5D, and 352 global utility scores could be generated using SF-6D and EQ-5D scoring functions.

The majority of the sample were women (56.5%), married or living together with someone else (60.5%), with low educational level (79.5%). The respondents' age ranged from 49 to 92 years, with a mean of 73 years (standard deviation, SD = 8.7 years). They were most frequently retired or manual workers, and living in urban areas (Table 1). Although the sample average income was low, i.e., less than €500 (71.0%), almost all respondents lived in their own houses (86.4%).

Comparison between utility measures

The distributions for SF-6D and EQ-5D are shown in Table 2. In SF-6D, attributes with 15% or more patients at the two lowest levels include role limitation, social functioning, pain, and mental health. There may be some potential for a floor effect in this measure, particularly for role limitation, because many patients feel that they are situated in this dimension's lowest level when compared with their own responses to usual activities on the EQ-5D. On the contrary, there is evidence of a ceiling effect in EQ-5D, that is, there is very little use of level 3 in three of its

Table 1 Demographic characteristics of patients

	n	%
Total sample	352	100.0
Sex		
Female	199	56.5
Male	153	43.5
Age group		
Less than 61 years	44	12.5
61–70 years	68	19.3
71–80 years	184	52.3
More than 80 years	56	15.9
Marital status		
Married/living together	213	60.5
Single	18	5.5
Divorced/separated	13	3.7
Widowed	108	30.7
Educational level		
Illiterate	80	22.7
Low	200	56.8
Middle	68	19.3
High	4	1.1
Employment status		
Unskilled manual workers	54	15.4
Retired	277	79.1
Housewives	14	4.0
Unemployed	5	1.4
Residence		
Urban area	248	70.7
Rural area	103	29.3
Income by month		
Less than €500	49	71.0
€1,000–1,499	12	17.4
€1,500–1,999	4	5.8
€2,000 or more	4	5.8

five dimensions. This suggests that one extreme problem in EQ-5D is much worse than any of the worst levels of the SF-6D [14].

Table 3 represents the Spearman correlation coefficients between the SF-6D and EQ-5D. As expected, all the similar dimensions had direct and high correlation (greater than 0.45): physical functioning and mobility, physical functioning and usual activities, role limitations and usual activities, social functioning and mobility, social functioning and usual activities, pain and pain/discomfort, and mental health and anxiety/depression. There were also high correlations between mental health and pain/discomfort, between vitality and mobility, between vitality and pain/discomfort, and between vitality and anxiety/depression. On the contrary, we found that role limitations and self-care, physical

Table 2 Distributions of responses to SF-6D and EQ-5D dimensions (percentages)

Level	SF-6D dimensions					
	Physical functioning	Role limitations	Social functioning	Pain	Mental health	Vitality
1	15.9	27.3	27.3	25	20.5	29.5
2	**35.2**	**46.6**	20.5	9.1	**46.6**	**40.9**
3	28.4	10.2	**33.0**	10.2	17.0	18.2
4	11.4	15.9	15.9	**36.4**	14.8	6.8
5	9.1	–	3.4	13.6	1.1	4.5
6	0.0	–	–	5.7	–	–

Level	EQ-5D dimensions				
	Mobility	Self-care	Usual activities	Pain/discomfort	Anxiety/depression
1	**62.5**	**89.8**	**62.5**	30.7	38.6
2	9.1	9.1	36.4	**61.4**	**54.5**
3	1.1	1.1	1.1	8.0	6.8

Modal level is in bold. The distribution adds up to 100% by columns

Table 3 Spearman correlation coefficients between SF-6D and EQ-5D dimensions*

SF-6D dimensions	EQ-5D dimensions				
	Mobility	Self-care	Usual activities	Pain/discomfort	Anxiety/depression
Physical functioning	**0.50**	0.28	**0.50**	0.40	0.41
Role limitations	0.41	0.24	**0.45**	0.30	0.41
Social functioning	**0.60**	0.38	**0.62**	0.37	0.44
Pain	0.42	0.30	0.42	**0.77**	0.43
Mental health	0.43	0.43	0.32	**0.52**	**0.70**
Vitality	**0.47**	0.29	0.40	**0.52**	**0.59**

* All the correlation coefficients are significant at the 1% level. High correlations (>0.45) are indicated in bold

functioning and self-care, and vitality and self-care had the lowest correlations (<0.30).

We also found that VAS values were lower than both EQ-5D and SF-6D scores. Utility scores for the three measures are reported in Table 4 and Fig. 1. In terms of the mean, SF-6D and EQ-5D provide very similar estimates; the VAS mean is systematically lower. Twenty-five percent of the sample registered at least 0.85 health state utility in EQ-5D, while the same percentage reported at least 0.77 in SF-6D and 0.69 in VAS.

