The HKU Scholars Hub   The University of Hong Kong   香港大學學術庫

Title	**A Comparison between the EQ-5D and the SF-6D in Patients with Chronic Obstructive Pulmonary Disease (COPD)**
Author(s)	**Chen, J; Wong, CKH; McGhee, S; Pang, PKP; Yu, WC**
Citation	**PLoS ONE, 2014, v. 9 n. 11, p. e112389**
Issued Date	**2014**
URL	**http://hdl.handle.net/10722/206880**
Rights	**Creative Commons: Attribution 3.0 Hong Kong License**

A Comparison between the EQ-5D and the SF-6D in Patients with Chronic Obstructive Pulmonary Disease (COPD)

Jing Chen[1]*, Carlos K. H. Wong[2], Sarah M. McGhee[1], Polly K. P. Pang[3], Wai-Cho Yu[3]

1 School of Public Health, The University of Hong Kong, Hong Kong, China, 2 Department of Family Medicine and Primary Care, The University of Hong Kong, Hong Kong, China, 3 Department of Medicine and Geriatrics, Princess Margaret Hospital, Hong Kong, China

Abstract

Background: The appropriate use of generic preference-based measures determines the accuracy of disease assessment and further decision on healthcare policy using quality adjusted life years. The discriminative capacity of different instruments would differ across disease groups. Our study was to examine the difference in utility scores for COPD patients measured by EQ-5D and SF-6D and to assist the choice of a proper instrument in this disease group.

Methods: Differences of mean utility scores of EQ-5D and SF-6D in groups defined by socio-demographic characteristics, comorbidities, health service utilisation and severity of illness were tested using Mann-Whitney test, t-test, Kruskal-Wallis test, Pearson's correlation coefficient and ANOVA, as appropriate. The discriminative properties of the two instruments were compared against indicators of quality of life using receiver operating characteristic curves. The statistical significance of the area under the curves (AUC) was tested by ANOVA and F-statistics used to compare the efficiency with which each instrument discriminated between disease severity groups.

Results: Mean utility scores of EQ-5D and SF-6D were 0.644 and 0.629 respectively in the 154 subjects included in the analysis. EQ-5D scores were significantly higher than SF-6D in groups less severe and these differences corresponded to a minimally important difference of greater than 0.03 ($p < 0.001$). EQ-5D and SF-6D scores were strongly correlated across the whole sample ($r = 0.677$, $p < 0.001$) and in pre-defined groups ($r > 0.5$ and $p < 0.05$ for all correlation coefficients). AUCs were above 0.5 against the indicators of health-related quality of life for both instruments. F-ratios suggested SF-6D was more efficient in discriminating cases of different disease severity than EQ-5D.

Conclusions: Both EQ-5D and SF-6D appeared to be valid preference-based measures in Chinese COPD patients. SF-6D was more efficient in detecting differences among subgroups with differing health status. EQ-5D and SF-6D measured different things and might not be used interchangeably in COPD patients.

Background

There is an increasing need to express effectiveness of interventions in terms of quality-adjusted life years (QALYs) in health economic evaluations. The QALY accounts for both length of life and health-related quality of life (HRQoL) in a single outcome measure. Preference-based measures of health are designed to construct QALYs. They consist of two parts, one a standardized descriptive system for people to describe their own health state, the other an algorithm for assigning values to each state. The algorithm is usually based on valuations from a general population sample of adults which are derived using techniques such as standard gamble (SG) or time trade off (TTO). Preference-based measures score health states on a scale of 0 to 1 equivalent to death and full health respectively [1].

There are several preference-based measures including the EQ-5D [2], the 15D [3], Health utility index (HUI) [4], Quality of Well-Being (QWB) [5] and most recent of all, the SF-6D [6]. The EQ-5D is currently the most popular preference-based instrument, used in about half of published studies [7]. It is a two part instrument, with the first part describing problems in five dimensions with three response levels for each. The second part consists of a vertical visual analogue scale (VAS) with zero at the bottom referring to the worst health state and 100 at the top referring to the best health state. The participant pointed to the level on the scale indicating their overall health on the day of the interview. The five dimensions (mobility, self-care, usual activities, pain/discomfort, anxiety/depression) each with three levels (no problems, some problems, extreme problems) can thus define 243 (3^5) different health states [1]. Using the TTO technique, country specific scoring algorithms have been developed for the EQ-5D in

UK [8], Norway [9], and Japan [10]. The EQ-5D has been translated into Chinese and used in China, Taiwan, Hong Kong, Singapore and Malaysia [11]. The SF-6D, which is increasingly used, was developed from the SF-36 [6]. Six dimensions (physical functioning, role limitation, social functioning, pain, mental health, vitality) were retained from the original SF-36, each with 4 to 6 levels and it can define 18 000 (6*6*5*5*5*4) different health states. Using the standard gamble (SG) technique, a scoring algorithm for the SF-6D was developed for the UK [6,12].

The EQ-5D and the SF-6D differ from one another in terms of their valuation techniques, classification systems, dimensions and items covered. There is therefore a possibility that they will generate a different utility score for a specific health state. Several studies have directly compared these two instruments in different populations, and these generally show disparities in the utility scores derived from the EQ-5D and the SF-6D [12–20]. The EQ-5D tends to generate higher scores than the SF-6D in healthy population or patients with mild conditions and vice versa in less healthy populations or among patients with severe conditions [13,17,19], leading to variation in the estimation of utility scores and QALYs which could influence decisions taken on resource use [21–23].

Harper *et al* showed that the SF-36 performed better in detecting changes in HRQoL in patients with mild COPD in outpatient settings than did the EQ-5D [24]. However, no study to our knowledge has directly compared the performance of EQ-5D and SF-6D for those with more severe stages of COPD. In order to assist clinicians and researchers in choosing the most appropriate instrument for discriminative and evaluative purposes we performed a head-to-head comparison of the EQ-5D and the SF-6D in a group of severe and very severe COPD patients. COPD was chosen because it is known to have a significant adverse impact on HRQoL [25–30]. The burden of the disease in terms of morbidity and mortality is projected to increase in the next few decades in East Asia and elsewhere making the ability to determine QALY loss important [31,32], thus a precise measure is required in estimate of QALYs and to better inform the healthcare policy using quality-adjusted life years as outcome.

