



Measuring health-related quality of life and well-being: a head-to-head psychometric comparison of the EQ-5D-5L, ReQoL-UI and ICECAP-A

Richard Huan Xu^{1,2} · Anju Devianee Keetharuth³ · Ling-ling Wang⁴ · Annie Wai-ling Cheung² · Eliza Lai-yi Wong²

Received: 11 March 2021 / Accepted: 20 July 2021 / Published online: 2 August 2021

© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2021

Abstract

Objective This study aimed to assess the psychometric properties of three generic preference-based measures and compare their performance in a sample of Hong Kong general population.

Methods Data used for this analysis were obtained from a cross-sectional telephone-based survey in July 2020. Participants were asked to complete several measures, including The EuroQol five-dimensional five levels (EQ-5D-5L), Recovering Quality of Life-Utility Index (ReQoL-UI) and ICEpop CAPability measure for adults (ICECAP-A). Acceptability, reliability, convergent and discriminant validity of three measures were assessed as well as the agreement between these instruments.

Results Based on data from 500 participants to the survey, a lower mean score of the ICECAP-A (mean = 0.85) was observed compared to the other two measures (mean_{ReQoL-UI} = 0.92; mean_{EQ-5D-5L} = 0.92). All three measures showed an acceptable internal consistency reliability (Cronbach's alpha = 0.74, 0.82 and 0.77, respectively) as well as good test–retest reliability (intra-class correlation coefficient = 0.74, 0.82 and 0.77, respectively). Correlation analyses confirmed satisfactory convergent validity and the ability of the measures to differentiate between participants with different health or from socioeconomic status groups. The Bland–Altman plot revealed poor agreement between the three measures.

Conclusions This study confirmed that EQ-5D-5L, ReQoL-UI and ICECAP-A were psychometrically robust to measure HRQoL in the general HK population. The EQ-5D-5L was more suitable for assessing physical HRQoL, whereas the ICECAP-A and ReQoL-UI were more appropriate for measuring interventions aimed at improving people's well-being and mental health.

Keywords EQ-5D-5L · ReQoL-UI · ICECAP-A · Psychometric evaluations

Eliza Lai-yi Wong considered as senior author.

✉ Richard Huan Xu
richard.xu@polyu.edu.hk

✉ Eliza Lai-yi Wong
lywong@cuhk.edu.hk

¹ Department of Rehabilitation Sciences, Hong Kong Polytechnic University, Hong Kong SAR, China

² Centre for Health Systems and Policy Research, Jockey Club School of Public Health and Primary Care, The Chinese University of Hong Kong, Hong Kong SAR, China

³ School of Health and Related Research, The University of Sheffield, Sheffield, UK

⁴ Department of Blood Transfusion, School of Medicine, Jinling Hospital, Nanjing University, Nanjing, China

Introduction

Economic evaluation is the most frequently used method in health care programme. It uses empirical techniques applied to cost and outcome measures to inform resource allocation in specific populations and settings [1]. Currently, the EQ-5D is one of the most widely used generic preference-based measures (GPBMs) assessing individual's health-related quality of life (HRQoL) to facilitate the economic evaluation of health care interventions [2]. It has been shown to be valid in different patient groups and settings [3]. Although the use of the EQ-5D has grown in recent decades, its ability to capture and assess people's mental HRQoL and well-being is questionable [4–6]. Recent studies have indicated that poor physical health is highly likely to lead to an increased risk of developing impaired mental health due to insecurity, confusion and emotional isolation [7, 8],

unsatisfied social well-being due to loss of wealth, work and school closure and shortage of acquiring adequate medical services [9, 10]. Pfefferbaum and North further indicated that impaired mental health and well-being may result in unhealthy behaviours and exacerbate people's physical health [11]. Several studies argued that the EQ-5D, which main focus is on certain aspects of physical health (four out of five items) may not adequately capture and measure the effectiveness of mental health, public health and social care interventions, which are issues certain to be echoed in populations affected by both acute and chronic diseases [12, 13].

The Recovering Quality of Life-Utility Index (ReQoL-UI) is a new GPBM that aims to capture changes in mental HRQoL [14]. It was developed on the basis of the theoretical framework established with considerable input from mental health service users, which is believed to provide different perspectives to the comparability across evaluations undertaken in physical and mental health [15]. The developers indicated that the ReQoL-UI has the advantage to detect psychometrically changes in HRQoL over time and differences across treatments. An alternative framework for measuring the cost-effectiveness of social care interventions is with the ICEpop CAPability (ICECAP) measures, which is theoretically grounded in Amartya Sen's capability approach [16]. It was designed to measure people's capability (what an individual can do) rather than function (what they actually do) to highlight the importance of freedom to choose. It focuses on well-being defined in a broader sense rather than health [17]. The ICECAP instruments have different versions, among them, the adult version (ICECAP-A) is validated in the Chinese population [18].

Although GPBMs are increasingly used to evaluate the effectiveness of health and social care interventions, there is little evidence to inform the selection of the most appropriate one for use in economic evaluations in the Hong Kong (HK) general population. Using reliable and appropriate instrument is vital to ensure the benefits of the interventions and policies are adequately capturing [19]. Thus, this study aimed to assess the psychometric properties of three GPBMs, the EQ-5D-5L, ReQoL-UI and ICECAP-A, and compare their performance in a sample of HK general population to inform instrument choice when conducting economic evaluation for public health and social care interventions, especially where mental health is an important component.

Methods

Sample size

For conducting psychometric analysis, a minimum of 300 respondents is required [20]. Given the possibility of missing

data, in this study, a target sample size of 500 from the HK general population was considered sufficient to perform such analysis.

Participants and data collection

A telephone survey was carried out in July 2020 to recruit participants. To minimize the sampling error, first, telephone numbers were selected randomly from the updated available public telephone directories as seed numbers. Another three sets of numbers were then generated using the randomization of the last two digits to recruit the unlisted numbers. Duplicate numbers were screened out, with the remaining numbers mixed in a random order to form the final sample. A total of 5,385 telephone numbers were sampled for the survey. The inclusion criteria for the study were HK permanent residents, ≥ 18 years, and able to speak Cantonese. Upon successful contact with a target household, the adult who have had their birthday most recently was selected to complete a questionnaire over the phone. Study protocol and informed consent was approved by the institutional review board of the Chinese University of Hong Kong (Ref. ID: SBRE-18-671).

Measurements

EQ-5D-5L

The Chinese EQ-5D-5L used in this study was approved by the EuroQol Group (www.euroqol.org). The descriptive system comprises five items (mobility, self-care, usual activities, pain/discomfort and anxiety/depression) with five levels (no problem to extreme problems) [21], which can be converted into a summarised utility score between 0 (death) and 1 (full health) to facilitate cost-utility analysis. The utility score was estimated based on HK population's preference weights [22]. We also administered the visual analogue scale (EQ-VAS) to describe individual's overall health status (0 [worst]–100 [best]).

