A comparison of the reliability and validity of SF-6D, EQ-5D and HUI3 utility measures in patients with schizophrenia and patients with depression in Singapore

Edimansyah Abdin[a,*], Siow Ann Chong[a], Esmond Seow[a], Chao Xu Peh[a], Jit Hui Tan[a], Jianlin Liu[a], Sophia Foo Si Hui[a], Boon Yiang Chua[a], Kang Sim[a], Swapna Verma[b], Janhavi Ajit Vaingankar[a], Mythily Subramaniam[a]

[a] Research Division, Institute of Mental Health, Buangkok Green Medical Park, 10 Buangkok View, 539747, Singapore
[b] Early Psychosis Intervention Programme, Institute of Mental Health, Singapore

A B S T R A C T

There is limited evidence of a direct comparison of the psychometric performance of generic preference-based measures in patients with mental illness in an Asian patient population. The current study aimed to compare the test–retest reliability, convergent and known-group validity and magnitude of change in scores of the EuroQol Five-Dimension, Health Utility Index Mark 3 (HUI3) and Short-Form Six-Dimension (SF-6D) measures in patients with depression and patients with schizophrenia spectrum disorder. 500 patients were recruited from a tertiary psychiatric institution in Singapore. The Schizophrenia Quality of Life Scale (SQLS), 8-item Patient Health Questionnaire, Quality of Life Enjoyment and Satisfaction Questionnaire (Q-LES-Q), and Positive and Negative Syndrome Scale were also included. In the schizophrenia sample, the SF-6D was found to have higher test–retest validity, convergent validity with SQLS domain scores, known-group validity and magnitude of change in scores over 6-month follow up than other measures. In the depression sample, the HUI3 was found to have higher test–retest reliability, convergent validity with Q-LES-Q, known group validity and magnitude of change in scores than other measures. Results suggest that the SF-6D and HUI3 to be more suitable as a utility measure for patients with schizophrenia and depression in an Asian patient population.

1. Introduction

The EuroQol Five-Dimension (EQ-5D) (EuroQol Group, 1990), the Short-Form Six-Dimension (SF-6D) (Brazier et al., 2002) and the Health Utilities Index Mark 3 (HUI3) (Feeny et al., 1995) are the three most widely used generic preference-based instruments to assess patients' health-related quality of life in clinical studies as well as to calculate quality-adjusted life years (QALYs) for cost-utility analyses (CUA) in economic evaluation of mental health treatments and programmes. With the significant increase in the use of these instruments in the economic evaluation of mental health interventions, there is growing interest to conduct a direct comparison of the psychometric performance of these instruments in psychiatric patients.

An earlier review on validity of the EQ-5D and SF-6D among patients with schizophrenia concluded that both instruments had good construct and known-group validity. However, the evidence for convergent validity and ability of the instruments to detect change over time as assessed by the magnitude of change in scores was mixed (Papaioannou et al., 2011). A review by Peasgood et al. (2012) found that both the EQ-5D and SF-6D demonstrated good construct validity

and magnitude of change in scores in patients with depression but demonstrated poor construct validity in patients with anxiety disorders (Peasgood et al., 2012). For people with psychosis, studies have reported that the evidence of validity of the EQ-5D was mixed (Brazier, 2010). A recent review by Brazier et al. (2014) concluded that the EQ-5D and SF-6D instruments appear to perform acceptably well in depression and personality disorder, but less so in anxiety, schizophrenia and bipolar disorder (Brazier et al., 2014). As a follow-up to these reviews, Mulhern et al. (2014) and Brazier et al. (2014) have further investigated the psychometric performance of the EQ-5D and SF-6D using seven existing datasets including people with depression, anxiety, mixed diagnosis of common disorders, personality disorders and schizophrenia and concluded that the EQ-5D and SF-6D were valid in samples with mild to moderate depression and anxiety. For schizophrenia, the findings were less clear, with small magnitude of change observed in scores of the two measures in comparison with the condition-specific measures (Brazier et al., 2014; Mulhern et al., 2014). The inconsistency of these findings in mental health population suggests that further work is required to establish the validity of these instruments in mental health populations locally.

---

* Corresponding author.
  E-mail address: Edimansyah_Abdin@imh.com.sg (E. Abdin).

Singapore is an island nation with a total population of 5.61 million. The Chinese (74.3%) form the majority of the population, followed by Malays (13.4%), Indians (9.0%) and those from other ethnic groups (3.2%). In Singapore, the EQ-5D 3 levels (EQ-5D-3L), SF-6D and HUI3 have been validated individually and mainly in the general population (Thumboo et al., 2002; Wee et al., 2004; Luo et al., 2007; Wee et al., 2007a, 2007b; Thumboo et al., 2013) and among those with chronic physical conditions (Luo et al., 2003; Xie et al., 2007; Gao et al., 2009; Luo et al., 2009a; Leung et al., 2013). However, there is limited evidence of a direct comparison of the reliability, validity and magnitude of change in scores of these measures among local patients with mental illnesses. The HUI3 has been validated in patient with schizophrenia (Luo et al., 2006). The ability of the EQ-5D-3L to detect improvement among patients who met criteria for remission over 1-year follow-up was reported in patients with first-episode psychosis (Subramaniam et al., 2014). However, there is limited evidence on the validity and responsiveness of the HUI3 and both EQ-5D-3L and 5 levels (EQ-5D-5L) in patients with depression and SF-6D in patients with schizophrenia. There are also other gaps in our knowledge about the psychometric performance of the utility scores derived from these instruments locally among patients with mental illnesses. First, although there were some differences between the psychometric performance of EQ-5D-3L, SF-6D and HUI3 reported previously, this evidence was limited and the studies mainly carried out in Western population using different samples (Brazier et al., 2014; Mulhern et al., 2014). It is not clear how comparable these measures are as compared to disease specific preference measures in such patients or whether one instrument has better psychometric properties than the other as head-to-head comparisons between these measures are lacking (Finch et al., 2018). Second, the magnitude of change in scores in these measures over time was not compared directly using the same sample and investigated thoroughly in the local context. Hence, the validation of these measures is vital. The current study aimed to compare the test–retest reliability, convergent validity, known-group validity and magnitude of change in scores of the four utility measures (EQ-5D-5L, EQ-5D-3L, HUI3 and SF-6D) in patients with depression and patients with schizophrenia spectrum disorders in Singapore.

2. Methods

A cross-sectional study was conducted on a convenient sample at the Institute of Mental Health (IMH) in Singapore. IMH is the national tertiary psychiatric care provider which serves a large number of patients with diverse mental disorders. Thus, its patient population lends itself well to the validation of these instruments. Participants were patients recruited from outpatient clinics at the IMH, Singapore citizens or permanent residents, aged 21 years and above, literate in English and having a clinical diagnosis of either depressive disorder or schizophrenia spectrum disorders. Written informed consent was obtained from all study participants. All participants were asked to complete a set of self-administered questionnaires that were designed to collect their sociodemographic, disease-specific and generic-preference QOL information. The severity of schizophrenia symptoms was assessed by a trained interviewer using the Positive and Negative Syndrome Scale (PANSS). Data on socio-demographic background, four utility measures, Schizophrenia Quality of Life Scale (SQLS), 8-item Patient Health Questionnaire (PHQ-8), Quality of Life Enjoyment and Satisfaction Questionnaire (Q-LES-Q), and Positive and Negative Syndrome Scale (PANSS) were collected from 500 participants (depression (n = 249) and schizophrenia spectrum disorders (n = 251)) at the baseline assessment. Data for four utility measures were also collected from 60 participants (30 from each diagnostic group) at 2 weeks follow-up for test–retest reliability and 180 participants (90 from each diagnostic group) at 6 months follow-up for magnitude of change analyses. We used the mean of EQ-5D utility scores at baseline (0.799) and one year follow-up (0.891) in patients with first episode psychosis to estimate

sample size for magnitude of change analyses (Subramaniam et al., 2014). The power and significance level was set at 0.9 and 0.05, respectively. We found that we need at least a sample size of 120 to be able to reject the null hypothesis that the magnitude of change in EQ-5D scores is small over the 6 months follow-up. After considering approximately 50% attrition rate, a sample size of 180 was required for the study. In terms of test–retest reliability analysis, we estimated that a sample size of 39 subjects was required to determine the intra-class correlation coefficient of 0.67 for SF-36 (Koh et al., 2006) with a power of 0.8 at a significance level of 0.05. After considering 50% attrition rates, a sample size of 60 was determined to be adequate for test retest analysis.