**Table 4** Descriptive statistics of SF-6D[a], EQ-5D[b], and VAS utility scores

Measure	Mean (SD)	Min.	Max.
SF-6D	0.69 (0.15)	0.38	1.00
EQ-5D	0.69 (0.27)	−0.18	1.00
VAS	0.58 (0.17)	0.20	1.00

[a] SF-6D scores based on the Brazier tariff [8]

[b] EQ-5D scores based on the York tariff [2]

The range and variability for EQ-5D and VAS were higher than those of SF-6D scores. EQ-5D presented negative values, denoting health states worse than death.

We also tested the sensitivity of health utility measures and VAS in terms of major patient characteristics. All measure scores were statistcally significantly lower in women (Table 5). As was expected, patients aged less than 61 years reported slightly higher levels of utility scores (except in VAS), and these differences were significant in all age groups. Contrarily to SF-6D, the EQ-5D measure seems to capture the general idea that older people tend to increase slightly the values they give to their health compared with individuals from the immediately lower age group. Mean utility scores were statistically significantly lower in the lower educational level than in the other educational levels. Statistically significant differences were found among people living in urban and rural areas; for the SF-6D people living in rural areas reported lower levels of utility scores. People married or living together with someone else reported higher levels of health utilities than

Fig. 1 Distribution of SF-6D, EQ-5D, and VAS utility scores
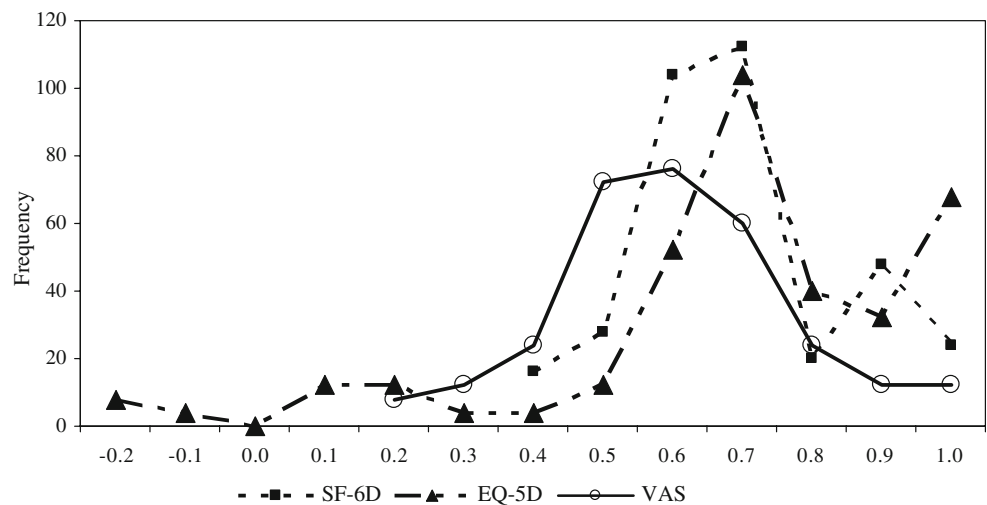
Table 5 Relationship between patients' characteristics and utility measures

	Utility score, mean (SD)		
	EQ-5D	SF-6D	VAS
Sex[a]			
Female	0.60 (0.30)	0.65 (0.13)	0.57 (0.14)
Male	0.80 (0.16)*	0.75 (0.16)*	0.61 (0.19)***
Age group[b,c]			
Less than 61 years	0.78 (0.24)	0.78 (0.14)	0.51 (0.15)
61–70 years	0.74 (0.23)	0.71 (0.15)	0.65 (0.18)
71–80 years	0.64 (0.30)	0.67 (0.16)	0.57 (0.17)
More than 80 years	0.70 (0.22)**	0.66 (0.13)*	0.61 (0.11)*
Educational level[c]			
Illiterate	0.60 (0.33)	0.63 (0.13)	0.56 (0.14)
Low	0.71 (0.22)	0.69 (0.15)	0.58 (0.16)
Middle	0.72 (0.31)	0.75 (0.14)	0.62 (0.21)
High	1.00 (0.00)*	1.00 (0.00)*	0.85 (0.00)*
Marital status[c]			
Single	0.57 (0.43)	0.72 (0.18)	0.64 (0.32)
Married or living together	0.73 (0.25)	0.70 (0.16)	0.57(0.15)
Divorced	0.44 (0.39)	0.61 (0.09)	0.45 (0.21)
Widowed	0.65 (0.43)*	0.67 (0.14)	0.62 (0.17)**
Employment status[b,c]			
Retired	0.67 (0.28)	0.67 (0.15)	0.58 (0.16)
Employed	0.79 (0.23)	0.78 (0.16)	0.61 (0.17)
Housewives	0.59 (0.32)*	0.65 (0.07)*	0.60 (0.17)*
Unemployed	1.00 (0.00)	0.93 (0.05)	0.56 (0.13)
Residence[a]			
Urban area	0.69 (0.27)	0.71 (0.15)	0.58 (0.17)
Rural area	0.68 (0.26)	0.64 (0.16)*	0.58 (0.16)
Income[c]			
Less than €500	0.82 (0.13)	0.79 (0.18)	0.65 (0.22)
€,1000–1,499	0.57 (0.27)	0.64 (0.10)	0.54 (0.24)
€1,500–1,999	0.77 (0.16)	0.72 (0.11)	0.65 (0.09)
€2,000 or more	1.00 (0.00)**	1.00 (0.00)**	0.85 (0.00)*

[a] T test used; [b] ANOVA used; [c] Kruskal–Wallis test used; * $P < 0.001$; ** $P < 0.01$; *** $P < 0.10$

single, widowed, and divorced or separated people, these differences being significant (except in SF-6D).