ReQoL-UI

The ReQoL-UI, which was developed based on the ReQoL-20, comprising six mental health items (activity; belonging and relationships; choice, control and autonomy; hope; self-perception; and well-being) and one physical health item was administered [14]. The ReQoL-UI has been translated to Chinese and adapted for use in HK with the necessary permissions [23]. In the absence of HK specific preference weights, we used the UK preference weights to calculate the utility score in this study. The weights were estimated from a sample of 305 UK general population using the time trade-off method [15]. The ReQoL-UI utility score ranges

between -0.195 and 1 , which reflects people's worst and best recovered HRQoL, respectively.

ICECAP-A

The ICECAP-A is a well-being measure assessing an adult's capability. The five attributes measured are stability, attachment, autonomy, achievement and enjoyment [16]. In this study, utility score of the ICECAP-A was calculated using the tariffs obtained from the UK general population using the best–worst scaling method [24]. The ICECAP-A utility score ranges between 0 (no capability) and 1 (full capability). The Chinese version of the ICECAP-A was approved by the University of Birmingham and its psychometric properties was reported by Tang et al. [18].

General anxiety disorder—7 items (GAD-7)

The GAD-7 is a self-rated scale to measure the severity of generalized anxiety disorder. It has seven items scored from zero (not at all) to three (nearly every day) [25]. Cut-off point of the GAD-7 for mild, moderate and severe anxiety are 5 , 10 and 15 , respectively. The psychometric properties of the Chinese GAD-7 was reported by Tong et al. [26].

Depression anxiety stress scales—21 (DASS-21)

The DASS-21 consists of three sub-scales to assess the emotional states of depression, anxiety and stress [27]. Scores of each item range from 0 (never applied to oneself) to 3 (very much/most of the time). Final scores are calculated by summing the scores for relevant items and then multiplied by two. The cut-off points identified for no clinical problems are 9 , 7 and 14 for three sub-scales, respectively [27]. Psychometric properties of the Chinese DASS-21 was reported by Gong et al. [28].

Sociodemographic characteristics and other indicators

Information about respondents' demographics (sex and age), socioeconomic status (marital status, educational level, employment, living status, government allowance and personal income), health conditions (chronic condition and cognitive ability) and social well-being (life satisfaction and social relationship) were collected.

Statistical analysis

R software was used to perform all statistical analyses [29]. The level of significance was set at p value ≤ 0.05 . The acceptability, reliability, discriminant and convergent validity and correlations and agreement between three measures were assessed in this study.

Acceptability

We assessed the completion rate of the three measures which we expected to be similar given their comparable length. In addition, the proportion of missing values, score ranges, the floor (percentage with lowest possible score) and ceiling effects (percentage with highest possible score) were reported to assess the acceptability.

Internal consistency and test–retest reliability

Cronbach's alpha (α) was used to assess the internal consistency reliability, where $\alpha > 0.7$ was identified as acceptable. A random sample of 50 respondents (10%) was invited to complete the measures two weeks later to evaluate the test–retest reliability of the measures using intra-class correlation coefficient (ICC, two-way mixed model, > 0.7 acceptable) [30]. The measures were expected to have similar reliability given they have similar response structure and number of items.

Convergent validity and hypothesized correlations between measures

Convergent validity was evaluated by investigating a priori hypothesized associations using Pearson correlation coefficient ($r \geq 0.7$, strong; $r > 0.5$, moderate; $r > 0.2$, weak) [31]. We hypothesized that the three measures would show a positive and moderate/strong association with participants' overall health status measured by the EQ-VAS. In addition, to test the convergent validity, we formulated the following hypotheses based on the concepts measured by each instrument: a. weak correlation among the utility scores of the three measures, as the concepts they are capturing are very different; b. moderate to strong negative correlation between EQ-5D utility and the physical item of the ReQoL-UI as the former has four items on physical health; c. moderate negative correlation between ReQoL-UI and ICECAP-A utility score and the anxiety and depression item of the EQ-5D; d. moderate negative association between the mental health items of ReQoL-UI and the ICECAP-A utility score.

Discriminant validity

Discriminant validity was assessed by examining the ability of the measures to differentiate people with different mental or physical health status, socioeconomic status and social well-being. We assumed that (a) respondents with no depression and no clinical signs using the GAD-7 and DASS-21 would report a high utility score; (b) respondents with no chronic conditions, and satisfied with their cognitive ability, life satisfaction and social relationship would report a high utility score; and (c) respondents with high socioeconomic

status (non-government allowance receivers, living with families, fully employed and well-paid) would report a high utility score.

Mann–Whitney *U* test (MW test) and Kruskal–Wallis one-way analysis of variance (KW test) were used to compare the differences between subgroups. Effect sizes (EZ) calculated based on *Z* score (MW test) and *H* score (KW test) were used to assess the discriminative power of the measures. Regarding the explanation of the EZ value, for MW test, $0.1 < EZ < 0.3$, $0.31 < EZ < 0.5$ and $EZ \geq 0.5$ were identified as weak, moderate and strong; for KW test, $0.01 < EZ < 0.059$, $0.06 < EZ < 0.139$ and $EZ > 0.14$ were identified as weak, moderate and strong [32, 33]. Separate multiple linear regression analyses were used to predict the utility score of three measures based on respondents' sociodemographic variables.

Agreement between measures

Agreement between measures was determined using Bland–Altman (B–A) plot and ICC. Regarding B–A plot, the y-axis represents the difference between utility scores of two measures and x-axis represents the mean of utility scores of two measures. The score distribution across the mean difference of two measures represent a good agreement. We assumed the agreement between three measures is poor given their different conceptual structures.

Results

Respondents' characteristics and feasibility

A total of 500 respondents responded to the survey and provided valid responses. 72.2% ($n = 361$) were female, 60.6% ($n = 303$) were older than 60 years, and over one third ($n = 174$) completed primary school-level or below education. Additionally, nearly 90% ($n = 448$) reported living with their families, 27.8% were fully employed and over two third ($n = 313$) reported an income of ≤ 5000 HKD (\$650 USD) per month (Table 1). All respondents completed the EQ-5D-5L, ReQoL-UI and ICECAP-A, an indication of the feasibility of administering the three measures.