2.1. Instruments

2.1.1. Socio-demographic variables

These included information on participants' age, gender, education, ethnicity, marital and employment status, housing type, and psychiatric diagnosis.

2.1.2. EQ-5D

The EQ-5D is a generic instrument for subjectively describing and valuing HRQoL that has been developed by the EuroQol Group (1990) and comprises of two versions. The EQ-5D-3L comprised five questions on mobility, self-care, pain, usual activities, and psychological status with three possible answers for each item (1 = no problem, 2 = moderate problem, 3 = severe problem) whilst the EQ-5D-5L has five possible answers for each item (1 = no problem, 2 = slight problems, 3 = moderate problems, 4 = severe problems, 5 = extreme problems). The utility scores of EQ-5D-3L and EQ-5D-5L were calculated using the scoring algorithms developed from the Singapore (Luo et al., 2014) and UK general populations (van Hout et al., 2012).

2.1.3. SF-6D

The SF-6D, derived from 11 items of the SF-36, is a multi-dimensional health classification system assessing the 6 health domains of physical functioning, role limitation, social functioning, pain, mental health, and vitality, with 4–6 levels for each domain. An SF-6D health state is defined by selecting 1 level from each domain, which results in a total of 18,000 health states. The SF-6D scoring algorithm was developed using the standard gamble (SG) method from a sample of 249 SF-6D health states from a representative sample of the UK population (Brazier et al., 2002). Utility scores generated by the SF-6D range from 0.29 to 1, with 1 representing full health and 0.29 representing the worst possible health state defined by the SF-6D (i.e., all domains being at the worst level). The utility scores derived from English and Chinese versions of the SF-6D have been demonstrated to be equivalent in Singapore (Wee et al., 2004).

2.1.4. HUI3

The HUI3 (Feeny et al., 1995) is a generic comprehensive health status classification and HRQoL utility scoring system, using a utility scoring function derived from a representative sample of the Canadian general population based on the SG and visual analog scale methods (Horsman et al., 2003). The utility score for HUI3 range from −0.36 to 1. The HUI3 classification system has 8 domains: vision, hearing, speech, ambulation, dexterity, emotion, cognition, and pain, with 5 to 6 levels per attribute derived from 15-multiple choice questions. The utility scores derived from Chinese and Malay versions of the HUI3 have been demonstrated to be equivalent in Singapore (Luo et al., 2007).

2.1.5. SQLS

The SQLS is a disease-specific instrument (Wilkinson et al., 2000), designed to assess the impact of schizophrenia and treatment regime on the daily life of patients. The SQLS has three sub scores (Psychosocial,

Motivation and energy and Symptoms and side-effects). Each sub score is transformed to a scaled score of 0 (the best quality of life status) to 100 (the worst quality of life status) (Wilkinson et al., 2000). The SQLS has been validated in Singaporean patients with schizophrenia (Luo et al., 2008).

2.1.6. PHQ-8

The PHQ-8 is a self-administered depression scale where participants indicate how often they have been bothered by each of the items (symptom), in the past two weeks using a 4-point scale, where 0 = not at all and 3 = nearly every day. Total scores range from 0 to 24, where scores of 10 and above indicate current depression (Kroenke et al., 2009). In the current study, patients with depression were divided into two subgroups at baseline using PHQ-8 cut-off points (where a score of ≥ 10 indicates clinical depression) (Kroenke et al., 2009).

2.1.7. Q-LES-Q

The Q-LES-Q is a disease-specific measure for assessing the impact that depression has on a patient's quality of life (Tuynman-Qua et al., 1997). The Q-LES-Q is a 16-item scale that assesses enjoyment and satisfaction across various life domains such as physical health, social relationships, and economic status over the past week. Each item is rated on a scale of 1 to 5 ("not at all or never" to "frequently or all the time"). The scoring of the Q-LES-Q involves summing the 14 items, with higher scores indicating a higher subjective QoL. The reliability of the scale was excellent in local sample (Ong et al., 2016).

2.1.8. PANSS

The PANSS (Kay et al., 1987) is a 30-item instrument designed to measure the severity of symptoms among those with schizophrenia spectrum disorders. The symptoms severity was assessed by a trained interviewer following a semi-structured interview with the participant. Each symptom is rated on a seven points scale represent increasing levels of psychopathology (1 = absent to 7 = extreme) with total scores ranging from 30 to 210. The construct validity of this scale has been validated in Singapore (Jiang et al., 2013). In the current study, patients with schizophrenia were divided into two subgroups at baseline using PANSS total cut-off scores – those who had PANSS total score ≤ 75 and PANSS total score ≥ 76 to denote moderately severe symptoms (Furukawa et al., 2015).

2.2. Statistical analyses

Ceiling and floor effects were calculated as the percentage of participants who achieved the maximum and minimum scores of each utility score. We adopted the standard approach for assessing the psychometric properties of the measures. This includes test–retest reliability, convergent validity, known-group validity and evaluation of magnitude of change in scores over the 6-month follow-up. Test–retest reliability was used to examine whether the utility scores would yield reproducible results if used repeatedly on stable patients within a two-week period of time. Test–retest reliability was assessed using intraclass correlation coefficient (ICC) based on a two-way mixed model with absolute agreement. ICC was interpreted using following criteria (Koo and Li, 2016): > 0.9, excellent reliability, 0.75 to 0.9, good, 0.5 to 0.75, moderate, and < 0.5, poor reliability. Validity of the utility scores was assessed using two validity tests including convergent validity and known-group validity. The convergence between the scores from three generic preference-based measures (EQ-5D, SF-6D and HUI3) and two disease specific preference measures (SQLS for schizophrenia and Q-LES-Q for depression) were examined using Spearman's correlation coefficients due to the skewed distribution of the data. We used the following categories for evidence of convergent validity: > 0.6, very strong; ≥ 0.5 to < 0.6, strong; < 0.5 to ≥ 0.3, moderate; and < 0.3, weak (Papaioannou et al., 2011). Locally weighted smoothing curve (LOWESS) fit lines were used to indicate nonlinear trends in the scatter

plots (Cleveland, 1979). Known-group validity was used to assess whether the four utility scores were able to discriminate between those people with mild and severe conditions schizophrenia and depression. Clinical importance of any difference was quantified using the effect size (ES) calculated as the differences in mean scores between groups divided by the standard deviation of the baseline scores. Magnitude of change in scores was defined as the extent to which an instrument was able to detect a clinically or practically important change in quality of life status (Walters, 2009). Magnitude of change in scores was quantified using mean change scores and standardized response mean (SRM) (the mean change in scores between baseline and 6 months follow-up divided by the standard deviation of the change in the utility scores) (Brazier et al., 2014). Effect sizes of 0.2 were defined using Cohen's criteria as small, 0.5 as moderate, and 0.8 as large. Correlation between changes in the four utility scores and changes in the Mental Component Summary (MCS) score and the Physical Component Summary (PCS) of the SF-36 questionnaire (Farivar et al., 2007) was assessed using Spearman's correlation coefficients. Minimal clinically important difference (MCID) defined as the smallest difference in score that patients perceive as important, either beneficial or harmful, and which lead the clinician to consider a change in the patient's management (Guyatt et al., 2002) was also examined. Wiebe et al. (2002) suggested that a 4.6 and 3.0 point changes scores can be used to indicate MCID scores for the MCS and the PCS, respectively (Wiebe et al., 2002). Hence, we have used these MCID values to classify patients as 'improved' (achieved positive MCID), 'worsened' (achieved negative MCID) or as 'no change' (not achieved either positive or negative MCID) (Fiest et al., 2014) and subsequently compared the mean change in the four utility scores over 6 months follow-up. According to previous studies, empirical distribution based estimates of MCID for utility scores can range between 0.01 and 0.08 with 0.03 being well represented in estimates for the generic preference-based measures (Feeny et al., 2012; Dickerson et al., 2018). Following previous work, the mean difference in utility scores between baseline and 6 months follow-up of 0.03 or more was considered to have achieved MCID. All statistically significant differences were evaluated at the 0.05 level using 2-sided tests.