Methods

Sample and data collection

The study was conducted between September 2010 and May 2011 in a respiratory specialist out-patient clinic (SOPC) in a public hospital (Princess Margret Hospital) in Hong Kong. All the patients with COPD who regularly attended the SOPC were screened by their medical records for the study. Patients who had severe (post-bronchodilator FEV_1 30–49% of predicted) or very severe (post-bronchodilator $FEV_1\% < 30\%$ of predicted) COPD according to the Global Obstructive Lung Disease (GOLD) standard [33] during the past one year were included. Those who had significant comorbidities which prevented their participation, such as dementia or Parkinson's disease, who did not have a pulmonary function test result within the past one year or who did not speak Cantonese were excluded and all the eligible patients were invited to participate in the study. From 166 patients approached, 2 patients were incommunicable (one was deaf and one was unable to speak), 5 refused to participate, and 159 (95.8%) provided written consent to participate and were interviewed by trained interviewers. Five of the interviewed participants were excluded from the analysis because their post FEV_1 was 50% of predicted which classified their severity stage as moderate. Thus, 154 patients were included in the final data analysis.

Subjects completed the respiratory disease specific Saint George's Respiratory Questionnaire (SGRQ), the SF-6D and the EQ-5D. They also answered questions on socio-demographic information (sex, age, education, housing type and income), self-rated health status ("Compared with past 12 months, what do you think about your present health condition? Better, The same, Worse") and life style information (e.g. smoking). The SF-6D and the EQ-5D were asked in a random sequence to avoid possible order effects. Utilization of health services (hospital admissions in the past 12 month) and presence of comorbidities (i.e. other than COPD, whether being diagnosed with hypertension, heart diseases, pneumonia, diabetes, asthma, rheumatoid arthritis, liver disease, cancers and other chronic diseases) were extracted directly from patients' medical records by one research nurse in the SOPC.

The study was approved by the Kowloon West Cluster Clinical Research Ethics Committee, Hong Kong. All participants provided written informed consent to participate in this study.

HRQoL Measures

The Cantonese Chinese version of the EQ-5D and the Chinese (Hong Kong) SF-6D questionnaire were used to measure the Health Related Quality of Life (HRQoL) of the sample patients [34]. Because a Hong Kong specific scoring algorithm was not available for the EQ-5D, the UK one was used for which direct valuations of 43 health states had been elicited from a sample of the general population using TTO and the theoretic utility scores ranged from −0.59 to 1 [35].

Direct valuations for 196 health states from the SF-6D HK had been valued by a representative sample of the Hong Kong population (n = 582) using SG and the theoretic utility scores ranged from 0.315 to 1 [36]. This algorithm was employed in the present study to generate utility scores but, in order to compare with scores from the EQ-5D, the UK scoring algorithm for the SF-6D [6] was also used.

The Hong Kong Chinese version of the St. George's Respiratory Questionnaire (SGRQ) was included as a measure of disease specific quality of life. The SGRQ contains 76 weighted responses divided into three sections covering distress due to respiratory symptoms, disturbance of physical activity and overall impact on daily life and well-being. A total score from 0 to 100 is summed from weighted positive responses and a higher score indicates worse HRQoL [37]. The Chinese version of the SGRQ had already been validated in patients with COPD in Hong Kong [38] and, although the weights applied to the responses have not been developed in Hong Kong, they appear to be similar in different countries and in different languages [39] so cultural or linguistic factors may not greatly affect responses to this questionnaire [40].

Statistical analysis

Utility scores for the EQ-5D and the SF-6D were both estimated using UK-derived scoring algorithms for direct comparison and a SGRQ total summary score was calculated. Possible ceiling and floor effects, represented by the proportion of participants with the best and worst theoretical scores respectively, were identified. Correlation between the SGRQ scores and EQ-5D or SF-6D or EQ-VAS scores were tested using Pearson's correlation coefficient.

The EQ-5D, SF-6D and SGRQ mean scores were compared for different groups defined by levels of socio-demographic variables, comorbidities or health service utilisation. Where the relevant variable had two levels, this between-group comparison used the Mann-Whitney test for the EQ-5D or t-test for the SF-6D

Table 1. Characteristics of the study sample (n = 154).

	Severe COPD		Very severe COPD		Total	
	n	%	n	%	n	%
Number of patients	100	65.0	54	35.0	154	100.0
Age (yrs, mean ±, sd)[¶]	74.8±7.9		69.4±7.4		72.9±8.1	
Gender (male)	98	98	54	100	152	98.7
COPD duration						
(yrs, mean ± sd)	6.8±4.2		7.9±4.8		7.2±4.4	
Lung function: post FEV$_1$% (mean ± sd)[¶]	38.1±5.7		22.6±5.1		32.7±9.2	
Education						
No/pre-primary	30	30.0	7	13.0	37	24.0
Primary	49	49.0	31	57.4	80	52.0
Secondary	20	20.0	15	27.8	35	22.7
Post-secondary	1	1.0	1	1.9	2	1.3
Monthly family income (HK$)						
No income	73	73.0	33	61.1	106	68.8
<$10,000	26	26.0	20	37.0	46	29.9
>$10,000	1	1.0	1	1.9	2	1.3
Housing						
Public or aided	76	76.0	47	87.0	123	79.9
Private or Old people's home	24	24.0	7	13.0	31	20.1
Smoking[‡]						
Current smoker	16	16.0	3	5.6	19	12.3
Ex-smoker	83	83.0	47	87.0	130	84.4
Never smoker	1	1.0	4	7.4	5	3.3
Comorbidity[II]						
No	35	35.0	15	27.8	50	32.5
Yes	65	65.0	39	72.2	104	67.5
Hospital admission in the past 12 month						
No	33	33.0	10	18.5	43	27.9
Yes	67	67.0	44	81.5	111	72.1
Perceived health (compared to previous year)						
Better	11	11.0	4	7.4	15	9.7
Same	18	18.0	7	13.0	25	16.2
Worse	71	71.0	43	79.6	114	74.0