Acceptability

The utility scores of the ReQoL-UI, EQ-5D-5L and ICECAP-A covered nearly the full possible range. The ICECAP-A showed a lower mean score of 0.85 (range: 0.29–1) than the other two measures (Mean_{ReQoL-UI} = 0.92 [0.34–1]; Mean_{EQ-5D-5L} = 0.92 [0.01–1]). Analysis at the item level showed that 88.6, 68 and 59.8% of respondents reported no

Table 1 Respondent's characteristics

	<i>n</i>	%
Overall	500	100
Sex		
Male	139	27.8
Female	361	72.2
Age		
18–49	107	21.4
50–59	90	18.0
60–69	138	27.6
≥ 70	165	33.0
Educational level		
Primary or below	174	34.8
Secondary/post-secondary	216	43.2
Tertiary or above	106	21.2
Marital status		
Single	67	13.4
Married	381	76.2
Divorced/widow(er)	50	10.0
Living status		
Living alone	52	10.4
Living with families	448	89.6
Employment		
Employed	139	27.8
Non-employed	179	35.8
Retired	182	36.4
Government allowance		
Receiver	177	35.4
Non-receiver	323	64.6
Personal income per month		
HKD $\leq 5000^a$	313	62.6
HKD 5001–20,000	93	18.6
HKD $\geq 20,001$	45	9.0
Refused to answer	49	9.8

^a1 HKD = 0.13 US dollar

problems on hope, belonging and relationship and choice and autonomy of the ReQoL-UI, respectively. Around 70.8–93.2% and 23.4–60.4% of respondents reported no problems on all items of the EQ-5D-5L and ICECAP-A, respectively (Table 2). No missing data were identified. The distributions of the HRQoL measurement scores are presented in Fig. 1.

Reliability

The Cronbach's alpha of the ReQoL-UI, EQ-5D-5L and ICECAP-A were 0.74, 0.82 and 0.77, respectively, which

Table 2 Descriptive statistics, responses and reliability

	Mean (sd ^a)	Median	Min	Max	Minimum value (%)	Maximum value (%)	Cronbach's alpha	ICC (95% CI ^b)
ReQoL-UI							0.74	0.61 (0.37–0.76)
Activity	–	–	1	5	8.0	16.6	–	–
Belonging and relationship	–	–	1	5	68.0	0.8	–	–
Choice and autonomy	–	–	1	5	59.8	0.4	–	–
Hope	–	–	1	5	88.6	1.0	–	–
Self-perception	–	–	1	5	6.2	18.6	–	–
Well-being	–	–	1	5	18.2	13.4	–	–
Physical health	–	–	1	5	68.0	0.4	–	–
Utility	0.92 (0.09)	0.93	0.34	1	0.2	6.0	–	–
EQ-5D-5L							0.82	0.78 (0.65–0.86)
Mobility	–	–	1	3	84.6	2.0	–	–
Self-care	–	–	1	3	93.2	0.8	–	–
Usual activities	–	–	1	3	86.0	1.4	–	–
Pain/discomfort	–	–	1	4	70.8	0.8	–	–
Anxiety/depression	–	–	1	4	85.8	0.4	–	–
Utility	0.92 (0.14)	1	0.01	1	0.2	63.8	–	–
ICECAP-A							0.77	0.85 (0.75–0.9)
Attachment	–	–	1	4	38.6	1.8	–	–
Stability	–	–	1	4	40.8	0	–	–
Autonomy	–	–	1	4	34.8	0.2	–	–
Achievement	–	–	1	4	23.4	1.8	–	–
Enjoyment	–	–	1	4	60.4	0.4	–	–
Utility	0.85 (0.13)	0.88	0.29	1	0.2	11.2	–	–

^asd standard deviation

^b95% CI 95% confidence interval

showed an acceptable internal consistency reliability (Table 2). The ICC for the ReQoL-UI (0.74), EQ-5D-5L (0.82) and ICECAP-A (0.77) exceeded the recommended threshold of 0.7, which indicated a satisfactory test–retest reliability.

Correlation and convergent validity

The three measures showed significant correlation with each other. The ReQoL-UI was moderately associated with the EQ-5D-5L and ICECAP-A utility scores ($r = 0.55$ and 0.49 , respectively) as well as the overall health ($r = 0.55$). The association of the ICECAP-A with the EQ-5D-5L and overall health was small ($r = 0.35$). The EQ-5D-5L utility score moderately associated with the physical health item of the ReQoL-UI ($r = -0.67$), the pain/discomfort of the EQ-5D-5L significantly associated with ReQoL-UI utility score ($r = -0.51$). The ICECAP-A

utility score exhibited weak correlation with all items of the EQ-5D-5L, and weak/moderate correlation with six out of seven ReQoL-UI items, respectively (Table 3).

Discriminant validity

The ReQoL-UI, EQ-5D-5L and ICECAP-A showed satisfactory discriminant validity (Table 4). Respondents without mental health problem based on the outcomes of the GAD-7 and DASS-21 and not receiving treatment from a psychiatrist reported higher utility scores. The EQ-5D-5L ($ES = 0.32$, $p = 0.007$) and ReQoL-UI ($EZ = 0.16$, $p < 0.001$) exhibited a stronger discriminative ability than ICECAP-A to differentiate respondents with/without chronic conditions and cognitive problems, respectively. The EQ-5D-5L also showed a stronger discriminatory power than the other two measures regarding respondents' government allowance status ($EZ = -0.43$, $p < 0.001$) and income levels ($EZ = 0.06$, $p < 0.001$).