3. Results

3.1. Descriptive statistics

The demographic characteristics of the participants are presented in Table 1. The sample included 500 participants with depression (n = 249) and schizophrenia spectrum disorders (n = 251). The proportion of depression among patients with schizophrenia spectrum disorders was 3.6%. The mean age of overall sample was 38.1 years (SD = 10.8), 64.4% were Chinese, 16.8% were Malays, 16.6% were Indians, and 2.2% belonged to other ethnicities (Table 1). The mean EQ-5D-5L, EQ-5D-3L, HUI3 and SF-6D utility scores were 0.76, 0.69, 0.55 and 0.65, respectively (Table 2). An inspection of the distribution

Table. 1
Characteristics of the sample.

	Overall (n = 500)	Schizophrenia spectrum disorder (n = 251)	Depression (n = 249)
	Mean(SD)	Mean(SD)	Mean(SD)
Age	38.06(10.81)	39.88(10.24)	36.23(11.07)
	N(%)	N	N
Gender			
Female	228(45.6)	110(43.8)	118(47.4)
Male	272(54.4)	141(56.2)	131(52.6)
Ethnicity			
Chinese	322(64.4)	148(59.0)	174(69.9)
Malay	84(16.8)	50(19.9)	34(13.6)
Indian	83(16.6)	46(18.3)	37(14.9)
Others	11(2.2)	7(2.8)	4(1.6)

Table. 2
Summary statistics and test–retest reliability of the utility scores in schizophrenia spectrum disorder and depression samples.

	Overall	Schizophrenia spectrum disorder	Depression
Utility scores			
EQ-5D-5L			
Mean (SD)	0.76(0.22)	0.81(0.22)	0.71(0.21)
Median	0.81	0.85	0.75
Ceiling effect (%)	22.6	36.7	8.4
ICC	0.71	0.62	0.74
EQ-5D-3L			
Mean (SD)	0.69(0.32)	0.77(0.31)	0.61(0.32)
Median	0.85	0.85	0.70
Ceiling effect (%)	28.4	43.0	13.7
ICC	0.63	0.35	0.62
HUI3			
Mean (SD)	0.55(0.33)	0.65(0.31)	0.46(0.33)
Median	0.64	0.74	0.48
Ceiling effect (%)	3.6	6.4	0.8
ICC	0.69	0.52	0.83
SF-6D			
Mean (SD)	0.65(0.14)	0.68(0.15)	0.61(0.12)
Median	0.62	0.66	0.60
Ceiling effect (%)	1.4	2.8	0.4
ICC	0.83	0.92	0.61

Table. 3
Convergent validity of the EQ-5D, HUI3 and SF-6D utility scores in schizophrenia spectrum disorder and depression samples.

Disorder-specific measure	EQ-5D-5L	EQ-5D-3L	HUI3	SF-6D
Schizophrenia spectrum disorder sample				
SQLS				
Psychosocial	−0.58*	−0.64*	−0.59*	−0.68*
Motivation and energy	−0.45*	−0.47*	−0.57*	−0.60*
Symptoms and side-effects	−0.59*	−0.60*	−0.54*	−0.66*
Depression sample				
Q-LES-Q total	0.65*	0.67*	0.79*	0.74*

* P value < 0.001.

6D utility scores were lowest (Median = 0.66) in the schizophrenia sample whereas the median HUI3 utility scores were lowest (Median = 0.48) in the depression sample (Table 2).

3.2. Test–retest reliability

The test–retest reliability coefficients for all utility scores ranged from poor (ICC = 0.35) to excellent reliability (ICC = 0.92) (Table 2). In the schizophrenia spectrum disorder sample, the highest test–retest reliability was observed in SF-6D with excellent reliability while the highest test retest reliability in depression was observed in HUI3 with good reliability.
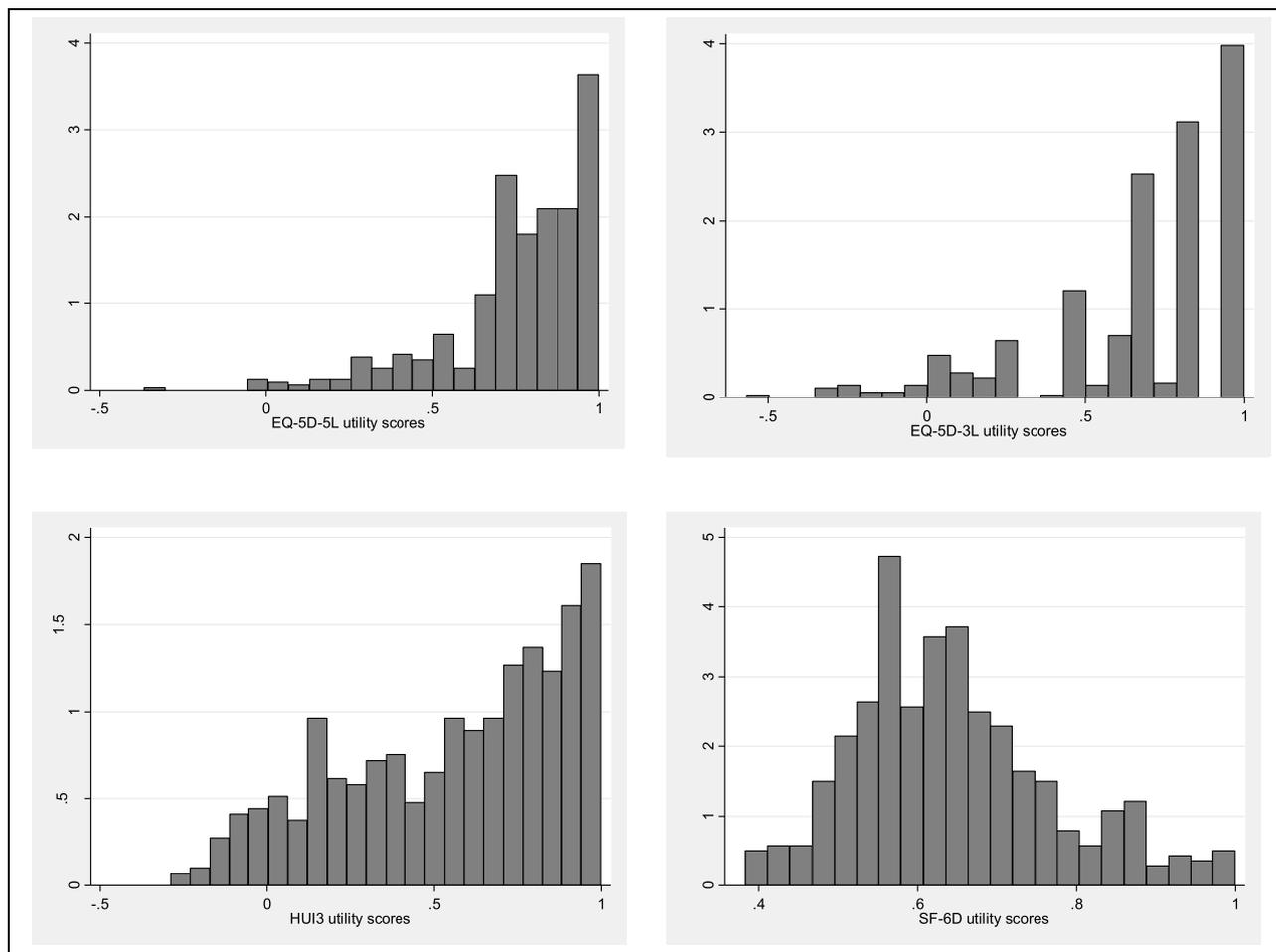
of the scores revealed that the EQ-5D-3L, EQ-5D-5L and HUI3 scores, in general, showed a skewed distribution whereas the SF-6D scores were approximately normally distributed (Fig. 1). The mean HUI3 utility scores were lowest in both schizophrenia spectrum disorders (Mean = 0.65) and depression (Mean = 0.46) samples. The median SF-