[II]Presence with any following disease: hypertension, heart diseases, pneumonia, diabetes, asthma, rheumatoid arthritis, liver disease, cancers and other chronic diseases.
[¶]$p < 0.001$, significance was tested by t-test.
[‡]$p < 0.05$, significance was tested by chi-square test.
Note: COPD = Chronic Obstructive Pulmonary Disease.
doi:10.1371/journal.pone.0112389.t001

and SGRQ. Where the relevant variable had three or more levels, this comparison used the Kruskal-Wallis test for the EQ-5D scores or the one way analysis of variance (ANOVA) for SF-6D scores and SGRQ. Association between the EQ-5D and SF-6D mean scores of each level of each variable was identified using Pearson's correlation coefficient and statistical significance of the difference was assessed by paired-sample t-test. A minimally important difference (MID) was defined as a difference of at least 0.03 between the EQ-5D and the SF-6D mean scores [20,41,42].

To determine whether the mean utility scores of the EQ-5D and SF-6D followed the pattern that less severe patients had higher scores on the EQ-5D and more severe patients scored higher on SF-6D, the correlations between EQ-5D and SGRQ as well as SF-6D and SGRQ were tested by simple linear regression. The

difference between the SF-6D and EQ-5D scores was regressed to SGRQ scores to further explore whether the two utility scores vary with severity differently. SGRQ was used as an external indicator of disease severity in COPD patients.

The discriminative properties of the EQ-5D and SF-6D were compared using receiver operating characteristic (ROC) curves. The area under the ROC curve (AUC) indicates the possibility of correctly discriminating between dichotomized groups (e.g. "severe" vs. "very severe"). In our analysis, the performance of the EQ-5D and SF-6D was evaluated against a couple of commonly used external indicators of HRQoL: the EQ-VAS scores and the SGRQ summary scores. These indicators were dichotomized using different cut-off points. EQ-VAS scores and SGRQ scores was grouped according to the percentiles of the
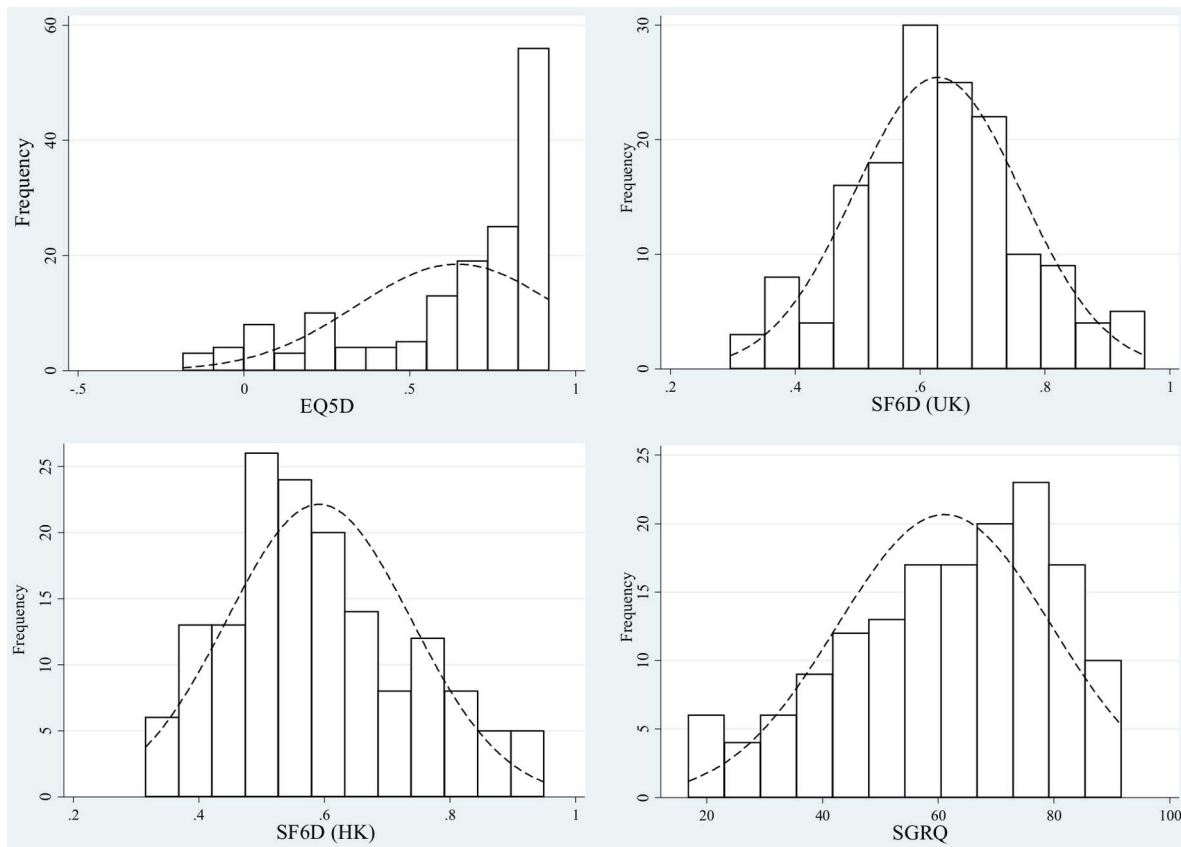
Figure 1. Distribution of EQ-5D, SF-6D (UK), SF-6D (HK) and SGRQ scores.
doi:10.1371/journal.pone.0112389.g001

score distribution. The statistical significance of the AUC (F-statistic) was tested by ANOVA for each instrument. F-statistics for the SF-6D and EQ-5D were compared by dividing the F-statistic of the SF-6D by that of the EQ-5D (F-ratio). A value greater than one indicates that the SF-6D is more efficient than the EQ-5D at detecting differences in quality of life. The significance level was set as p<0.05 and all analyses were performed using STATA (StataCorp. 2013. Stata Statistical Software: Release 13. College Station, TX: StataCorp LP).