Nonparametric tests showed that health utility values were significantly related to employment: those who were retired and housewives reported lower utility values than employed and unemployed people. We also found statistically significant differences between people with different levels of income: those who earned €2,000 or more showed higher levels of utility than the others.

Agreement among utility measures

SCA was applied to a hypercontingency table formed by the six dimensions of the SF-6D and the five dimensions of the EQ-5D, a 31 × 15 matrix. Displaying the instruments' dimensions, by deriving coordinates to represent categories of the variables involved, the first three axes explained 91% of the total variance (the first axis explained 72.5% of the total variance, the second 13.4%, and the third 5.2%). Figure 2 shows the strong contributions (above the mean) of the instruments' levels to the axes formation.

Table 6 synthesizes the strong associations between pairs of levels, obtained through each cell contribution to the chi-squared measure (dimensions with levels highly correlated are marked in grey). This table also shows the sense of the associations: direct associations are marked with (+) and inverse associations are marked with (−).

Figure 2 and Table 6 show a high inverse association between the levels "Your health does not limit you in vigorous activities" ($6D_{11}$) and "Some problems walking about" ($5D_{12}$), meaning that a person with no limitations in vigorous activities in SF-6D will not probably state some problems in walking in EQ-5D. On the contrary, a person with little limitation in bathing and dressing ($6D_{15}$) will probably refer to some problems in walking ($5D_{12}$). According to this line of thinking, someone with no limitations in vigorous activities ($6D_{11}$) will almost certainly refer to having no problems with performing usual activities related to work, study, housework, family or leisure activities ($5D_{31}$), and will not refer to some problems with performing usual activities in the same measure ($5D_{32}$). Also, someone with little limitation in bathing and dressing ($6D_{15}$) will answer that he/she is unable to perform usual activities ($5D_{33}$).

An individual without problems with his/her work or other regular daily activities as a result of physical health or any emotional problems ($6D_{21}$) will tend to say that he/she has no problems with performing usual activities ($5D_{31}$) and will not state problems with performing usual activities ($5D_{32}$). An individual who is limited in the kind of work or other activities as a result of his/her physical health and accomplishes less than he/she would like as a result of emotional problems ($6D_{24}$) will not mention having no problems with performing usual activities

($5D_{31}$), but will most probably say that he/she has some problems with performing usual activities ($5D_{32}$) or that he/she is unable to perform usual activities ($5D_{33}$).

Someone who is not limited in his/her social activities ($6D_{31}$) will not indicate that he/she has some problems walking about ($5D_{12}$). Similarly, a person who is limited in his/her social activities most of the time ($6D_{34}$) will not say that he/she has no problems walking about ($5D_{11}$), but will state some problems walking about ($5D_{12}$). A person not limited in his/her social activities ($6D_{31}$) will probably state no problems with performing usual activities ($5D_{31}$) and, naturally, will not refer to having some problems with performing usual activities ($5D_{32}$). Any person whose health limits his/her social activities most of the time ($6D_{34}$) will not only state that he/she has some problems with performing usual activities ($5D_{32}$), but also that he/she is unable to perform usual activities ($5D_{33}$) and will not, most certainly, state no problems with performing usual activities ($5D_{31}$).

In terms of pain, it is possible to see that someone who has no pain ($6D_{41}$) will report no pain or discomfort ($5D_{41}$) or moderate pain or discomfort ($5D_{42}$). Therefore, people reporting having pain that moderately interferes with their normal work ($6D_{44}$) will report moderate pain or discomfort ($5D_{42}$) and, obviously, they will not report no pain or discomfort in EQ-5D ($5D_{41}$). Hence, people reporting having pain that interferes quite a bit with their normal work ($6D_{45}$) will refer to having extreme pain or discomfort ($5D_{43}$).

Regarding mental health, answers referring to feeling tense or downhearted and low none of the time ($6D_{51}$) are directly related to people with no pain or discomfort ($5D_{41}$) and inversely related to people with moderate pain or discomfort ($5D_{42}$). Anyone who is feeling tense or downhearted and low most of the time ($6D_{54}$) will state having extreme pain or discomfort ($5D_{43}$). Feeling tense or downhearted and low none of the time ($6D_{51}$) means not being anxious or depressed ($5D_{51}$), since individuals' answers are related. However, naturally, this means the contrary of being moderately anxious or depressed ($5D_{52}$). Someone who is feeling tense or downhearted and low some of the time ($6D_{53}$) will not refer to being not anxious or depressed ($5D_{51}$), as will not refer to feeling tense or downhearted and low most of the time ($6D_{54}$). In fact, a person in this last situation ($6D_{54}$), will report being extremely anxious or depressed ($5D_{53}$).