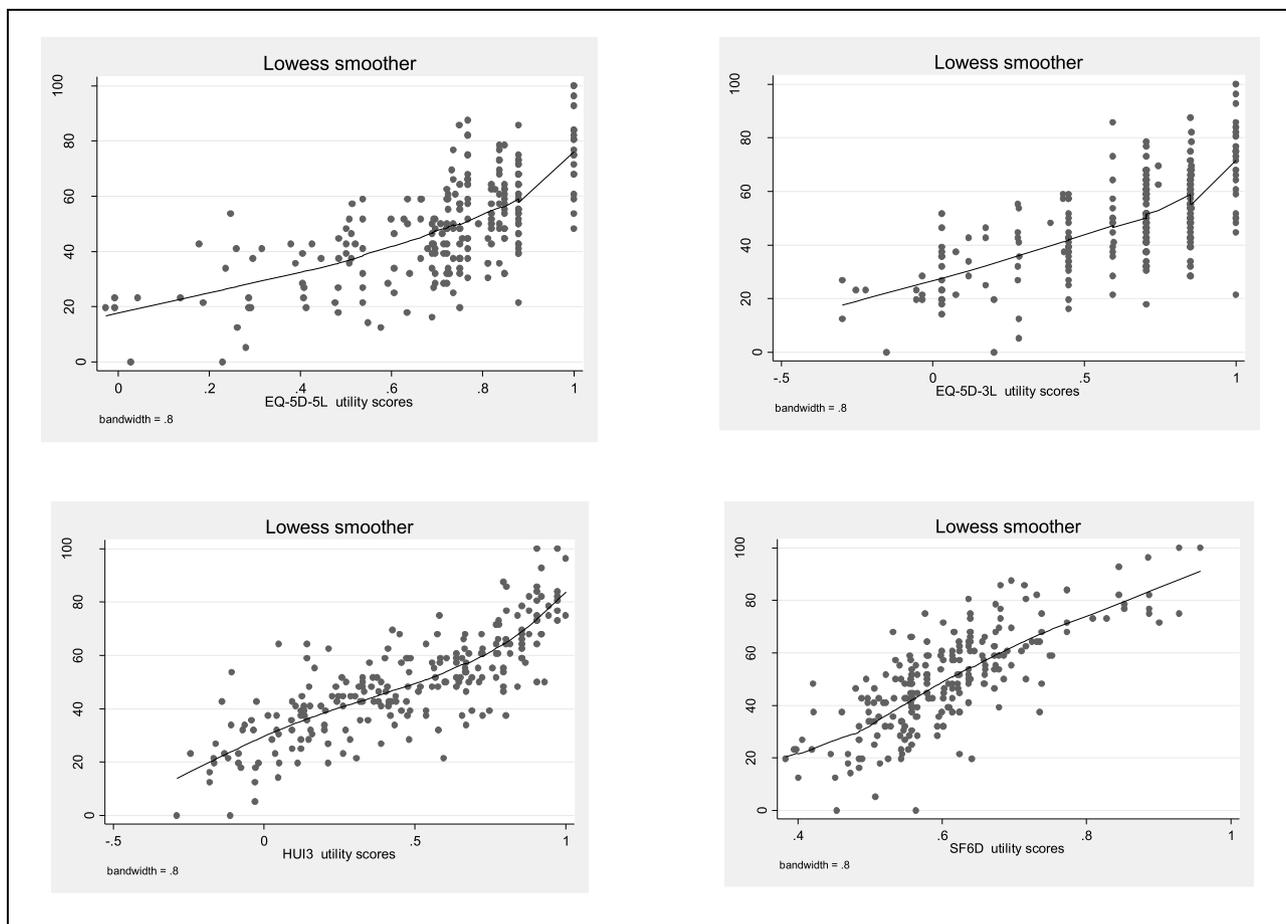**Fig. 1.** EQ-5D-3L, EQ-5D-5L, HUI3 and SF-6D distribution in overall sample.

Fig. 2. Scatterplots and LOWESS lines for four utility scores and Q-LES-Q in depression sample.

3.3. Convergent validity

Table 3 showed the convergent validity of the four utility scores in schizophrenia spectrum disorder and depression samples. In the schizophrenia spectrum disorder sample, the correlation coefficients between the four utility scores and SQLS's subdomain scores ranged from moderate (0.45) to very strong (0.68) convergence. The SF-6D utility scores had the highest correlation coefficient value (SF-6D with SQLS Psychosocial = 0.68) than other utility scores. In the depression sample, the correlation coefficient values between the utility scores and Q-LES-Q total scores ranged within very strong (0.65 to 0.79) level of convergence. The HUI3 utility scores had the highest correlation coefficient value as compared to the other utility scores. Overall, the LOWESS fitting lines indicated that the relationships between four utility scores and two disease specific preference measures were different across their severity scale. For example, Fig. 2 shows that the relationship between the SF-6D and Q-LES-Q is more consistent across the severity scale, with a linear relationship between the scores. However, a score of 1 on the EQ-5D-3L is associated with a wide range of Q-LES-Q scores, meaning that EQ-5D-3L displays a slightly larger ceiling effect than EQ-5D-5L.

3.4. Known-group validity

In the schizophrenia spectrum disorder sample, those with a mild condition (PANSS total score ≤75) were found to have higher utility scores than those with a moderate condition (PANSS total score ≥76). The effect size of mean differences between the two groups according to Cohen's criteria ranged from moderate (1.24) to large (1.54). SF-6D utility scores had the highest effect size value as compared to the other

utility scores. In the depression sample, those with a mild condition (PHQ-8 total score ≤9) were found to have higher utility scores than those with a more severe condition (PHQ total score ≥10). The HUI3 had the highest effect size value as compared to other utility scores according to Cohen's criteria (Table 4).

3.5. Magnitude of change in scores

The mean change for the four utility scores from baseline to 6 months follow-up in schizophrenia spectrum disorder sample is presented in Table 5. In the schizophrenia spectrum disorder sample, except for HUI3, all utility scores tended to have higher mean scores after 6 months follow-up with a small effect size. The SF-6D seems to have slightly higher SRM than other utility scores. However, none of the mean difference showed clinically meaningful improvement. In the depression sample, similarly all utility scores tended to have higher mean scores after 6 months follow-up with a small effect size according to Cohen's criteria. The EQ-5D-5L seems to have slightly higher SRM than other utility scores. All utility scores achieved clinically meaningful improvements in the overall sample. The mean change in the HUI3 utility scores had the highest MCID (Mean Difference = 0.09) as compared to the other utility scores. Further subgroup analysis indicates that the HUI3 seems to have higher SRM among patients with severe depression than other utility scores. Table 6 summarizes correlation and magnitude of change in the EQ-5D-5L, EQ-5D-3L, HUI3, SF-6D utility scores and in the MCS and PCS scores. Spearman correlation coefficients of changes scores were strong between MCS and SF-6D in the schizophrenia spectrum disorder sample and between MCS and SF-6D and HUI3 in the depression sample. The SF-6D had the highest mean change scores among schizophrenia spectrum disorder patients while

Table. 4

Known group validity of the EQ-5D, HUI3 and SF-6D utility scores in schizophrenia spectrum disorder and depression samples.