Results

Almost all the patients were male (98.7%), 65.0% were in the severe COPD stage and, on average, had been diagnosed with COPD for 7 years (SD 4.4) (Table 1). More than 80% were ex-

smokers while 12% were current smokers. Fewer than 30% had secondary education or higher and 68.8% had no family income while 67.5% of them were with comorbid conditions and 72.1% had been admitted to hospital at least once in the past 12 months. Around three-quarters of the group (74.0%) reported their health to be worse compared with the previous year.

The mean scores on the EQ-5D and the SF-6D (UK) were 0.644 and 0.629 respectively. The distribution of EQ-5D, SF-6D (UK and HK) and SGRQ scores were generally skewed, particularly noticeable for the EQ-5D (Figure 1). The EQ-5D showed a ceiling effect with 22.1% of patients reporting the highest possible scores. No ceiling effect was observed in the SF-6D or SGRQ. Both EQ-5D and SF-6D scores were negatively associated with SGRQ since high scores indicated good HRQoL on EQ-5D and SF-6D but poor HRQoL on the SGRQ. Correlation was

Table 2. Distribution of EQ-5D, SF-6D (UK & HK), EQ-VAS and SGRQ scores in the sample subjects and their correlations with SGRQ.

Measures	Theoretical range	Observed range	Mean ± sd	Ceiling effect (%)	Floor effect (%)	Pearson correlation[¶]
EQ-5D	−0.590, 0.919	−0.184, 0.919	0.644±0.306	34 (22.1)	0	−0.583
SF-6D (HK)	0.315, 1	0.315, 0.950	0.591±0.147	0	2 (1.3)	−0.745
SF-6D (UK)	0.296, 1	0.296, 0.960	0.629±0.133	0	1 (0.6)	−0.728
EQ-VAS	0, 100	0, 100	55.28±20.42	4 (2.6)	5 (3.2)	−0.437
SGRQ	100, 0	91.62, 16.84	61.07±18.53	0	0	–

[¶]EQ-5D, SF-6D and EQ-VAS correlated with SGRQ; p<0.001 for all Pearson's correlation coefficient.
Note: SGRQ = Saint George's Respiratory Questionnaire.
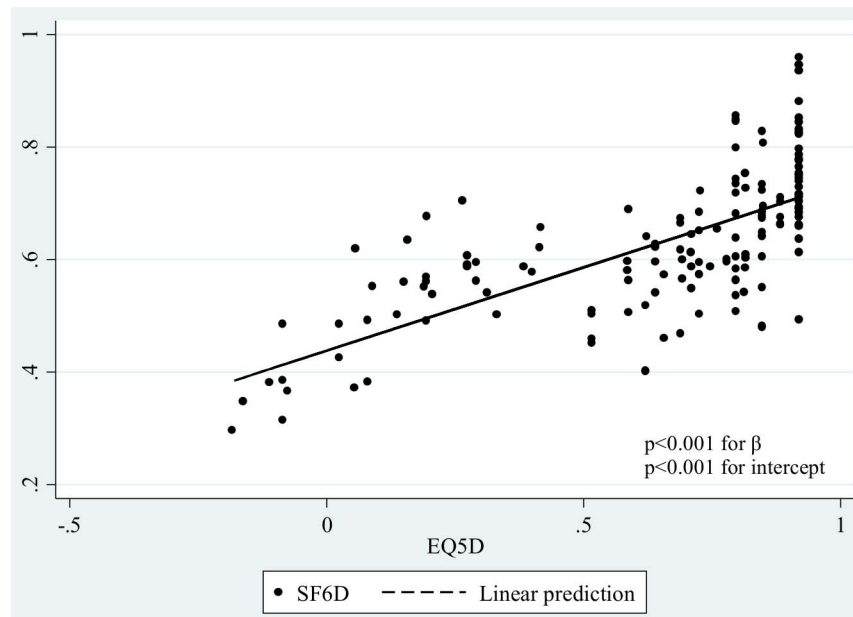doi:10.1371/journal.pone.0112389.t002

Figure 2. Scatter plot of EQ-5D and SF-6D.
doi:10.1371/journal.pone.0112389.g002

higher between SF-6D (HK and UK) and SGRQ (r = −0.745 and −0.728, p<0.001) than between EQ-5D and SGRQ (r = −0.583, p<0.001) (Table 2).

The EQ-5D and SF-6D mean scores were strongly correlated across the whole sample (r = 0.677, p<0.001) and a scatter plot further showed this when regressing EQ-5D to SF-6D (the coefficient ß = 0.30, p<0.001, and the constant = 0.44, p< 0.001) (Figure 2). In subgroups defined by disease severity, socio-demographic characteristics and utilization of health services, the EQ-5D and SF-6D mean scores were strongly correlated as well (Table 3). Both measures were sensitive to differences within subgroups in disease severity, hospital admission and presence or not of comorbidities and differences detected were statistically significant.

Direct comparison of the EQ-5D and SF-6D showed that in the less severe cases indicated by lower SGRQ scores, the EQ-5D scores were higher than SF-6D (Figure 3 above). On the contrary, the SF-6D scores were higher than EQ-5D in more severe cases indicated by a higher SGRQ scores. There appeared a cross-over of these two utility scores. The correlation between the difference of the SF-6D and EQ-5D (SF-6D – EQ-5D) and SGRQ score was positive (β = 0.004, p<0.001) (Figure 3 below). The difference in most cases (88.3%) were greater than the MID of 0.03.