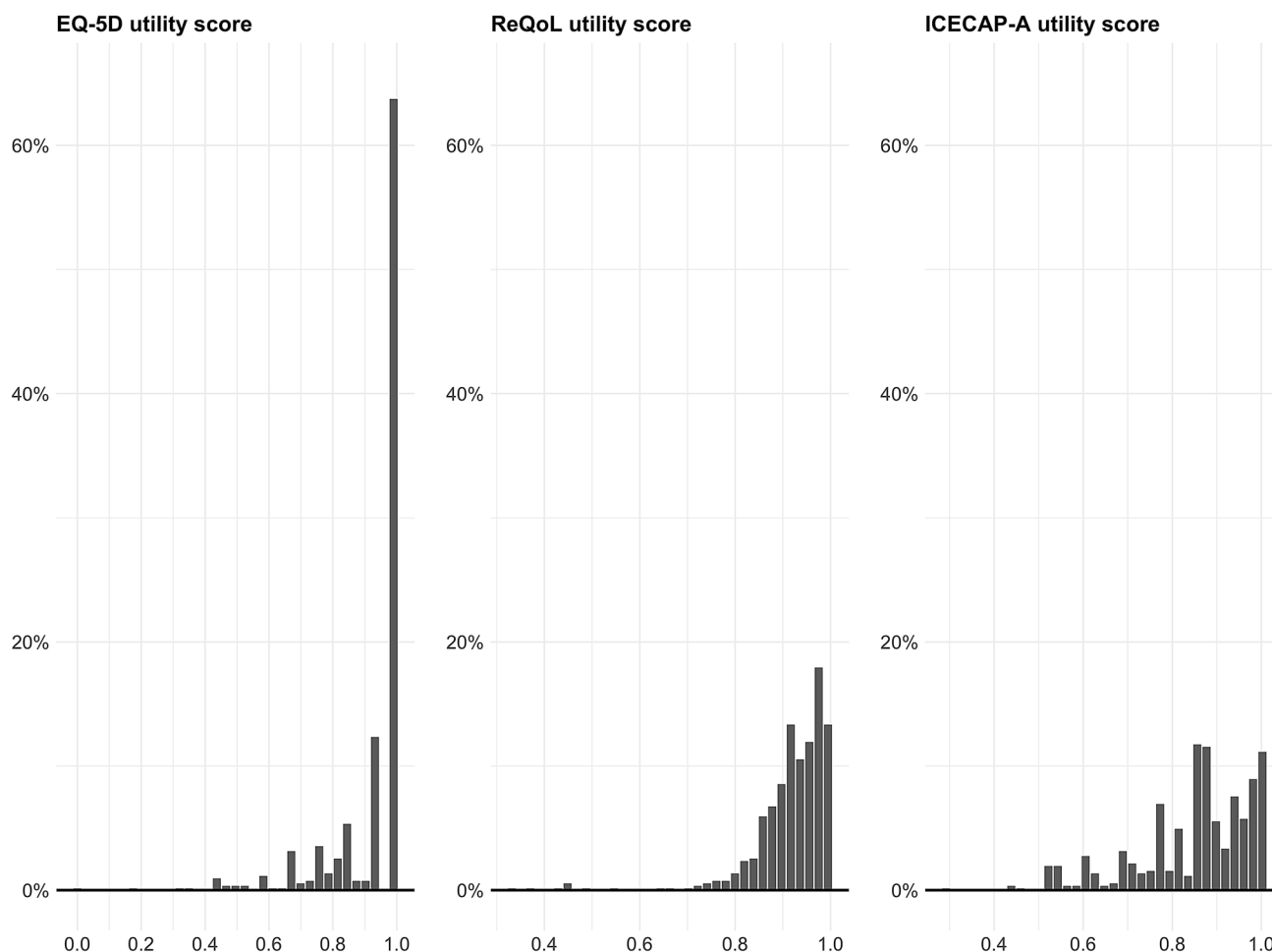


Fig. 1 Histogram of the EQ-5D-5L, ReQoL-UI and ICECAP-A utility scores

Results of multiple regression analysis

Figure 2 shows that education was a significant predictor for estimating the change of utility score of all three measures. Respondents, who were highly educated, showed a good HRQoL and well-being. Respondents, who were divorced/widowed, obtained a low EQ-5D-5L (coefficient = -0.095 , $p=0.002$) and ReQoL-UI (coefficient = -0.054 , $p=0.01$) utility score. Respondents with a good pay tended to report a high ICECAP-A utility score (coefficient = 0.064 , $p=0.04$).

Agreement between measures

The agreement between three measures was poor. The ICC of the EQ-5D-5L and ReQoL-UI was 0.5, which was higher than that of the other two pairs of comparison. The B–A plot demonstrates a wide limit of agreement interval between measures. A systematic difference of the agreement between the low utility scores of measures was observed, which

indicated respondents with poor health status/well-being are more likely to report less consistent utility scores (Fig. 3).

Discussion

This was the first study that directly compared the psychometric properties and performances of three GPBMs, the EQ-5D-5L, ReQoL-UI and ICECAP-A, in the HK general population. All of them exhibited satisfactory feasibility, reliability and validity to assess the population's HRQoL and well-being related outcomes. The ICECAP-A showed a stronger discriminant ability to differentiate people reporting different mental health status. However, the EQ-5D-5L outperformed the other two measures in subpopulation with different physical health and socioeconomic status. Given the conceptual structures of these measures were different, the low agreement between them was expected. Overall, the psychometric properties of three measures are relatively

Table 3 Correlations between and convergent validity of the EQ-5D, ReQoL-UI and ICECAP-A

	ReQoL-UI utility score	EQ-5D-5L utility score	ICECAP utility score
ReQoL-UI utility score	–	0.55***	–
ICECAP-A utility score	0.49***	0.35***	–
EQ-VAS (overall health)	0.55***	0.48***	0.35***
EQ-5D-5L			
Mobility	– 0.39***	–	– 0.2***
Self-care	– 0.37***	–	– 0.25***
Usual activities	– 0.34***	–	– 0.2***
Pain/discomfort	– 0.51***	–	– 0.33***
Anxiety/depression	– 0.46***	–	– 0.36***
ReQoL-UI			
Activity	–	– 0.19***	– 0.45***
Belonging and relationship	–	– 0.2***	– 0.2***
Choice and autonomy	–	– 0.19***	– 0.17***
Hope	–	– 0.04	– 0.45***
Self-perception	–	– 0.3***	– 0.5***
Well-being	–	– 0.1*	– 0.43***
Physical health	–	– 0.67***	– 0.37***
ICECAP-A			
Attachment	– 0.34***	– 0.13**	–
Stability	– 0.37***	– 0.23***	–
Autonomy	– 0.37***	– 0.23***	–
Achievement	– 0.34***	– 0.31***	–
Enjoyment	– 0.27***	– 0.39***	–

* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$

sound with EQ-5D-5L performing better than the other two measures in our sample of HK general population.

All the measures showed good feasibility and acceptability, because no missing data were detected, no ceiling or floor effect were observed, and utility values covered a nearly full score range. The values of α confirmed that internal consistency reliability of three measures were acceptable, which the EQ-5D-5L showed a good internal consistency reliability with 0.82, and the ICECAP-A and ReQoL-UI exhibited an acceptable reliability of 0.77 and 0.74, respectively. This finding was not unexpected because a great number of studies have confirmed the good performance of the EQ-5D-5L in HK Chinese population [34–37]. However, no empirical evidence about the other two measures in HK population was found, especially that this is the first paper using the ReQoL-UI in the HK population [15]. Additionally, the ICECAP-A and EQ-5D-5L exhibited acceptable to good test–retest reliability but the ICC for the ReQoL-UI was poor.

Regarding the correlation and convergent validity, all the measures showed a significant association with each other and moderately correlated with the respondents' overall health. This is consistent with findings in the literature. For example, a Hungarian study indicated that a correlation of

0.57 between the ICECAP-A and the EQ-5D-5L among the general public, and the correlation between the EQ-5D-5L and item achievement and enjoyment of the ICECAP-A was stronger than with the other items [38]. Another multi-centre study also exhibited a moderate association between the EQ-5D-5L and ICECAP-A in UK ($r=0.36$) and German (0.35) healthy population [39]. Nevertheless, no study about the relationship between the ICECAP-A and EQ-5D-5L in the Chinese population was found.