			EQ-5D-5 index		EQ-5D-3 index		HUI3 index		SF-6D index	
Schizophrenia spectrum disorder										
PANSS										
≤75	(n = 238)	Mean(SD)	0.84	(0.18)	0.80	(0.26)	0.69	(0.30)	0.69	(0.15)
≥76	(n = 13)	Mean(SD)	0.38	(0.37)	0.19	(0.47)	0.25	(0.29)	0.54	(0.10)
		Mean difference	0.46		0.61		0.42		0.15	
		Effect size	1.24		1.31		1.42		1.54	
		P value	< 0.001		< 0.001		< 0.001		< 0.001	
Depression										
PHQ-8										
≤9	(n = 97)	Mean(SD)	0.83	(0.13)	0.81	(0.20)	0.71	(0.22)	0.67	(0.11)
≥10	(n = 151)	Mean(SD)	0.63	(0.22)	0.49	(0.32)	0.30	(0.29)	0.56	(0.08)
		Mean difference	0.20		0.31		0.42		0.11	
		Effect size	0.92		0.98		1.45		1.43	
		P value	< 0.001		< 0.001		< 0.001		< 0.001	

Table. 5

EQ-5D, HUI3 and SF-6D utility scores at baseline and after 6 months follow-up.

Schizophrenia spectrum disorder (n = 90)	EQ-5D-5L Mean	SD	EQ-5D-3L Mean	SD	HUI3 Mean	SD	SF-6D Mean	SD
Baseline	0.84	0.19	0.76	0.33	0.68	0.28	0.69	0.15
6 months	0.84	0.16	0.78	0.29	0.67	0.30	0.71	0.15
Difference	0.005	0.14	0.02	0.27	−0.01	0.25	0.01	0.11
	SRM		SRM		SRM		SRM	
Overall (n = 90)	0.03		0.07		−0.05		0.13	
PANSS ≤ 75 (n = 85)	−0.01		0.08		−0.04		0.10	
PANSS > 75 (n = 5)	0.38		−0.01		−0.23		0.62	
Depression (n = 90)	EQ-5D-5L Mean	SD	EQ-5D-3L Mean	SD	HUI3 Mean	SD	SF-6D Mean	SD
Baseline	0.67	0.23	0.57	0.35	0.40	0.33	0.59	0.12
6 months	0.74	0.21	0.62	0.31	0.48	0.31	0.62	0.12
Difference	0.07	0.22	0.05	0.37	0.09	0.31	0.03	0.12
	SRM		SRM		SRM		SRM	
Overall (n = 90)	0.32*		0.15		0.27*		0.28*	
≤9 (n = 35)	−0.10		−0.27		−0.34		−0.01	
≥10 (n = 55)	0.54*		0.28*		0.69*		0.53*	

* *P* value < 0.05, SRM: Standardised Response Mean.

the HUI3 had the highest mean change scores among depression patients who have achieved improvement in the MCS.

4. Discussions

This is one of few studies that has conducted a direct comparison of the reliability, validity, and magnitude of change in the four utility

scores in patients with schizophrenia spectrum disorders and patients with depression in a multiethnic Asian population.

The findings provide evidence that the SF-6D performs better as compared to other utility measures in terms of test–retest reliability, convergent validity, known-group validity and magnitude of change in scores over the 6-month follow-up in the schizophrenia sample. The analysis confirmed that the SF-6D test–retest reliability was high (ICC = 0.92), implying its ability to produce reproducible results from patients if the instrument is used repeatedly within a short period of time. This finding appears to be consistent with a previous study, which also reported that the test–retest reliability for the SF-6D demonstrated a higher ICC rate (0.82) among patients with systematic sclerosis (Khanna et al., 2007).

In terms of validity, the SF-6D was significantly and negatively correlated with the SQLS domain scores indicating that increases in SF-6D utility scores are associated with decreases in SQLS scores where lower scores indicate a higher subjective QoL. Previous evidence on convergent validity with SQLS is limited. There is one study that has explored the correlation between the SQLS and SF-36 which suggested that both the measures are addressing related, but not identical areas (Wilkinson et al., 2000). The results of known group validity support evidence of discriminant ability of the SF-6D utility scores to differentiate people with mild and moderately ill conditions in schizophrenia sample. A previous review paper had showed mixed evidence on the ability of the GPBMs such as EQ-5D and SF-6D to reflect schizophrenia-specific symptoms (Brazier et al., 2014). They found that the SF-6D scores discriminated people with different severity groups based on Brief Psychiatric Rating Scale Expanded (BPRS-E) cut-off developed by Leucht et al. (2005) better than that the EQ-5D (Brazier et al., 2014).

Our findings also consistently showed that the SF-6D provides a more sensitive measure than other utility scores to detect change before and after 6 months follow-up. However, the effect size of change in SF-

Table. 6

Correlation and magnitude of change in the EQ-5D-5L, EQ-5D-3L, HUI3, SF-6D utility scores and in the MCS and PCS scores.

	Mental component summary score (MCS)				Physical component summary score (PCS)			
	Correlation coefficient	Mean change scores			Correlation coefficient	Mean change scores		
		Improved	No change	Worsen		Improved	No change	Worsen
Schizophrenia spectrum disorder (N,%)		(27, 30%)	(41, 45.6%)	(22, 24.4%)		(28, 31.1%)	(28, 31.1%)	(34, 37.8%)
EQ-5D-5L	0.24*	0.03	0.01	−0.03	−0.14	−0.01	−0.01	0.02
EQ-5D-3L	0.24*	0.07	0.05	−0.10	0.14	0.07	0.06	−0.06
HUI3	0.32*	0.05	0.01	−0.15	0.25*	0.01	0.05	−0.08
SF-6D	0.54*	0.08	0.02	−0.07	0.42*	0.06	0.03	−0.04
Depression (N, %)		(38, 42.2%)	(36, 40%)	(16, 17.8%)		(36, 40%)	(31, 34.4%)	(23, 25.6%)
EQ-5D-5L	0.40*	0.17	0.03	−0.08	0.41*	0.14	0.04	0.00
EQ-5D-3L	0.35*	0.18	0.04	−0.20	0.43*	0.23	−0.04	−0.10
HUI3	0.59*	0.26	0.04	−0.22	0.08	0.08	0.10	0.08
SF-6D	0.69*	0.10	0.00	−0.07	0.39*	0.07	0.03	−0.03

* *P* value < 0.05.

6D utility scores at 6 months follow-up was small according to Cohen's effect size criteria. In the current study, the mean (SD) PANSS total scores in schizophrenia spectrum disorder sample were much lower (47.8(5.4) vs. 75.5(19.1), Mean difference = 27.7) than that reported in first episode psychosis patients within 3-month of entering the treatment programme (Subramaniam et al., 2014). It is thus possible that the small difference in the utility scores between the initial visit and after 6 months of follow up may be simply due to the fact that the patients were stable patients at the point of recruitment. Furukawa et al. (2015) suggested that the more severe the illness at baseline, the larger would be the benefits of the treatment received by the patients. They found that the effect size for patients who were mildly ill (baseline PANSS scores up to 75), moderately ill (baseline PANSS scores of 76 to 95), and severely ill (greater than 95) at baseline was 0.2, 0.5, and 0.8 at 6 weeks follow up which roughly correspond with small, medium, and large effects, respectively. Our subgroup analyses using Furukawa et al. (2015) cutoff also found that the effect size between baseline and follow up visit increased with the severity of the illness in the schizophrenia spectrum disorder sample. For example the effect size of change in SF-6D utility scores at 6 months follow-up for patients who were mildly ill and moderately ill at baseline was 0.10 and 0.62, respectively, and the effect size in moderately ill group was much higher than the effect size in overall sample (0.62 vs. 0.13). However, we would like to highlight that we did not collect sufficient data to enable all the severity groups to be fully assessed in the current study as none of the patients were assessed to be severely ill according to a PANSS cut-off score of greater than 75. We have compared the magnitude of change in the four utility scores between baseline and 6-month follow-up in relation to the changes in the MCS and PCS scores. We found that the SF-6D had the highest magnitude of change in scores among schizophrenia spectrum disorder patients who have achieved improvement in the MCS. Hence, further studies are warranted to confirm this finding by performing a direct comparison of magnitude of change in scores among chronic schizophrenia patients.