The AUC showed the discriminative ability of the EQ-5D and SF-6D to distinguish between very severe and severe COPD as defined by EQ-VAS and SGRQ summary scores. Against both indicators of HRQoL, the AUC scores for both instruments were above 0.5 suggesting that they can detect the difference between severe and very severe cases (Table 4). For most of the comparisons, the SF-6D generated a larger AUC than the EQ-5D. Most of the F-ratios was greater than 1 which showed SF-6D to be more efficient in discriminating cases of differed disease severity than EQ-5D.

Discussion

It is important to compare the available preference-based measures of health because the choice of instrument can greatly influence the eventual decision-making based on the results. Different instruments can have differing sensitivity to disease-related quality of life impacts and can display ceiling or floor effects in some populations. This comparative study of the EQ-5D and the SF-6D in people with severe or very severe COPD confirmed the ceiling effect of the EQ-5D, that is, a substantial group of subjects scored the highest possible score, even though all had COPD classed as severe by lung function tests. This ceiling effect of the EQ-5D has been noted in other studies [12,13,16]. Both of the instruments demonstrated a moderate capacity to distinguish between previously defined groups in terms of disease severity as assessed by lung function, by whether they had a recent hospital admission and by whether they had any comorbidities. Neither instrument was found to be clearly superior. When compared with the SGRQ, the SF-6D showed stronger correlation coefficients and generated a larger AUC than EQ-5D implying that SF-6D may be better at reflecting the health status of those with severe COPD.

Both the EQ-5D and the SF-6D discriminated well between pre-defined subgroups with mainly significant differences between mean scores for groups defined by disease severity, hospital admissions and presence of comorbidities. Thus there is evidence of construct validity for each measure in identifying HRQoL of people with severe and very severe COPD. The EQ-5D and SF-6D were strongly positively correlated (Pearson's correlation coefficient >0.5 and statistically significant) throughout the entire sample and pre-defined subgroups. Despite this, the difference in mean scores on the two measures in pre-defined subgroups often constituted a MID.

Even though the EQ-5D and SF-6D generated similar mean scores in our sample, the distributions were very different. The SF-6D utility scores followed a relative normal distribution, unlike the EQ-5D scores which were negatively skewed with a clustering at the maximum score of one. It was clear from the analysis that the EQ-5D appeared with larger ceiling effect (22.1%) compared with the SF-6D (UK and HK) in our sample. Previous studies in populations with chronic diseases have also reported ceiling effects for the EQ-5D [43–45]. It suggested that the EQ-5D, in which

Table 3. EQ-5D and SF-6D (UK) mean scores, mean score difference between EQ-5D and SF-6D, and correlation between EQ-5D and SF-6D scores by group.

Variables	n (%)	Mean scores			Mean score difference (EQ-5D – SF-6D)		Pearson correlation between EQ-5D and SF-6D	
		SGRQ	EQ-5D	SF-6D	Estimate	p value for md[#]	Estimate	p value for pcc
Total sample	154	61.07	0.644	0.629	0.015	0.21	0.677	<0.001
Stage of COPD								
Severe	100 (64.9)	57.76	0.686	0.646	0.041	0.03	0.688	<0.001
Very severe	54 (35.1)	67.19	0.565	0.597	−0.032	0.19	0.633	<0.001
p value[§]		0.001	0.007	0.015				
Age								
50–59	12 (7.8)	58.27	0.808	0.650	0.158	<0.001	0.684	0.014
60–69	39 (25.3)	61.69	0.655	0.611	0.044	0.12	0.652	<0.001
70–79	75 (48.7)	61.43	0.611	0.629	−0.018	0.27	0.716	<0.001
80+	28 (18.2)	60.43	0.646	0.643	0.002	0.48	0.701	<0.001
p value[II]		0.98	0.54	0.72				
Education								
No/pre-primary	37 (24.0)	62.32	0.618	0.610	0.008	0.42	0.784	<0.001
Primary	80 (52.0)	62.72	0.657	0.634	0.022	0.19	0.657	<0.001
Secondary	35 (22.7)	56.76	0.630	0.630	0.0003	0.50	0.604	<0.001
Post-sec/degree	2 (1.3)	47.20	0.833	0.708	0.125	0.08	1.0000	<0.001
p value[II]		0.28	0.82	0.67				
Hospital admission								
No	43 (27.9)	52.07	0.772	0.688	0.084	<0.001	0.533	<0.001
Yes	111 (72.1)	64.55	0.594	0.605	−0.011	0.32	0.689	<0.001
p value[§]		<0.001	0.002	<0.001				
Comorbidities								
No	50 (32.5)	59.77	0.706	0.650	0.056	0.03	0.745	<0.001
Yes	104 (67.5)	61.69	0.614	0.618	−0.004	0.43	0.642	<0.001
p value[§]		0.27	0.04	0.04				