In this study, all three measures exhibited moderate ability to discriminate between people with different health and socioeconomic status. For instance, the EQ-5D-5L showed a stronger discriminatory power than the other measures in distinguishing people with different physical health and socioeconomic status, which was consistent with findings of previous studies [35, 40, 41]. However, some hypotheses were not confirmed. For example, although the ReQoL-UI and ICECAP-A were mainly designed to assess people's mental HRQoL and well-being, we found that the ICECAP-A showed a higher ability to differentiate between people with mental problems than the ReQoL-UI. In addition, compared with the ICECAP-A, the EQ-5D-5L and ReQoL-UI showed a stronger discriminatory power in differentiating people with satisfactions in social life, where we assumed

Table 4 Discriminant validity of the EQ-5D-5L, ReQoL-UI and ICECAP-A

	<i>n</i>	EQ-5D-5L	ReQoL-UI	ICECAP-A
Visited psychiatrists				
Yes	22	0.81 (0.26)	0.85 (0.15)	0.72 (0.18)
No	478	0.93 (0.13)	0.92 (0.08)	0.86 (0.12)
Z value (effect size)		2.75 (0.12)	2.92 (0.13)	3.7 (0.17)
<i>p</i> value		0.006	0.003	<0.001
GAD-7				
No depression (<5)	437	0.93 (0.13)	0.93 (0.07)	0.86 (0.12)
Mild (5–10)	45	0.88 (0.19)	0.87 (0.2)	0.75 (0.14)
Moderate or above (≥ 10)	18	0.81 (0.19)	0.82 (0.1)	0.75 (0.18)
KW H value (effect size)		21.53 (0.04)	23.71 (0.04)	24.53 (0.05)
<i>p</i> value		<0.001	<0.001	<0.001
DASS—Depression				
Clinical (≤9)	28	0.82 (0.21)	0.8 (0.15)	0.69 (0.17)
Non-clinical (>9)	472	0.93 (0.13)	0.92 (0.09)	0.86 (0.12)
Z value (effect size)		4.33 (0.19)	4.73 (0.21)	5.11 (0.23)
<i>p</i> value		<0.001	<0.001	<0.001
DASS—Anxiety				
Clinical (≤7)	30	0.85 (0.24)	0.84 (0.16)	0.75 (0.15)
Non-clinical (>7)	470	0.93 (0.12)	0.92 (0.08)	0.86 (0.13)
Z value (effect size)		4.35 (0.2)	3.49 (0.16)	3.68 (0.17)
<i>p</i> value		0.001	<0.001	<0.001
DASS—Stress				
Clinical (≤14)	6	0.75 (0.24)	0.68 (0.23)	0.66 (0.09)
Non-clinical (>14)	494	0.93 (0.14)	0.92 (0.08)	0.85 (0.13)
Z value (effect size)		4.33 (0.19)	4.73 (0.21)	5.11 (0.23)
<i>p</i> value		<0.001	<0.001	<0.001
Chronic conditions				
Yes	189	0.86 (0.18)	0.9 (0.11)	0.84 (0.13)
No	311	0.96 (0.08)	0.93 (0.06)	0.85 (0.13)
Z value (effect size)		7.19 (0.32)	2.13 (0.1)	0.83 (0.04)
<i>p</i> value		<0.001	0.03	0.41
Cognitive ability				
Satisfied	409	0.94 (0.12)	0.92 (0.09)	0.86 (0.13)
Not satisfied	91	0.87 (0.18)	0.9 (0.09)	0.82 (0.14)
Z value (effect size)		2.67 (0.12)	3.59 (0.16)	2.21 (0.1)
<i>p</i> value		0.007	<0.001	0.03
Life satisfaction				
Satisfied	337	0.94 (0.11)	0.93 (0.08)	0.87 (0.11)
Not satisfied	163	0.89 (0.17)	0.9 (0.1)	0.81 (0.15)
Z value (effect size)		3.49 (0.16)	4.44 (0.2)	4.02 (0.18)
<i>p</i> value		<0.001	<0.001	<0.001
Social relationship				
Satisfied	365	0.94 (0.12)	0.92 (0.08)	0.86 (0.12)
Not satisfied	135	0.88 (0.17)	0.9 (0.09)	0.83 (0.14)
Z value (effect size)		2.97 (0.13)	2.45(0.11)	2.32 (0.1)
<i>p</i> value		0.003	0.01	0.02
Government allowance				
Receiver	177	0.84 (0.19)	0.9 (0.12)	0.84 (0.14)
Non-receiver	323	0.97 (0.07)	0.93 (0.06)	0.86 (0.12)
Z value (effect size)		9.61 (0.43)	2.03 (0.1)	1.49 (0.07)
<i>p</i> value		<0.001	0.04	0.14
Living status				
Living alone	52	0.79 (0.22)	0.88 (0.11)	0.82 (0.15)
Living with families	448	0.94 (0.12)	0.92 (0.08)	0.85 (0.13)
Z value (effect size)		6.26 (0.28)	3.23 (0.14)	1.34 (0.06)
<i>p</i> value		<0.001	0.001	0.18

Table 4 (continued)

	<i>n</i>	EQ-5D-5L	ReQoL-UI	ICECAP-A
Employment				
Employed	139	0.98 (0.05)	0.93 (0.05)	0.87 (0.13)
Non-employed	361	0.93 (0.12)	0.93 (0.07)	0.85 (0.12)
Z value (effect size)		4.31 (0.24)	- 0.55 (- 0.03)	1.65 (0.09)
<i>p</i> value		<0.001	0.58	0.09
Income per month				
HKD ≤ 5000	313	0.89 (0.16)	0.9 (0.1)	0.84 (0.13)
HKD 5001–20,000	93	0.97 (0.07)	0.91 (0.08)	0.85 (0.12)
HKD ≥ 20,001	45	0.98 (0.08)	0.95 (0.06)	0.9 (0.12)
KW H value (effect size)		30.97 (0.06)	9.7 (0.02)	10.19 (0.02)
<i>p</i> value		<0.001	0.007	0.006

Kruskal–Wallis H effect size: “weak” 0.01–0.059, “moderate” 0.06–0.139, “strong” > 0.14; Mann–Whitney *U* effect size: “weak” 0.11–0.30, “moderate” 0.31–0.50, “strong” > 0.5