The findings provide evidence that the HUI3 has good test–retest reliability, convergent validity and known-group validity than other utility measures in depression sample. The analysis confirmed that the HUI3 test–retest reliability was high (ICC = 0.83). This finding appears to be consistent with previous studies, which also reported that the test–retest reliability for the HUI3 was in the acceptable range, with ICC of 0.70 or higher in patients recovering from hip fracture and patients with rheumatic disease (Luo et al., 2003; Jones et al., 2005). We found that the HUI3 utility scores had the highest correlation with Q-LES-Q total scores (0.79) as compared to the other utility measures. Our findings are also similar to that of the study by Harnam et al. (2011) among people with generalized anxiety disorder that found high correlations between HUI3 and Q-LES-Q. Hence, these results suggest that the HUI3 utility scores were psychometrically valid for use in depression with high level of convergence with the Q-LES-Q. Known-group validity results also support evidence of discriminant ability of the HUI3 utility measures to differentiate people with mild and severe conditions in depression sample. The SRM values indicate that the performance of the HUI3 utility was very strong in this sample. The current findings correspond well with the study by Heintz et al. (2012) and Espallargues et al. (2005), which showed that the HUI3 is more sensitive than that EQ-5D and a visual analogue scale in detecting differences. However, the findings seems to partially support the findings of Luo et al. (2009a, 2009b), who compared the utility scores produced by the EQ-5D, HUI2, and HUI3 in the general population in the United States and found evidence that the EQ-5D, HUI2, and HUI3 utility scores were generally comparable in determining health burden of chronic conditions in the population.

We found mixed evidence for the magnitude of change in the four utility scores in depression sample. Although the EQ-5D-5L utility scores had slightly higher magnitude of change in scores than other utility scores in the overall sample, results from subgroup analysis

indicate that the HUI3 was more sensitive in detecting change in the quality of life status among patients with severe depression (PHQ-8 scores ≥10) after they underwent 6 months outpatient treatment follow-up than EQ-5D-5L. We found that the effect size of HUI3 at the 6 months follow up for patients who were clinically depressed at baseline was higher than EQ-5D-5L (0.69 vs. 0.54). We also found that the HUI3 seem to have higher magnitude of change in scores among depression patients who have achieved improvement in the MCS. The HUI3 utility scores achieved clinically significant improvement and had the highest mean difference between baseline and 6-month follow-up as compared to the other utility scores. The current study has adopted the distribution based approach to determine MCID. Yet many studies in the literature prefer to use anchor based approach instead of the distribution approach because a statistically significant improvement derived from distribution based approach may not actually be meaningful to the patient. However, it should be noted that the use of MCID values based on anchor based approach also has limitations. The MCID values can vary due many factors including the characteristics of sample, instrument, calculation method (anchor or ROC), type of anchor and study design (Katz et al., 2015). Hence, it is recommended that the best MCID values should be used from a similar study and context. To our knowledge, few studies have been conducted to investigate direct comparison of HUI3 against other common utility measures in mental health population. Wee et al. (2007a, 2007b) similarly compared EQ-5D-3L, HUI2, HUI3 and SF-6D in a primary care setting in Singapore and found that the SF-6D utility scores were generally higher than the EQ-5D, HUI2, or HUI3, but suggested that despite a wide range of difference in utility estimates the variations have relatively small impact on CUA according to their hypothetical decision trees analysis (Wee et al., 2007b). Furthermore, the authors mentioned that the findings need to be interpreted with caution as direct comparisons were only based on cross-sectional data in a primary care population and suggested that the impact of differences in change scores on CUA should be investigated in future in longitudinal studies and in other clinical populations.

In the current study we found that slightly higher rates of patients reported perfect health in the EQ-5D-3L than the EQ-5D-5L which suggest that the EQ-5D-3L seems to have a higher ceiling effect than the EQ-5D-5L. This is consistent with previous findings which found the EQ-5D-5L to be a more precise and valid measure than the EQ-5L-3L (Janssen et al., 2018). We found that a higher proportion of patients reported perfect health in the schizophrenia spectrum disorders than in the depression samples. These results were consistent with other studies conducted in clinical and general populations (Sagadevan et al., 2018; Goppoldova et al., 2008; Penner-Goeke et al., 2015). Sagadevan et al. (2018), for instance, have reported that patients with schizophrenia reported the highest quality of life, followed by patients with pathological gambling, anxiety disorders and depressive disorders. Similarly, Penner-Goeke et al. (2015) found that mood disorders were associated with highest decrease in quality of life as compared to psychotic illness. Several factors may explain these findings. Previous studies have suggested that majority of patients with schizophrenia lack awareness or have poor insight into their illness and this was found to be significantly associated with higher quality of life (Karow and Pajonk, 2006; Lysaker et al., 2018). It was suggested that poor insight may be associated with more severe symptoms and poorer community function while better insight may lead to depression, demoralization and low levels of quality of life (Lysaker et al., 2018). However, interpretation of aforementioned findings might not be applicable to our current study due to fact that patients with schizophrenia had predominantly mild to moderate symptom severity. They were stable outpatients as assessed by psychiatrists at the point of recruitment who were cognitively capable of providing consent and self-administered the questionnaires. The current results showed that the proportion of moderately severe symptoms among schizophrenia spectrum disorder patients (5.1%, $n = 13$) was significantly lower than the proportion of clinical depression (60%,

n = 151) among patients with depression. Hence, it is possible that the proportion of patients reporting perfect health in the schizophrenia sample might have been much lower as compared to patients with depression if inpatients or patients with more severe conditions had been included in the current study. However, given that the present study has limited data to confirm these hypotheses, future studies are needed to test these associations in our patient population.

Our study is not without limitations. First, given that we only included patients with schizophrenia and patients with depression, the generalizability of our findings to other mental disorders and more severely ill hospitalized patients with schizophrenia needs further investigation. Second, our classification of the patients into severity levels according to the PHQ-8 and PANSS could be questioned as different cutoffs may produce different effect size in utility scores. Third, the measures were administered only in English. Hence, the performance of the generic-preference based measures among those who were not fluent in English remains unclear. Another limitation of the study is that we were unable to assess responsiveness of utility scores by groups i.e., among patients who improve, get worse, or remain stable based on severity scales between the baseline and the 6-month follow-up. For instance, Dickerson et al. (2018) used the Children's Depression Rating Scale-Revised to classify patients as improved between the baseline and 12 weeks follow-up and based the estimates of responsiveness of the change in utility scores among patients who improved. In the current study, the severity of symptoms among patients with schizophrenia and patients with depression as assessed by the PANSS and the PHQ-8 are only available at baseline. Hence, we are unable to classify whether patients were stable over the two-week period during the test re-test assessment and at 6-month follow-up based on severity scales. Lastly, the EQ-5D-5L, HUI3 and SF-6D utility scores were derived from other populations which may not be appropriate because health preferences may vary between countries and cultures (Abdin et al., 2015). These limitations notwithstanding, to our knowledge, this is the first study in a multi-ethnic Asian population that directly compares the reliability, validity, and magnitude of change in the four utility scores (EQ-5D-5L, EQ-5D-3L, HUI3, SF-6D) in patients with schizophrenia spectrum disorders and patients with depression. The current study provides important evidence regarding the performance of GPBMs as head-to-head comparisons between GBPMs have been lacking and the findings on comparative study on GBPMs reported by the reviews were often weak (Finch et al., 2018).