[#]Significance test by paired-samples t-test.
[§]P value for between-group comparison, significance test by t-test for SF-6D and SGRQ; Mann-Whitney test for EQ-5D.
[II]P value for between-group comparison, significance test by ANOVA for SF-6D and SGRQ; Kruskal-Wallis test for EQ-5D.
Note: COPD = Chronic Obstructive Pulmonary Disease; SGRQ = Saint George's Respiratory Questionnaire.
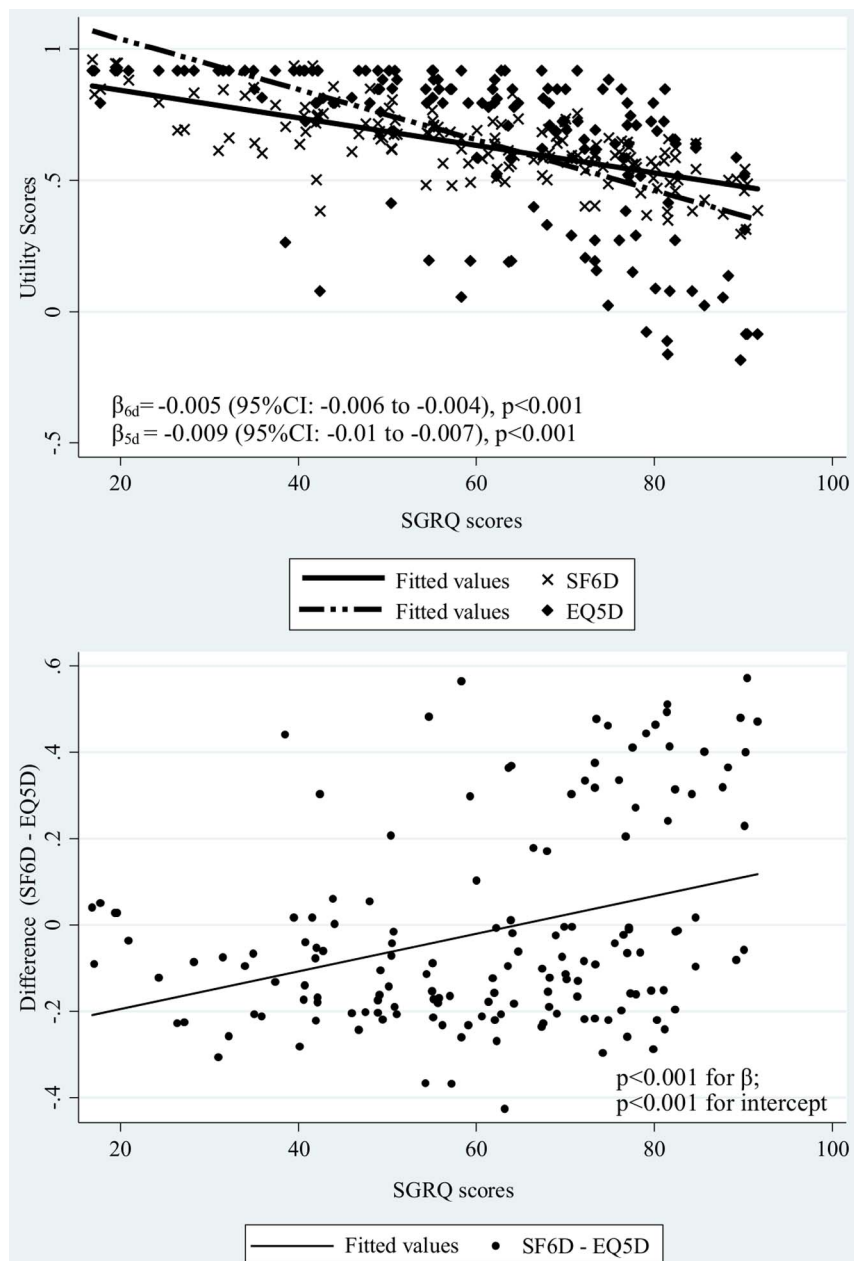doi:10.1371/journal.pone.0112389.t003

β_{6d} = -0.005 (95%CI: -0.006 to -0.004), p<0.001
β_{5d} = -0.009 (95%CI: -0.01 to -0.007), p<0.001

p<0.001 for β;
p<0.001 for intercept

Figure 3. Scatter plots of SF-6D vs. SGRQ, EQ-5D vs. SGRQ (above) and difference (SF-6D – EQ-5D) vs. SGRQ (below).
doi:10.1371/journal.pone.0112389.g003

22.1% of our sample patients scored the highest possible score, may not be as efficient as SF-6D in reflecting the disease severity. Further direct comparison of the EQ-5D and the SF-6D scores against disease severity indicated by SGRQ scores also confirmed this. The relationship observed between the SF-6D and SGRQ differed from that between the EQ-5D and SGRQ, where the SF-6D scored higher in more severe cases and the EQ-5D scored higher in less severe cases. This may be due to valuation techniques, which is TTO for the EQ-5D, as well as scaling properties. Others have found that SG generally produces higher values than TTO for the same condition [35,46] but TTO usually gives higher values than SG for milder health states [47]. This suggests that there might be a "cross-over" point for TTO and SG scores along the axis of illness severity and hence, also, for EQ-5D

and SF-6D values. We did indeed see this crossover point in our study.

Moreover, consistent with other studies [19,21,45], the SF-6D was more efficient in detecting HRQoL differences between subgroups than the EQ-5D with the SF-6D generating larger AUCs and the F-ratio of SF-6D to EQ-5D being greater than 1 in most cases in our study. Specifically, Harper *et al* had already shown that SF-36 was better than EQ5D in detecting minor changes in HRQoL for those with mild COPD in outpatient settings [24] and our study adds evidence that in severe and very severe stages of COPD, the EQ-5D and the SF-6D show different discriminative capacities with SF-6D being more efficient in detecting changes in HRQoL. However, some studies have still preferred to use the EQ-5D for chronic disease in clinic settings and for comparing HRQoL pre and post-intervention [43]. Stel *et*

Table 4. Area under receiver operating characteristic curves (AUC) for EQ-5D and SF-6D with different cut-off points in external indicators.

	SF-6D			EQ-5D			
	AUC	95%CI	F statistics	AUC	95%CI	F statistics	F ratio
EQ-VAS scores							
>=50 vs. <50	0.718	0.631 0.804	18.37	0.724	0.629 0.818	24.13	0.76
>=60 vs. <60	0.672	0.587 0.756	16.21	0.649	0.562 0.735	14.11	1.15
>=70 vs. <70	0.695	0.602 0.788	18.02	0.652	0.559 0.746	8.15	2.21
>=80 vs. <80	0.733	0.610 0.856	13.11	0.687	0.563 0.812	5.32	2.46
>=90 vs. <90	0.763	0.592 0.934	6.41	0.755	0.610 0.899	4.13	1.55
SGRQ scores							
>49 vs. <=49	0.864	0.789 0.938	67.26	0.826	0.749 0.902	23.88	2.82
>64 vs. <=64	0.835	0.769 0.900	66.30	0.850	0.788 0.911	54.97	1.21
>77 vs. <=77	0.867	0.809 0.925	60.02	0.846	0.784 0.908	54.34	1.10

Note: AUC = Area under receiver operating characteristics curves; CI = Confidence interval; SGRQ = Saint George's Respiratory Questionnaire.
doi:10.1371/journal.pone.0112389.t004

al found that the EQ-5D had a substantially lower percentage of missing data than the SF-6D both at baseline and post-intervention assessments when assessing the impact of different surgical interventions on patients with coronary heart disease. Our study did not find such a difference in completeness. Thus along with Harper *et al* [24], the SF-6D would be a more precise measurement in estimate of QALYs in patients with COPD than the EQ-5D. It is evident that the EQ-5D and the SF-6D yield different utility values and measured different things in COPD patients. There two instruments might not be used interchangeably, so a careful choice needs to be made. It is therefore important for clinical and health technology assessment authorities to choose the most appropriate measurement to better inform the healthcare policy using QALYs as outcome. More comparative studies would be useful.