Another pattern can be found in terms of vitality, since an individual who feels a lot of energy all of the time ($6D_{61}$) or a little of the time ($6D_{64}$), will not have some problems walking about ($5D_{11}$). On the contrary, anyone who reports having a lot of energy none of the time ($6D_{65}$) will report some problems walking about ($5D_{12}$) and will not say that he/she has no problems walking about ($5D_{11}$). Someone with a

Fig. 2 SCA scatter plot above the first and the second factorial axes (diamonds for SF-6D levels, filled triangles for EQ-5D levels). The levels are labeled 6D or 5D according to the instrument. We also used two indices in the labels, the first denoting the dimension and the second the level (for instance, $6D_{12}$ represents the second level of dimension 1 of SF-6D).
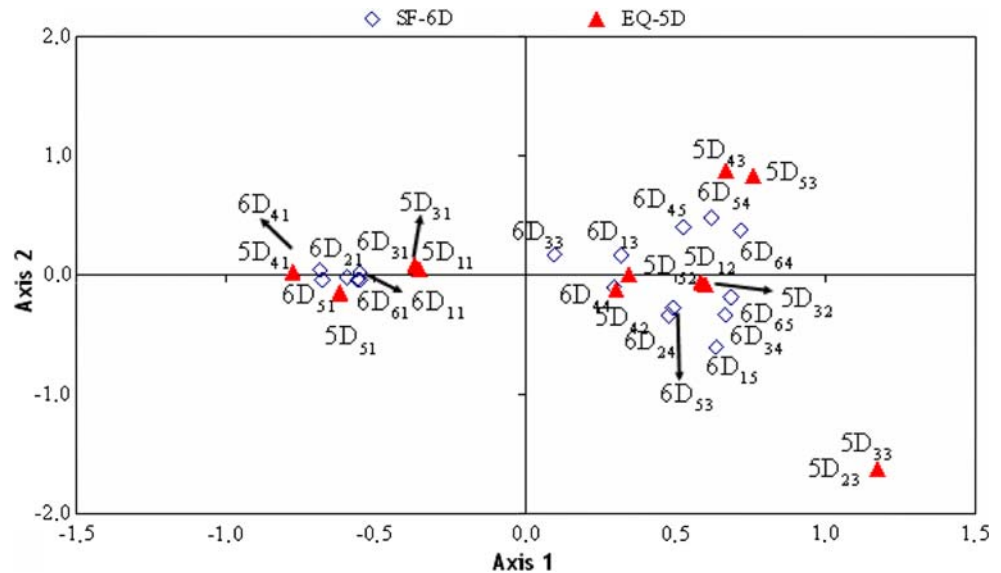
lot of energy all of the time ($6D_{61}$) refers to not having pain or discomfort ($5D_{41}$) and not having moderate pain or discomfort ($5D_{42}$). Having a lot of energy a little of the time ($6D_{64}$) is stated by those who experience extreme pain or discomfort ($5D_{43}$) and by those who are extremely anxious or depressed ($5D_{53}$). Similarly, a person who mentions a lot of energy all of the time ($6D_{61}$) will not refer to being moderately ($5D_{52}$) or extremely anxious or depressed ($5D_{53}$).

In a more detailed analysis at the level of the third main axis, we can also observe some more associations, shown in Table 6. In fact, it is possible to find a high positive association between the levels $6D_{32}$ and $5D_{31}$, meaning that a person whose health limits his/her social activities a little of the time ($6D_{32}$), will probably refer to not having problems in performing usual activities ($5D_{31}$).

Feeling tense or downhearted and low a little of the time ($6D_{52}$) means having moderate pain or discomfort ($5D_{42}$) and being moderately anxious or depressed ($5D_{52}$), since individuals' answers are related. And these individuals will not say they have no pain or discomfort ($5D_{41}$), as they will not state being extremely anxious or depressed ($5D_{53}$). Someone who has a lot of energy most of the time ($6D_{62}$) will not refer to having no pain or discomfort ($5D_{41}$) or being extremely anxious or depressed ($5D_{53}$). In fact, a person in this situation will report having moderate pain or discomfort ($5D_{42}$) and being moderately anxious or depressed ($5D_{52}$).

Clustering SF-6D and EQ-5D levels

The Ward method, as well as the furthest neighbour and the within-groups methods, pointed to five clusters of homogeneous levels. The solution of the k-means cluster analysis (Table 7) was similar, but more consistent.

Levels belonging to the same group are homogeneous, as they are associated to each other. The first group is mainly formed by levels denoting no problems in physical or mental health. Levels referring to some problems belong to the second group. The third cluster is mainly formed by levels related to a lot of problems in physical or mental health, while the forth cluster includes the levels which define extreme health problems. Finally, the fifth cluster gathers the levels that define very extreme health problems.