5. Conclusions

In summary, the current study provides important evidence to clinicians and researchers about the best choice of generic preference-based measures that can be used to assess health-related quality of life of patients with mental illness in an Asian patient population. The direct comparison of four utility scores in patients with schizophrenia spectrum disorder and patients with depression demonstrated that the SF-6D and HUI3 utility scores could produce reliable, valid and high magnitude of change results when used in economic evaluation for measuring benefits of mental health interventions including improvement in health-related quality of life in patients with schizophrenia spectrum disorders and patients with depressive disorders respectively.

Funding

Conflict of interest

The authors declare that they have no conflict of interest.

Ethical approval

Informed consent

Informed consent was obtained from all individual participants included in the study.

Supplementary materials

References

Abdin, E., Subramaniam, M., Vaingankar, J.A., Luo, N., Chong, S.A., 2015. Population norms for the EQ-5D index scores using Singapore preference weights. Qual. Life Res. 24 (6), 1545–1553.

Brazier, J., 2010. Is the EQ-5D fit for purpose in mental health? Br. J. Psychiatry 197 (5), 348–349.

Brazier, J., Connell, J., Papaioannou, D., Mukuria, C., Mulhern, B., Peasgood, T., et al., 2014. A systematic review, psychometric analysis and qualitative assessment of generic preference-based measures of health in mental health populations and the estimation of mapping functions from widely used specific measures. Health Technol. Assess. 18 (34), 1–188.

Brazier, J., Roberts, J., Deverill, M., 2002. The estimation of a preference-based measure of health from the SF-36. J. Health Econ. 21 (2), 271–292.

Cleveland, W.S., 1979. Robust locally weighted regression and smoothing scatterplots. J. Am. Stat. Assoc. 74 (368), 829–836.

Dickerson, J.F., Feeny, D.H., Clarke, G.N., MacMillan, A.L., Lynch, F.L., 2018. Evidence on the longitudinal construct validity of major generic and utility measures of health-related quality of life in teens with depression. Qual. Life Res. 27 (2), 447–454.

Espallargues, M., Czoski-Murray, C.J., Bansback, N.J., Carlton, J., Lewis, G.M., Hughes, L.A., et al., 2005. The impact of age-related macular degeneration on health status utility values. Invest. Ophthalmol. Vis. Sci. 46 (11), 4016–4023.

EuroQol Group, 1990. EuroQol – a new facility for the measurement of health-related quality of life. Health Policy 16 (3), 199–208.

Farivar, S.S., Cunningham, W.E., Hays, R.D., 2007. Correlated physical and mental health summary scores for the SF-36 and SF-12 Health Survey, V.1. Health Qual. Life Outcomes 5 (54), 1–8.

Feeny, D., Furlong, W., Boyle, M., Torrance, G.W., 1995. Multi-attribute health status classification systems. Health Utilities Index. Pharmacoeconomics. 7 (6), 490–502.

Feeny, D., Spritzer, K., Hays, R.D., Liu, H., Ganiats, T., Kaplan, R.M., et al., 2012. Agreement about identifying patients who change over time: cautionary results in cataract and heart failure patients. Med. Decis. Mak. 32 (2), 273–286.

Fiest, K.M, Sajobi, T.T., Wiebe, S., 2014. Epilepsy surgery and meaningful improvement in quality of life: results from a randomized controlled trial. Epilepsia 55 (6), 886–892.

Finch, A.P., Brazier, J.E., Mukuria, C., 2018. What is the evidence for the performance of generic preference-based measures? A systematic overview of reviews. Eur. J. Health Econ. 19 (4), 557–570.

Furukawa, T.A., Levine, S.Z., Tanaka, S., Goldberg, Y., Samara, M., Davis, J.M., et al., 2015. Initial severity of schizophrenia and efficacy of antipsychotics: participant-level meta-analysis of 6 placebo-controlled studies. JAMA Psychiatry 72 (1), 14–21.

Gao, F., Ng, G.Y., Cheung, Y.B., Thumboo, J., Pang, G., Koo, W.H., et al., 2009. The Singaporean English and Chinese versions of the EQ-5D achieved measurement equivalence in cancer patients. J. Clin. Epidemiol. 62 (2), 206–213.

Guyatt, G.H., Osoba, D., Wu, A.W., Wyrwich, K., Norman, G.R., Clinical Significance Consensus Meeting Group, 2002. Methods to explain the clinical significance of health status measures. Mayo Clin. Proc. 77 (4), 371–383.

Goppoldova, E., Dragomirecka, E., Motlova, L., Hajek, T., 2008. Subjective quality of life in psychiatric patients: diagnosis and illness-specific profiles. Can. J. Psychiatry 53, 587–593.

Harnam, N., Wyrwich, K.W., Revicki, D., Locklear, J.C., Endicott, J., 2011. The measurement of health-related quality of life in a population with generalized anxiety disorder—findings from the QUEST study (Editor) In: Rijeka, S.S. (Ed.), Different Views of Anxiety Disorder. InTech publisher, pp. 199–216.

Heintz, E., Wirehn, A.B., Peebo, B.B., Rosenqvist, U., Levin, L.A., 2012. QALY weights for diabetic retinopathy–a comparison of health state valuations with HUI-3, EQ-5D, EQ-VAS, and TTO. Value Health 15 (3), 475–484.

Horsman, J., Furlong, W.D., Feeny, D., Torrance, G., 2003. The Health Utilities Index (HUI): concepts, measurement properties and applications. Health Qual. Life Outcomes 1 (54), 1–13.

Janssen, M.F., Bonsel, G.J., Luo, N., 2018. Is EQ-5D-5L better than EQ-5D-3L? A head-to-head comparison of descriptive systems and value sets from seven countries. Pharmacoeconomics 36 (6), 675–697.

Jiang, J., Sim, K., Lee, J., 2013. Validated five-factor model of positive and negative syndrome scale for schizophrenia in Chinese population. Schizophr. Res. 143 (1),

38–43.

Jones, C.A., Feeny, D., Eng, K., 2005. Test–retest reliability of health utilities index scores: evidence from hip fracture. Int. J. Technol. Assess. Health Care 21 (3), 393–398.

Kay, S.R., Fiszbein, A., Opler, L.A., 1987. The positive and negative syndrome scale (PANSS) for schizophrenia. Schizophr. Bull. 13 (2), 261–276.

Karow, A., Pajonk, F.G., 2006. Insight and quality of life in schizophrenia: recent findings and treatment implications. Curr. Opin. Psychiatry 19 (6), 637–641.

Katz, N.P., Pailard, F.C., Ekman, E., 2015. Determining the clinical importance of treatment benefits for interventions for painful orthopedic conditions. J. Orthop. Surg. Res. 10 (24), 1–11.

Khanna, D., Furst, D.E., Wong, W.K., Tsevat, J., Clements, P.J., Park, G.S., Scleroderma Collagen Type 1 Study, 2007. Reliability, validity, and minimally important differences of the SF-6D in systemic sclerosis. Qual. Life Res. 16 (6), 1083–1092.

Koh, E.T., Leong, K.P., Tsou, I.Y.Y., Lim, V.H., Pong, L.Y., Chong, S.Y., Tan Tock Seng Hospital Rheumatoid Arthiritis (TTSH RA) Study Group, 2006. The reliability, validity and sensitivity to change of the Chinese version of SF-36 in oriental patients with rheumatoid arthiritis. Rheumatology 45, 1023–1028.