Bold figures indicated moderate or above effect size

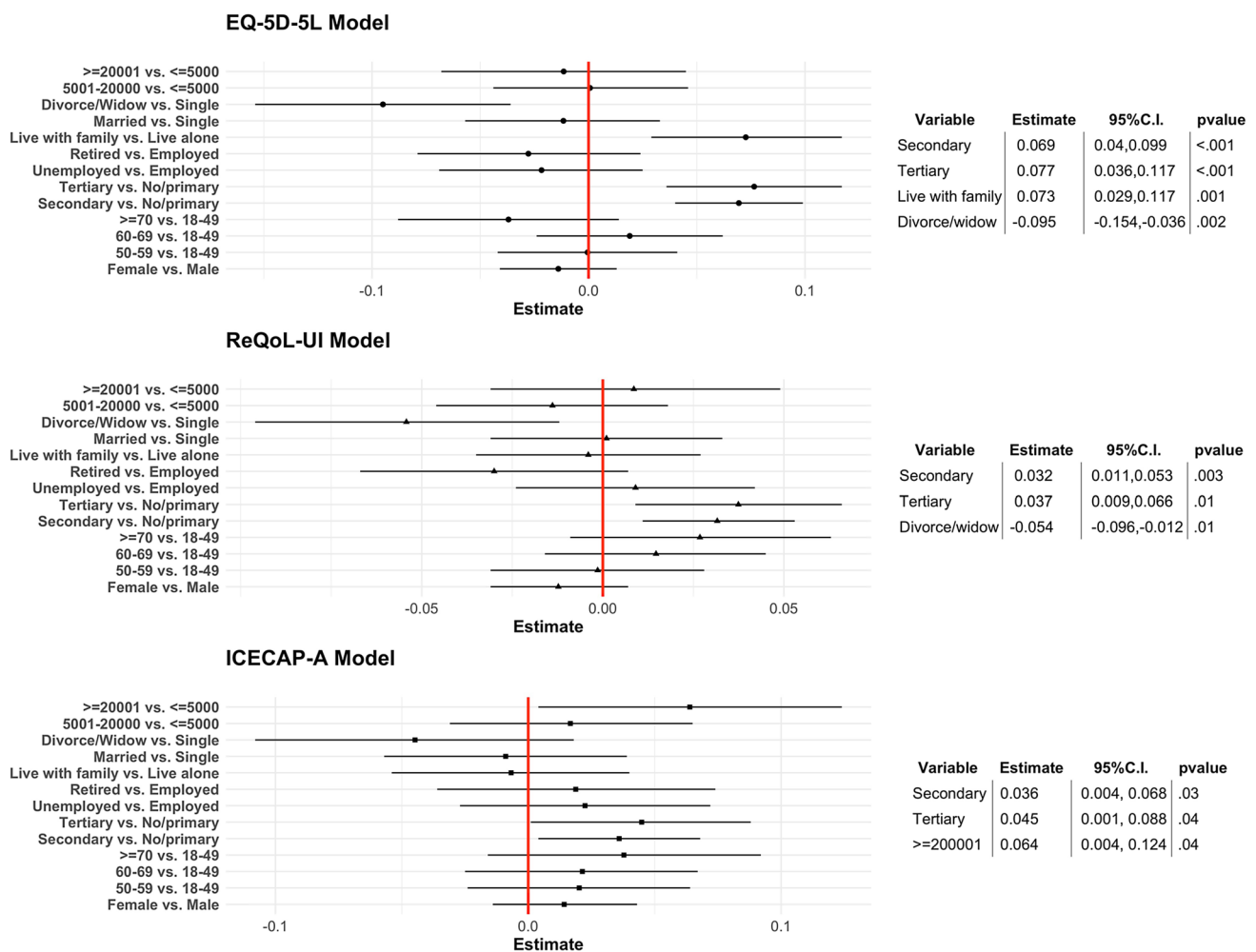


Fig. 2 Multiple regression models of the EQ-5D-5L, ReQoL-UI and ICECAP-A utility scores and selected sociodemographic characteristics