Discussion

In the literature of HRQL measures, there is an overall concern regarding differences in terms of results between instruments. Several studies that attempt to compare different instruments have been published [10, 12, 14–24, 27, 29, 30]. This study compares the SF-6D and the EQ-5D and investigates the differences in agreement between them. It also attempts to understand the possible reasons for the divergences found and to explore their implications. It was not our purpose just to compare the instruments, but also to apply different methodologies to understand the pattern of an individual when answering the SF-6D and EQ-5D, i.e., how would he/she respond for a certain dimension of EQ-5D, given that he/she gave a particular answer to a certain dimension of SF-6D. Although the SF-6D and EQ-5D provide very similar estimates at the mean level, the range and variability for EQ-5D were higher than those of SF-6D. The results showed evidence of a potential floor effect in the SF-6D and of a ceiling effect in the EQ-5D.

As expected, Spearman correlation coefficients revealed direct and high correlations between all the similar dimensions (physical functioning and mobility, physical functioning and usual activities, role limitations and usual activities, social functioning and mobility, social functioning and usual activities, pain and pain/discomfort, and mental

Table 6 Strong associations between the levels of the highly correlated dimensions

SF-6D			EQ-5D															
			M			SC			UA			PD			AD			
			11	12	13	21	22	23	31	32	33	41	42	43	51	52	53	
PF	11			-					+	-								
	12																	
	13																	
	14																	
	15			+							+							
	16																	
RL	21								+	-								
	22																	
	23																	
	24								-	+	+							
SF	31			-					+	-								
	32								+									
	33																	
	34		-	+					-	+	+							
	35																	
P	41											+	+					
	42																	
	43																	
	44											-	+					
	45													+				
MH	51											+	-		+	-		
	52											-	+			+	-	
	53														-			
	54													+		-	+	
	55																	
V	61			-								+	-		+	-	-	
	62											-	+			+	-	
	63																	
	64		-	+										+	+		+	
	65		-	+														

EQ-5D dimensions: M, mobility; SC, self-care; UA, usual activities; PD, pain/discomfort; AD, anxiety/depression

SF-6D dimensions: PF, physical functioning; RL, role limitations; SF, social functioning; P, pain; MH, mental health; V, vitality

Direct associations: +; inverse associations: −

health and anxiety/depression). There were also high correlations between mental health and pain/discomfort, between vitality and mobility, between vitality and pain/discomfort, and between vitality and anxiety/depression and low correlations between role limitations and self-care, physical functioning and self-care, and vitality and self-care.

Nonparametric tests showed that health utility values were significantly related to sex, age, marital status, educational level, employment status, residence, and income: women; patients aged 60 years or more; single, widowed, and divorced or separated people; patients with low educational levels; retired people and housewives; people living in rural areas; and those who earned less than €2,000 reported lower levels of utility than men; patients aged less than 60 years; those who were married or living together with someone else; patients with high educational levels; employed and unemployed individuals; people living in urban areas; and those who earned €2,000 or more.

SCA was used to assess the agreement among the instruments' descriptive systems and to investigate similarities between the levels of their dimensions. This enabled us to identify the levels most associated to each other and therefore to describe patterns of the individuals' answers. For instance, it is now possible to say that an individual who accomplishes less than he/she would like, as a result of emotional problems in the SF-6D, will not answer that he/she has no problems with performing usual activities in the EQ-5D, but will most probably say that he/she has some problems with performing usual activities or that he/she is unable to perform usual activities, in the EQ-5D.

Cluster analysis was used to classify SF-6D and EQ-5D levels into homogeneous groups. The first group is mainly formed by levels denoting no problems in physical or mental health and the second by levels referring to some problems. Levels related to a lot of problems in physical or mental health belong to the third cluster, while in the forth