Strengths and limitations

To our knowledge, our study was the first head to head comparison of the SF-6D and the EQ-5D among Chinese subjects with a severe chronic disease. Classifying health status using SGRQ and EQ-VAS scores could capture the underlying disease severity and provide reliable and validated external measures of HRQoL. Existing clinical conditions and related health care utilization data were extracted directly from medical records which minimized errors and bias in recall.

However, there were several potential limitations in our study. Firstly, the UK-derived scoring algorithms were used for comparing the EQ-5D and the SF-6D since there was no local algorithm available for the EQ-5D. This may fail to reflect the true preference in a Chinese population. Secondly, our sample was mainly male (98%) which may limit the external validity of our results. However, evidence on any gender difference in HRQoL is not conclusive with many studies showed non-significant differences [25,27,29,48,49]. A local study did not find any difference in the HRQoL of males and females with COPD who were attending community health centers (unpublished data). Finally, the cross-sectional nature of this study renders it unable to test the ability of the instruments to detect changes in HRQoL.

Conclusions

Both the EQ-5D and the SF-6D appeared to be valid and discriminative preference-based measures of health in Chinese patients with severe and very severe COPD. Our study added evidence that the SF-6D was more efficient in detecting differences among subgroups with differing disease severity of COPD. The EQ-5D and the SF-6D measured different things and might not be used interchangeably in COPD patients. Further research would be needed to determine other psychometric properties of the EQ-5D and the SF-6D such as responsiveness to assess ability to detect longitudinal changes.

Acknowledgments

Author Contributions

Conceived and designed the experiments: JC SMM WCY. Performed the experiments: JC SMM WCY PP. Analyzed the data: JC CW SMM. Wrote the paper: JC CW SMM PP WCY.