Koo, T.K., Li, M.Y., 2016. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. J. Chiropr. Med. 15 (2), 155–163.

Kroenke, K., Strine, T.W., Spitzer, R.L., Williams, J.B., Berry, J.T., Mokdad, A.H., 2009. The PHQ-8 as a measure of current depression in the general population. J. Affect. Disord. 114 (1–3), 163–173.

Leucht, S., Kane, J.M., Kissling, W., Hamann, J., Etschel, E., Engel, R.R., 2005. What does the PANSS mean? Schizophr. Res. 79 (2–3), 231–238.

Leung, Y.Y., Png, M.E., Wee, H.L., Thumboo, J., 2013. Comparison of EuroQol-5D and short form-6D utility scores in multiethnic Asian patients with psoriatic arthritis: a cross-sectional study. J. Rheumatol. 40 (6), 859–865.

Luo, N., Chew, L.H., Fong, K.Y., Koh, D.R., Ng, S.C., Yoon, K.H., et al., 2003. Validity and reliability of the EQ-5D self-report questionnaire in Chinese-speaking patients with rheumatic diseases in Singapore. Ann. Acad. Med. Singap. 32 (5), 685–690.

Luo, N., Johnson, J.A., Shaw, J.W., Coons, S.J., 2009a. Relative efficiency of the EQ-5D, HUI2, and HUI3 index scores in measuring health burden of chronic medical conditions in a population health survey in the United States. Med. Care. 47 (1), 53–60.

Luo, N., Low, S., Lau, P.N., Au, W.L., Tan, L.C., 2009b. Is EQ-5D a valid quality of life instrument in patients with Parkinson's disease? A study in Singapore. Ann. Acad. Med. Singap. 38 (6), 521–528.

Luo, N., Seng, B.K., Thumboo, J., Feeny, D., Li, S.C., 2006. A study of the construct validity of the Health Utilities Index Mark 3 (HUI3) in patients with schizophrenia. Qual. Life Res. 15 (5), 889–898.

Luo, N., Seng, B.K., Xie, F., Li, S.C, Thumboo, J., 2008. Psychometric evaluation of the Schizophrenia Quality of Life Scale (SQLS) in English- and Chinese-speaking Asians in Singapore. Qual. Life Res. 17 (1), 115–122.

Luo, N., Wang, P., Thumboo, J., Lim, Y.W., Vrijhoef, H.J., 2014. Valuation of EQ-5D-3L health states in Singapore: modeling of time trade-off values for 80 empirically observed health states. Pharmacoeconomics 32 (5), 495–507.

Luo, N., Wang, Q., Feeny, D., Chen, G., Li, S.C., Thumboo, J., 2007. Measuring health preferences for Health Utilities Index Mark 3 health states: a study of feasibility and preference differences among ethnic groups in Singapore. Med. Decis. Mak. 27 (1), 61–70.

Lysaker, P.H., Pattison, M.L., Leonhardt, B.L., Phelps, S., Vohs, J.L., 2018. Insight in schizophrenia spectrum disorders: relationship with behavior, mood and perceived quality of life, underlying causes and emerging treatments. World Psychiatry 17 (1), 12–23.

Mulhern, B., Mukuria, C., Barkham, M., Knapp, M., Byford, S., Soeteman, D., et al., 2014. Using generic preference-based measures in mental health: psychometric validity of the EQ-5D and SF-6D. Br. J. Psychiatry 205 (3), 236–243.

Ong, C., Sagayadevan, V., Lee, S.P., Ong, R., Chong, S.A., Frost, R., et al., 2016. Hoarding among outpatients seeking treatment at a psychiatric hospital in Singapore. J. Obsessive-Compuls. Relat. Disord. 8, 56–63.

Papaioannou, D., Brazier, J., Parry, G., 2011. How valid and responsive are generic health status measures, such as EQ-5D and SF-36, in schizophrenia? A systematic review. Value Health 14 (6), 907–920.

Peasgood, T., Brazier, J., Papaioannou, D., 2012. A Systematic Review of the Validity and Responsiveness of EQ-5D and SF-6D for Depression and Anxiety. The University of Sheffield HEDS Discussion Paper No.12.15.

Penner-Goeke, K., Henriksen, C.A., Chateau, D., Latimer, E., Sareen, J., Katz, LY., 2015. Reductions in quality of life associated with common mental disorders: results from a nationatially representative sample. J. Clin. Psychiatry 76 (11), 1506–1512.

Sagadevan, V., Lee, S.P., Ong, C., Abdin, E., Chong, S.A., Subramaniam, M., 2018. Quality of life across mental disorders in psychiatric outpatients. Ann. Acad. Med. Singap. 47 (7), 243–252.

Subramaniam, M., Abdin, E., Poon, L.Y., Vaingankar, J.A., Lee, H., Chong, S.A., et al., 2014. EQ-5D as a measure of programme outcome: results from the Singapore early psychosis intervention programme. Psychiatry Res. 215 (1), 46–51.

Thumboo, J., Fong, K.Y., Chan, S.P., Machin, D., Feng, P.H., Thio, S.T., et al., 2002. The equivalence of English and Chinese SF-36 versions in bilingual Singapore Chinese. Qual. Life Res. 11 (5), 495–503.

Thumboo, J., Wu, Y., Tai, E.S., Gandek, B., Lee, J., Ma, S., et al., 2013. Reliability and validity of the English (Singapore) and Chinese (Singapore) versions of the short-form 36 version 2 in a multi-ethnic urban Asian population in Singapore. Qual. Life Res. 22 (9), 2501–2508.

Tuynman-Qua, H., de Jonghe, F., McKenna, S.P., 1997. Quality of life in depression scale (QLDS). Development, reliability, validity, responsiveness and application. Eur. Psychiatry 12 (4), 199–202.

van Hout, B., Janssen, M.F., Feng, Y.S., Kohlmann, T., Busschbach, J., Golicki, D., et al., 2012. Interim scoring for the EQ-5D-5L: mapping the EQ-5D-5L to EQ-5D-3L value sets. Value Health 15 (5), 708–715.

Walters, S., 2009. Quality of Life Outcomes in Clinical Trials and Health-Care Evaluation: A Practical Guide to Analysis and Interpretation. A John Wiley and Sons, Ltd.

Wee, H.L., Cheung, Y.B., Fong, K.Y., Luo, N., Machin, D., Thumboo, J., 2004. Are English- and Chinese-language versions of the SF-6D equivalent? A comparison from a population-based study. Clin. Ther. 26 (7), 1137–1148.

Wee, H.L., Loke, W.C., Li, S.C., Fong, K.Y., Cheung, Y.B., Machin, D., et al., 2007a. Cross-cultural adaptation and validation of Singapore Malay and Tamil versions of the EQ-5D. Ann. Acad. Med. Singap. 36 (6), 403–408.

Wee, H.L., Machin, D., Loke, W.C., Li, S.C., Cheung, Y.B., Luo, N., et al., 2007b. Assessing differences in utility scores: a comparison of four widely used preference-based instruments. Value Health 10 (4), 256–265.

Wiebe, S., Matijevic, S., Eliasziw, M., Derry, P.A., 2002. Clinically important change in quality of life in epilepsy. J. Neurol. Neurosurg. Psychiatry 73, 116–120.

Wilkinson, G., Hesdon, B., Wild, D., Cookson, R., Farina, C., Sharma, V., et al., 2000. Self-report quality of life measure for people with schizophrenia: the SQLS. Br. J. Psychiatry 177, 42–46.

Xie, F., Li, S.C., Luo, N., Lo, N.N., Yeo, S.J., Yang, K.Y., et al., 2007. Comparison of the EuroQol and short form 6D in Singapore multiethnic Asian knee osteoarthritis patients scheduled for total knee replacement. Arthritis Rheumatol. 57 (6), 1043–1049.