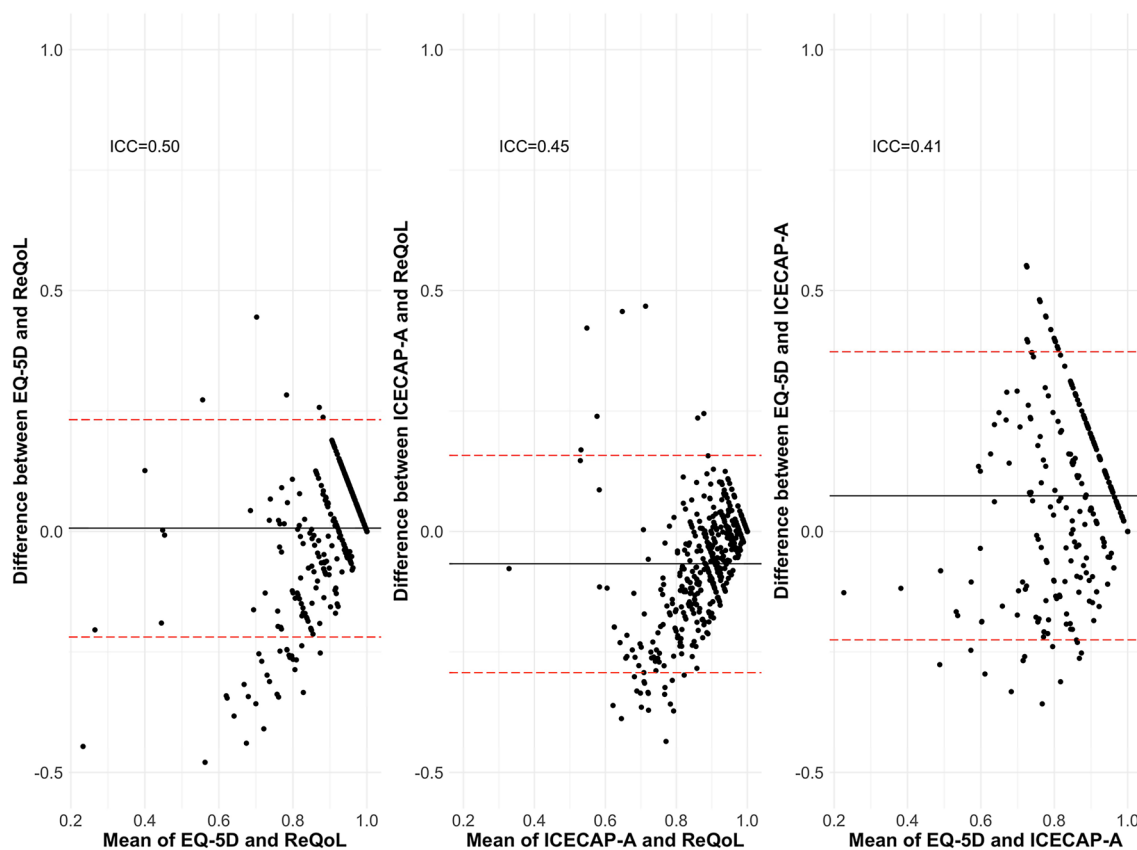


Fig. 3 Bland–Altman plot of the EQ-5D-5L, ReQoL-UI and ICECAP-A

that ICECAP-A should outperform the other measures. One possible explanation is the utility scores of the ReQoL-UI and ICECAP-A were calculated based on the UK people's preference weights. Given both mental HRQoL and well-being are subjectively concepts [42], using the UK population's preference may undermine measures' validity in Chinese population. Further, there is evidence that the country weights impact on GPBMs' utility values [43, 44]. The development of local value set for the ICECAP-A and ReQoL-UI is, therefore, recommended. Furthermore, regression analysis exhibited that educational attainment is an important predictor affecting people's HRQoL and well-being, which was in line with previous findings [41, 45–47].

Although our preliminary results showed that all three measures exhibited satisfactory psychometric properties, the choice of measures depends on the concept that the study intends to measure. For instance, if measuring physical health is the focus of an intervention, then EQ-5D-5L, which covers most aspects of physical health of HRQoL, is preferred. However, if the objective of the intervention is to measure the impact of treatments on with a focus to improve mental health or well-being [14, 48], the other two measures may be appropriate given their conceptualization and constructs, though their performance in HK population needs further exploration.

A strength of this study is it directly compared the performance of three GPBMs in a same sample of HK general population, supporting the generalizability of our findings to conduct the economic evaluations for all HK population. Moreover, the data collected during the COVID-19 pandemic may increase the sensitivity of these measures to detect people's mental HRQoL and well-being, facilitating the assessment of reliability and validity of the ICECAP-A and ReQoL-UI. However, several limitations need to be addressed. First, utility scores of the ICECAP-A and ReQoL-UI were calculated using the UK preference weights, which may generate bias in assessing the validity of those two measures. Another limitation was our sample was not representative of the HK general population in terms of age as it was hard to recruit younger people through the telephone survey. Last, despite, in this study, the utility score of three measures ranged between 0 and 1, which indicated the direct comparison between them is reasonable, the lower limit of the EQ-5D-5L and ReQoL-UI utility score can be smaller than 0, which may raise some methodological issues in explaining outcomes when using different measures. This issue should be further explored.

Conclusions

This study confirmed that the EQ-5D-5L, ICECAP-A and ReQoL-UI performed psychometrically well in this sample of HK general population though the agreement between them was poor. Considering their distinct theoretical structures, the selection of measure to facilitate the economic evaluation depends on the nature and objective of the intervention. Using the EQ-5D-5L to measure health benefits from a mental health intervention may fail to capture its benefits leading to misallocation of resources to mental health services. Additionally, studies are needed to further investigate the psychometric performance of the ICECAP-A and ReQoL-UI using preferences elicited from the HK general population.

Author contributions RHX: study concept and design; data analysis and interpretation; software; writing-original draft; writing-review and editing. ADK: study concept and design; data analysis and interpretation; writing-review and editing. LLW: software; visualization; writing-review and editing. AWC: provision of study materials or patients; collection and assembly of data; writing-review and editing. ELW: study concept and design; provision of study materials or patients; collection and assembly of data; supervision; writing-review and editing.

Funding None.

Data availability The data that support the findings of this study are available from the corresponding author upon reasonable request.

Code availability Not applicable.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

Ethical approval All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki Declaration and its later amendments or comparable ethical standards. The institutional review board of the Chinese University of Hong Kong approved the study protocol and informed consent (Ref no.: SBRE-18-671).

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

Consent to publish Not applicable.

References

1. Roberts, S.L.E., Healey, A., Sevdalis, N.: Use of health economic evaluation in the implementation and improvement science fields—a systematic literature review. *Implement Sci* **14**, 72 (2019)
2. Rabin, R., de Charro, F.: EQ-5D: a measure of health status from the EuroQol Group. *Ann Med* **33**, 337–343 (2001). <https://doi.org/10.3109/07853890109002087>
3. Brazier, J., Connell, J., Papaioannou, D., et al.: 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 (Rockv)* (2014). <https://doi.org/10.3310/hta18340>
4. Lin, Z., Matteson, M., Li, X., et al.: Older adults' eHealth literacy and the role libraries can play. *J Librariansh Inf Sci* (2020). <https://doi.org/10.1177/0961000620962847>
5. Beisani, M., Vilallonga, R., Petrola, C., et al.: Effects of COVID-19 lockdown on a bariatric surgery waiting list cohort and its influence in surgical risk perception. *Langenbeck's Arch Surg* **406**(20), 393–400 (2020)
6. Lim, S.L., Woo, K.L., Lim, E., et al.: Impact of COVID-19 on health-related quality of life in patients with cardiovascular disease: a multi-ethnic Asian study. *Health Qual Life Outcomes* **18**, 387 (2020)
7. Kapilashrami, A., Bhui, K.: Mental health and COVID-19: Is the virus racist? *Br J psychiatry* **217**, 405–407 (2020)
8. Ornell, F., Schuch, J.B., Sordi, A.O., et al.: 'Pandemic fear' and COVID-19: mental health burden and strategies. *Rev Bras Psiquiatr* **42**, 232–235 (2020)
9. Parihar, A.: Doing the day's work well: my unlikely COVID-19 renaissance. *Br J Gen Pract* **70**, 344 (2020)
10. Alradhawi, M., Shubber, N., Sheppard, J., et al.: Effects of the COVID-19 pandemic on mental well-being amongst individuals in society- A letter to the editor on "The socio-economic implications of the coronavirus and COVID-19 pandemic: a review." *Int J Surg* **78**, 147–148 (2020)
11. Pfefferbaum, B., North, C.S.: Mental health and the Covid-19 pandemic. *N Engl J Med* **383**, 510–512 (2020)
12. Saarni, S.I., Viertiö, S., Perälä, J., et al.: Quality of life of people with schizophrenia, bipolar disorder and other psychotic disorders. *Br J psychiatry* **197**, 386–394 (2010)
13. Brazier, J.: Is the EQ-5D fit for purpose in mental health? *Br J psychiatry* **197**, 348 (2010). <https://doi.org/10.1192/bjp.bp.110.082453>
14. Keetharuth, A.D., Brazier, J., Connell, J., et al.: Recovering Quality of Life (ReQoL): a new generic self-reported outcome measure for use with people experiencing mental health difficulties. *Br J Psychiatry* **212**, 42–49 (2018). <https://doi.org/10.1192/bjp.2017.10>
15. Keetharuth, A.D., Rowen, D., Bjorner, J.B., et al.: Estimating a preference-based index for mental health from the recovering quality of life measure: valuation of recovering quality of life utility index. *Value Heal* (2020). <https://doi.org/10.1016/j.jval.2020.10.012>
16. Al-Janabi, H., Flynn, T., Coast, J.: Development of a self-report measure of capability wellbeing for adults: the ICECAP-A. *Qual Life Res* **21**, 167–176 (2012). <https://doi.org/10.1007/s11136-011-9927-2>
17. Goranitis, I., Coast, J., Day, E., et al.: Measuring health and broader well-being benefits in the context of opiate dependence: the psychometric performance of the ICECAP-A and the EQ-5D-5L. *Value Heal* **19**, 820–828 (2016)
18. Tang, C., Xiong, Y., Wu, H., et al.: Adaptation and assessments of the Chinese version of the ICECAP-A measurement. *Health Qual Life Outcomes* **16**, 11–45 (2018). <https://doi.org/10.1186/s12955-018-0865-3>
19. Finch, A.P., Brazier, J.E., Mukuria, C.: What is the evidence for the performance of generic preference-based measures? A systematic overview of reviews. *Eur J Heal Econ* **19**, 557–570 (2017)