References

1. Brazier J (2007) Measuring and valuing health benefits for economic evaluation. Oxford; New York: Oxford University Press. xi, 344 p.
2. Rabin R, de Charro F (2001) EQ-5D: a measure of health status from the EuroQol Group. Ann Med 33: 337–343.
3. Sintonen H (2001) The 15D instrument of health-related quality of life: properties and applications. Ann Med 33: 328–336.
4. Horsman J, Furlong W, Feeny D, Torrance G (2003) The Health Utilities Index (HUI): concepts, measurement properties and applications. Health Qual Life Outcomes 1: 54.
5. Kaplan RM, Anderson JP (1988) A general health policy model: update and applications. Health Serv Res 23: 203–235.
6. Brazier J, Roberts J, Deverill M (2002) The estimation of a preference-based measure of health from the SF-36. J Health Econ 21: 271–292.
7. Rasanen P, Roine E, Sintonen H, Semberg-Konttinen V, Ryynanen OP, et al. (2006) Use of quality-adjusted life years for the estimation of effectiveness of health care: A systematic literature review. Int J Technol Assess Health Care 22: 235–241.
8. Dolan P, Gudex C, Kind P, William A (1995) A social tariff for EuroQol: Results from a UK general population study. Discussion paper 138. York: The University of York.
9. Nord E (1991) EuroQol: health-related quality of life measurement. Valuations of health states by the general public in Norway. Health Policy 18: 25–36.
10. Tsuchiya A, Ikeda S, Ikegami N, Nishimura S, Sakai I, et al. (2002) Estimating an EQ-5D population value set: the case of Japan. Health Econ 11: 341–353.
11. EuroQol Group (2013) About EQ-5D Rotterdam. Available: http://www.euroqol.org/about-eq-5d.html
12. Brazier J, Roberts J, Tsuchiya A, Busschbach J (2004) A comparison of the EQ-5D and SF-6D across seven patient groups. Health Econ 13: 873–884.
13. Cunillera O, Tresserras R, Rajmil L, Vilagut G, Brugulat P, et al. (2010) Discriminative capacity of the EQ-5D, SF-6D, and SF-12 as measures of health status in population health survey. Qual Life Res 19: 853–864.
14. Gerard K, Nicholson T, Mullee M, Mehta R, Roderick P (2004) EQ-5D versus SF-6D in an older, chronically Ill patient group. Appl Health Econ Health Policy 3: 91–102.
15. Kontodimopoulos N, Argiriou M, Theakos N, Niakas D (2011) The impact of disease severity on EQ-5D and SF-6D utility discrepancies in chronic heart failure. Eur J Health Econ 12: 383–391.
16. Kontodimopoulos N, Pappa E, Papadopoulos AA, Tountas Y, Niakas D (2009) Comparing SF-6D and EQ-5D utilities across groups differing in health status. Qual Life Res 18: 87–97.
17. Lamers LM, Bouwmans CA, van Straten A, Donker MC, Hakkaart L (2006) Comparison of EQ-5D and SF-6D utilities in mental health patients. Health Econ 15: 1229–1236.
18. Longworth L, Bryan S (2003) An empirical comparison of EQ-5D and SF-6D in liver transplant patients. Health Econ 12: 1061–1067.
19. Szende A, Leidy NK, Stahl E, Svensson K (2009) Estimating health utilities in patients with asthma and COPD: evidence on the performance of EQ-5D and SF-6D. Qual Life Res 18: 267–272.
20. Walters SJ, Brazier JE (2005) Comparison of the minimally important difference for two health state utility measures: EQ-5D and SF-6D. Qual Life Res 14: 1523–1532.
21. Grieve R, Grishchenko M, Cairns J (2009) SF-6D versus EQ-5D: reasons for differences in utility scores and impact on reported cost-utility. Eur J Health Econ 10: 15–23.
22. Joore M, Brunenberg D, Nelemans P, Wouters E, Kuijpers P, et al. (2010) The impact of differences in EQ-5D and SF-6D utility scores on the acceptability of cost-utility ratios: results across five trial-based cost-utility studies. Value Health 13: 222–229.
23. Scott NW, Fayers PM, Aaronson NK, Bottomley A, de Graeff A, et al. (2009) The practical impact of differential item functioning analyses in a health-related quality of life instrument. Qual Life Res 18: 1125–1130.
24. Harper R, Brazier JE, Waterhouse JC, Walters SJ, Jones NM, et al. (1997) Comparison of outcome measures for patients with chronic obstructive pulmonary disease (COPD) in an outpatient setting. Thorax 52: 879–887.
25. Engstrom CP, Persson LO, Larsson S, Ryden A, Sullivan M (1996) Functional status and well being in chronic obstructive pulmonary disease with regard to clinical parameters and smoking: a descriptive and comparative study. Thorax 51: 825–830.
26. Ferrer M, Alonso J, Morera J, Marrades RM, Khalaf A, et al. (1997) Chronic obstructive pulmonary disease stage and health-related quality of life. The Quality of Life of Chronic Obstructive Pulmonary Disease Study Group. Ann Intern Med 127: 1072–1079.
27. Ketelaars CA, Schlosser MA, Mostert R, Huyer Abu-Saad H, Halfens RJ, et al. (1996) Determinants of health-related quality of life in patients with chronic obstructive pulmonary disease. Thorax 51: 39–43.
28. Lam CL, Lauder IJ (2000) The impact of chronic diseases on the health-related quality of life (HRQOL) of Chinese patients in primary care. Fam Pract 17: 159–166.
29. Stahl E, Lindberg A, Jansson SA, Ronmark E, Svensson K, et al. (2005) Health-related quality of life is related to COPD disease severity. Health Qual Life Outcomes 3: 56.
30. Voll-Aanerud M, Eagan TM, Wentzel-Larsen T, Gulsvik A, Bakke PS (2008) Respiratory symptoms, COPD severity, and health related quality of life in a general population sample. Respir Med 102: 399–406.
31. Buist AS, Vollmer WM, McBurnie MA (2008) Worldwide burden of COPD in high- and low-income countries. Part I. The burden of obstructive lung disease (BOLD) initiative. Int J Tuberc Lung Dis 12: 703–708.
32. World Health Organization Quick COPD facts. Available: http://www.who.int/respiratory/copd/en/.
33. GOLD Executive Committee (2011) Global initiative for Chronic Obstructive Lung Disease: Global strategy for the diagnosis, management, and prevention of Chronic Obstructive Pulmonary Disease. USA.
34. Lam CL, Brazier J, McGhee SM (2008) Valuation of the SF-6D Health States Is Feasible, Acceptable, Reliable, and Valid in a Chinese Population. Value Health 11: 295–303.
35. Dolan P, Gudex C, Kind P, Williams A (1996) Valuing health states: a comparison of methods. J Health Econ 15: 209–231.
36. McGhee SM, Brazier J, Lam CL, Wong LC, Chau J, et al. (2011) Quality adjusted life years: Population specific measurement of the quality component. Hong Kong Med J 17: 17–21.
37. Jones PW, Quirk FH, Baveystock CM, Littlejohns P (1992) A self-complete measure of health status for chronic airflow limitation. The St. George's Respiratory Questionnaire. Am Rev Respir Dis 145: 1321–1327.
38. Yu TW, Scudds RJ, Scudds RA (2004) Reliability and Validity of a Hong Kong Chinese Version of the St George's Respiratory Questionnaire in Patients with COPD. 22: 33–39.
39. Quirk FH, Baveystock CM, Wilson R, Jones PW (1991) Influence of demographic and disease related factors on the degree of distress associated with symptoms and restrictions on daily living due to asthma in six countries. Eur Respir J 4: 167–171.
40. Jones PW (1994) Quality of life, symptoms and pulmonary function in asthma: long-term treatment with nedocromil sodium examined in a controlled multicentre trial. Nedocromil Sodium Quality of Life Study Group. Eur Respir J 7: 55–62.
41. Sullivan PW, Lawrence WF, Ghushchyan V (2005) A national catalog of preference-based scores for chronic conditions in the United States. Med Care 43: 736–749.
42. Walters SJ, Brazier JE (2003) 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.
43. van Stel HF, Buskens E (2006) Comparison of the SF-6D and the EQ-5D in patients with coronary heart disease. Health Qual Life Outcomes 4: 20.
44. McCrone P, Patel A, Knapp M, Schene A, Koeter M, et al. (2009) A comparison of SF-6D and EQ-5D utility scores in a study of patients with schizophrenia. J Ment Health Policy Econ 12: 27–31.
45. Zhao FL, Yue M, Yang H, Wang T, Wu JH, et al. (2010) Validation and comparison of EuroQol and short form 6D in chronic prostatitis patients. Value Health 13: 649–656.
46. Green C, Brazier J, Deverill M (2000) Valuing health-related quality of life. A review of health state valuation techniques. Pharmacoeconomics 17: 151–165.
47. Dolan P, Sutton M (1997) Mapping visual analogue scale health state valuations onto standard gamble and time trade-off values. Soc Sci Med 44: 1519–1530.
48. Renwick DS, Connolly MJ (1996) Impact of obstructive airways disease on quality of life in older adults. Thorax 51: 520–525.
49. Shoup R, Dalsky G, Warner S, Davies M, Connors M, et al. (1997) Body composition and health-related quality of life in patients with obstructive airways disease. Eur Respir J 10: 1576–1580.