Table 7 Clusters of SF-6D and EQ-5D levels

Group	Contents
1. No problems	Your health does not limit you in vigorous activities ($6D_{11}$)
	Your health limits you a little in vigorous activities ($6D_{12}$)
	No problems walking about ($5D_{11}$)
	You have no problems with your work or other regular daily activities as a result of your physical health or any emotional problems ($6D_{21}$)
	No problems with performing usual activities (e.g., work, study, housework family or leisure activities) ($5D_{31}$)
	Your health limits your social activities none of the time ($6D_{31}$)
	You have no pain ($6D_{41}$)
	You have pain but it does not interfere with your normal work (both outside the home and housework) ($6D_{42}$)
	No pain or discomfort ($5D_{41}$)
	You feel tense or downhearted and low none of the time ($6D_{51}$)
	Not anxious or depressed ($5D_{51}$)
	You have a lot of energy all of the time ($6D_{61}$)
2. Some problems	Your health limits you a little in moderate activities ($6D_{13}$)
	Your health limits you a lot in moderate activities ($6D_{14}$)
	No problems with self-care ($5D_{21}$)
	You are limited in the kind of work or other activities as a result of your physical health ($6D_{22}$)
	You accomplish less than you would like as a result of emotional problems ($6D_{23}$)
	Your health limits your social activities a little of the time ($6D_{32}$)
	Your health limits your social activities some of the time ($6D_{33}$)
	You have pain that interferes with your normal work (both outside the home and housework) a little bit ($6D_{43}$)
	You have pain that interferes with your normal work (both outside the home and housework) moderately ($6D_{44}$)
	Moderate pain or discomfort ($5D_{42}$)
	You feel tense or downhearted and low a little of the time ($6D_{52}$)
	Moderately anxious or depressed ($5D_{52}$)
	You have a lot of energy most of the time ($6D_{62}$)
3. A lot of problems	Some problems walking about ($5D_{12}$)
	Your health limits you a little in bathing and dressing ($6D_{15}$)
	You are limited in the kind of work or other activities as a result of your physical health and accomplish less than you would like as a result of emotional problems ($6D_{24}$)
	Some problems with performing usual activities ($5D_{32}$)
	Your health limits your social activities most of the time ($6D_{34}$)
	You have pain that interferes with your normal work (both outside the home and housework) extremely ($6D_{46}$)
	You feel tense or downhearted and low some of the time ($6D_{53}$)
	You feel tense or downhearted and low all of the time ($6D_{55}$)
	You have a lot of energy some of the time ($6D_{63}$)
	You have a lot of energy none of the time ($6D_{65}$)
4. Extreme problems	Some problems washing or dressing myself ($5D_{22}$)
	Your health limits your social activities all of the time ($6D_{35}$)
	You have pain that interferes with your normal work (both outside the home and housework) quite a bit ($6D_{45}$)
	Extreme pain or discomfort ($5D_{43}$)
	You feel tense or downhearted and low most of the time ($6D_{54}$)
	Extremely anxious or depressed ($5D_{53}$)
	You have a lot of energy a little of the time ($6D_{64}$)
5. Very extreme problems	Confined to bed ($5D_{13}$)
	Unable to wash or dress myself ($5D_{23}$)
	Unable to perform usual activities ($5D_{33}$)

cluster are the levels that define extreme health problems. Finally, the fifth cluster gathers the levels that define very extreme health problems.

It should, however be noted that the level "Your health limits you a lot in bathing and dressing" ($6D_{16}$) does not appear in this fifth group, where it should be. The explanation for this may be an incorrect answer to SF-6D [the four individuals who answered "Confined to bed" ($5D_{13}$) in the EQ-5D, chose the level above in the SF-6D] or that the last levels of mobility and self-care in EQ-5D and physical functioning in SF-6D do not measure the same concepts. Does this mean that the SF-6D is not able to identify very extreme problems, or is the explanation only the individuals' misunderstanding?

Both instruments showed consistency, namely in the agreement found in some dimensions and in some levels of each dimension. However, it seemed that they measure different concepts, at least to some extent. Actually, some levels of both instruments agreed, while others, contrarily to what was expected, disagreed. These findings generally support the results of Tsuchiya et al. [20] and Brazier et al. [14] in terms of the major difference between the two instruments: the differences in the descriptive systems account for at least a part of the major differences in the range of the two instruments. Indeed, using cluster analysis we found some levels from both instruments that were supposed to measure the same concepts, but where individuals answered in a different way. This means that apparently similar levels are in fact different and contribute to the differences found in terms of the indices computed from both descriptive systems. These differences between the descriptive systems of the instruments found in our study should be further investigated.

Conclusion

The aim of this paper was to compare EQ-5D and SF-6D and to investigate the differences in agreement between them. It was also our purpose to understand the possible reasons for divergences found and to explore their implications. To our knowledge, no study has yet compared EQ-5D and SF-6D in terms of their descriptive systems, analyzing the extent of agreement or disagreement of their dimensions' levels.

A major strength of the research reported here is its newness in terms of the way the comparison was addressed and the type of methodology used. Over the past years there have been other studies comparing SF-6D and EQ-5D [14–24, 27, 30]. Whereas most of them present comparisons in terms of dimensions, none of them compared the levels of those dimensions. Moreover, none presented findings reporting the probable pattern of the answers of a

particular individual to a certain level of a dimension of EQ-5D, given that he/she gave a particular answer to a certain level of a dimension of SF-6D. Bearing in mind that the methods adopted in this paper have not been widely used before in looking at the comparison of preference-based instruments, it can be said that its originality has to be balanced by a negative counterpart. In fact, though the methods used are robust and applicable to this problem, the existence of several levels in both instruments (31 in SF-6D and 15 in EQ-5D) leads to several possible relations between them, generating sctatterplots that are not easy to analyze. Although we only analyzed the strong associations between the instruments' levels, it was still a real challenge to identify the patterns of the individuals' answers. However, it is our conviction that this research identified some areas of agreement, as well as disagreement, between both instruments, and we hope that it helps shed some light on the issue of the comparability between instruments, which is a topic currently in vogue in the HRQL literature. Another limitation of the research is that data were collected from a relatively small and unusual sample of respondents. Whilst the type of patients could condition the results, since it is an elderly population, it should, however, be stressed that this could also be seen as a strength of the study, since studies comparing EQ-5D and SF-6D using data from old and visually impaired patients are not common. Nevertheless, to address this issue we intend to apply this methodology to a larger sample from the general population. Patients suffering from other diseases should also be used in future analysis to confirm the consistency of the findings reported herein.

This study provided evidence that both instruments are consistent, although it seems that they measure different concepts, at least to some extent. These findings are particularly important since these instruments are usually employed in HRQL studies and in economic evaluations. Furthermore, this reinforces the importance of the research carried out on mapping between instruments [27, 31–33] and the need for more investigation in this field. Further research is needed to overcome the differences between EQ-5D and SF-6D: revisions of one or both descriptive systems or of their scoring algorithm are necessary to enable the interchangeably use of both instruments. Brazier et al. [14] suggest adding more intermediate levels to the EQ-5D or adding lower levels to the SF-6D dimensions, at least for the physical functioning and role limitations. Our current research is centred on this last suggestion of those authors—adding lower levels to two of the SF-6D dimensions—in order to correct its floor effect and to try to have extreme levels similar in both instruments.

Further studies should compare the performance of the SF-6D with that of other preference-based measures, such as HUI, and compare utility scores provided by SF-6D,

EQ-5D and HUI with the ones obtained by elicitation techniques, such as SG or TTO. In fact, there is already some literature on these matters and on mapping between instruments [14–21, 27–33], although not specifically comparing the SF-6D utility scores to utilities generated by SG or TTO.

References

1. Brooks, R. (1996). EuroQol: The current state of play. *Health Policy (Amsterdam), 37*, 53–72. doi:10.1016/0168-8510(96)00822-6.

2. Dolan, P. (1997). Modelling valuations for EuroQol health states. *Medical Care, 35*, 1095–1108. doi:10.1097/00005650-199711000-00002.

3. Kind, P., Hardman, G., & Macran, S. (1999). *UK Population Norms for EQ-5D*. Discussion Paper 172. University of York: Centre for Health Economics.

4. Torrance, G., Furlong, W., Feeny, D., & Boyle, M. (1995). Multi-attribute preference functions: Health utilities index. *Pharmaco-Economics, 7*, 503–520. doi:10.2165/00019053-199507060-00005.

5. Torrance, G., Feeny, D., Furlong, W., Barr, R., Zhang, Y., & Wang, Q. (1996). Multi-attribute utility function for a comprehensive health status classification system: Health utilities index mark 2. *Medical Care, 34*(7), 702–722. doi:10.1097/00005650-199607000-00004.

6. McCabe, C., Stevens, K., Roberts, J., & Brazier, J. (2005). Health state values for the HUI2 descriptive system: Results from a UK survey. *Health Economics, 14*, 231–244. doi:10.1002/hec.925.

7. Brazier, J., Usherwood, T., Harper, R., & Thomas, K. (1998). Deriving a preference-based single index from the UK SF-36 health survey. *Journal of Clinical Epidemiology, 51*(11), 1115–1128. doi:10.1016/S0895-4356(98)00103-6.

8. Brazier, J., Roberts, J., & Deverill, M. (2002). The estimation of a preference-based measure of health from the SF-36. *Journal of Health Economics, 21*, 271–292. doi:10.1016/S0167-6296(01)00130-8.

9. Kaplan, R. M., Bush, J. W., & Berry, C. C. (1976). Health status: Types of validity and the index of well-being. *Health Services Research, 11*(4), 478–507.

10. Kaplan, R. M., Ganiats, T. G., Sieber, W. J., & Anderson, J. P. (1998). The quality of well-being scale: Critical similarities and differences with SF-36. *International Journal for Quality in Health Care, 10*, 509–520. doi:10.1093/intqhc/10.6.509.

11. Anderson, J. P., Kaplan, R. M., Berry, C. C., Bush, J. W., & Rumbaut, R. G. (1989). Interday reliability of function assessment for a health status measure: The quality of well-being scale. *Medical Care, 27*, 1076–1083. doi:10.1097/00005650-198911000-00008.

12. Osborne, R., Hawthorne, G., Lew, E., & Gray, L. (2003). Quality of life assessment in the community-dwelling elderly: Validation of the assessment of quality of life (AQoL) instrument and comparison with the SF-36. *Journal of Clinical Epidemiology, 56*(2), 138–147. doi:10.1016/S0895-4356(02)00601-7.

13. Hawthorne, G., & Osborne, R. (2005). Population norms and meaningful differences for the assessment of quality of life (AQoL) measure. *Australian and New Zealand Journal of Public Health, 29*(2), 136–142. doi:10.1111/j.1467-842X.2005.tb00063.x.

14. Brazier, J., Roberts, J., Tsuchiya, A., & Busschbach, J. (2004). A comparison of the EQ-5D and SF-6D across seven patient groups. *Health Economics, 13*, 873–884. doi:10.1002/hec.866.

15. Petrou, S., & Hockley, C. (2005). An investigation into the empirical validity of the EQ-5D and SF-6D based on hypothetical preferences in a general population. *Health Economics, 14*(11), 1169–1189. doi:10.1002/hec.1006.

16. Stavem, K., Frøland, S. S., & Hellum, K. B. (2005). Comparison of preference-based utilities of the 15D, EQ-5D and SF-6D in patients with HIV/AIDS. *Quality of Life Research, 14*, 971–980. doi:10.1007/s11136-004-3211-7.

17. Lamers, L., Bouwmans, C., van Straten, A., Donker, M., & Hakkaart, L. (2006). Comparison of EQ-5D and SF-6D utilities in mental health patients. *Health Economics, 15*(11), 1229–1236. doi:10.1002/hec.1125.

18. Marra, C., Woolcott, J., Kopec, J., Shojania, K., Offer, R., Brazier, J., et al. (2005). A comparison of generic, indirect utility measures (the HUI2, HUI3, SF-6D, and the EQ-5D) and disease-specific instruments (the RAQoL and The HAQ) in rheumatoid arthritis. *Social Science & Medicine, 60*, 1571–1582. doi:10.1016/j.socscimed.2004.08.034.

19. Feeny, D., Wu, L., & Eng, K. (2004). Comparing short form 6D, standard gamble and health utilities index mark 2 and mark 3 utility scores: Results of total hip arthroplasty patients. *Quality of Life Research, 13*(10), 1659–1670. doi:10.1007/s11136-004-6189-2.

20. Tsuchiya, A., Brazier, J., & Roberts, J. (2006). Comparison of valuation methods used to generate the EQ-5D and the SF-6D value sets. *Journal of Health Economics, 25*(2), 334–346. doi:10.1016/j.jhealeco.2005.09.003.

21. Marra, C., Esdaile, J., Guh, D., Kopec, J., Brazier, J., & Koehler, B. (2004). Chalmers, A., Anis, A. A comparison of four indirect methods of assessing utility values in rheumatoid arthritis. *Medical Care, 42*(11), 1125–1131. doi:10.1097/00005650-200411000-00012.

22. Pickard, A., Simon, J., Jeffrey, A., & Feeny, D. H. (2005). Responsiveness of generic health-related quality of life in stroke. *Quality of Life Research, 14*, 207–219. doi:10.1007/s11136-004-3928-3.

23. Longworth, L., & Bryan, S. (2003). An empirical comparison of EQ-5D and SF-6D in liver transplant patients. *Health Economics, 12*(12), 1061–1067. doi:10.1002/hec.787.

24. Hawthorne, G., Richardson, J., & Atherton Day, N. (2001). A comparison of the assessment of quality of life (AQoL) with four other generic utility instruments. *Annals of Medicine, 33*, 358–370. doi:10.3109/07853890109002090.

25. Kopec, J., & Willison, K. (2003). A comparative review of four preference-weighted measures of health-related quality of life. *Journal of Clinical Epidemiology, 56*(4), 317–325. doi:10.1016/S0895-4356(02)00609-1.

26. Everitt, B. S., & Dunn, G. (2001). *Applied Multivariate Data Analysis*. London: Arnold.

27. Bryan, S., & Longworth, L. (2005). Measuring health-related utility: Why the disparity between EQ-5D and SF-6D? *The European Journal of Health Economics, 50*, 253–260. doi:10.1007/s10198-005-0299-9.

28. Holland, R., Smith, R., Harvey, I., Swift, L., & Lenaghan, E. (2004). Assessing quality of life in the elderly: A direct

comparison of the EQ-5D and AQoL. *Health Economics, 13*(8), 793–805. doi:10.1002/hec.858.

29. O'Brien, B., Spath, M., Blackhouse, G., Severens, J., Dorian, P., & Brazier, J. (2003). A view from the bridge: Agreement between the SF-6d utility algorithm and the health utilities index. *Health Economics, 12*(11), 975–981. doi:10.1002/hec.789.

30. Gerard, K., Nicholson, T., Mulle, M., Mehta, R., & Roderick, P. (2004). EQ-5D versus SF-6D in an older, chronically Ill patient group. *Applied Health Economics and Health Policy, 3*(2), 91–102. doi:10.2165/00148365-200403020-00005.

31. Franks, P., Lubetkin, E., Gold, M., Tancredi, D., & Jia, H. (2004). Mapping the SF-12 to the EuroQol EQ-5D index in a national US

sample. *Medical Decision Making, 24*(3), 247–254. doi:10.1177/0272989X04265477.

32. Gray, A., Rivero-Arias, O., & Clarke, P. (2006). Estimating the association between SF-12 responses and EQ-5D utility values by response mapping. *Medical Decision Making, 26*(18), 18–29. doi:10.1177/0272989X05284108.

33. Tsuchiya, A., Brazier, J., McColl, E., & Parkin, D. (2002). Deriving preference-based single indices from non-preference based condition-specific instruments: Converting AQLQ into EQ5D indices. Discussion Paper 02/1. The University of Sheffield: Sheffield Health Economics Group.