20. DeVellis, R.F.: Scale development : theory and applications, 4th edn. SAGE, Los Angeles (2017)
21. Herdman, M., Gudex, C., Lloyd, A., et al.: Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L). *Qual Life Res* **20**, 1727–1736 (2011). <https://doi.org/10.1007/s11136-011-9903-x>
22. Wong, E.L.Y., Ramos-Goñi, J.M., Cheung, A.W.L., et al.: Assessing the use of a feedback module to model EQ-5D-5L health states values in Hong Kong. *Patient* **11**, 235–247 (2018). <https://doi.org/10.1007/s40271-017-0278-0>
23. Xu, R.H., Keetharuth, A.D., Wang, L., et al.: Psychometric evaluation of the Chinese Recovering Quality of Life (ReQoL) outcome measure and assessment of health-related quality of life during the COVID-19 pandemic. *Front Psychol* (2021). <https://doi.org/10.3389/fpsyg.2021.663035>
24. Flynn, T.N., Huynh, E., Peters, T.J., et al.: Scoring the ICECAP-A capability instrument. Estimation of a UK general population tariff. *Health Econ* **24**, 258–269 (2015). <https://doi.org/10.1002/hec.3014>
25. Kroenke, K., Spitzer, R.L., Williams, J.B.W., et al.: Anxiety disorders in primary care: prevalence, impairment, comorbidity, and detection. *Ann Intern Med* **146**, 317 (2007). <https://doi.org/10.7326/0003-4819-146-5-200703060-00004>
26. Tong, X., An, D., McGonigal, A., et al.: Validation of the Generalized Anxiety Disorder-7 (GAD-7) among Chinese people with epilepsy. *Epilepsy Res* **120**, 31–36 (2015)
27. Lovibond, S.H.: Manual for the depression anxiety stress scales, 2nd edn. Psychology Foundation of Australia, Sydney, NSW (1995)
28. Gong, X., Xie, X., Xu, R., et al.: Psychometric properties of the Chinese versions of DASS-21 in Chinese college students. *Chinese J Clin Psychol* **18**, 443–446 (2010)
29. R Core Team. A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. 2013. <https://www.r-project.org/>
30. Xu, R.H., Zhou, L.M., Wong, E.L., et al.: Psychometric evaluation of the Chinese version of the decision regret scale. *Front Psychol* **11**, 3101 (2020). <https://doi.org/10.3389/fpsyg.2020.583574>
31. Cohen, J.: Statistical power analysis for the behavior science, 2nd edn. Lawrence Erlbaum Associates, New York (1988)
32. Tomczak, M., Tomczak, E.: The need to report effect size estimates revisited. An overview of some recommended measures of effect size. *Trends Sport Sci* **1**, 19–25 (2014)
33. Fritz, C.O., Morris, P.E., Richler, J.J.: Effect size estimates: current use, calculations, and interpretation. *J Exp Psychol* **141**, 2–18 (2012)
34. Cheung, P.W.H., Wong, C.K.H., Cheung, J.P.Y.: Differential Psychometric Properties of EuroQoL 5-Dimension 5-Level and Short-Form 6-Dimension Utility Measures in Low Back Pain. *Spine (Phila Pa 1976)* **44**, E679–E686 (2019). <https://doi.org/10.1097/BRS.0000000000002939>
35. Wong, E., Xu, R., Cheung, A.: Health-related quality of life in elderly people with hypertension and the estimation of minimally important difference using EQ-5D-5L in Hong Kong SAR, China. *Eur J Heal Econ* **21**, 869–879 (2020)
36. Wong, E.L.-Y., Cheung, A.W.-L., Wong, A.Y.-K., et al.: Normative profile of health-related quality of life for Hong Kong general population using preference-based instrument EQ-5D-5L. *Value Heal* **22**, 916–924 (2019). <https://doi.org/10.1016/j.jval.2019.02.014>
37. Wong, E., Xu, R., Cheung, A.: Measurement of health-related quality of life in patients with diabetes mellitus using EQ-5D-5L in Hong Kong, China. *Qual life Res* **29**, 1913–1921 (2020). <https://doi.org/10.1186/1477-7525-8-18>
38. Baji, P., Farkas, M., Dobos, Á., et al.: Capability of well-being: validation of the Hungarian version of the ICECAP-A and ICECAP-O questionnaires and population normative data. *Qual life Res* **29**, 2863–2874 (2020)
39. Linton, M.-J., Mitchell, P.M., Al-Janabi, H., et al.: Comparing the German translation of the ICECAP-A capability wellbeing measure to the original English version: psychometric properties across healthy samples and seven health condition groups. *Appl Res Qual Life* **15**, 651–673 (2018)
40. Wong EL-Y, Ho K-F, Wong SY-S et al (2020) Views on workplace policies and its impact on health-related quality of life during coronavirus disease (COVID-19) pandemic: cross-sectional survey of employees. *Int J Heal policy Manag* <https://doi.org/10.34172/ijhpm.2020.127>
41. Xu, R., Cheung, A., Wong, E.: Examining the health-related quality of life using EQ-5D-5L in patients with four kinds of chronic diseases from specialist outpatient clinics in Hong Kong SAR, China. *Patient Prefer Adherence* **11**, 1565–1572 (2017). <https://doi.org/10.2147/PPA.S143944>
42. Jokisaari, M.: Regret appraisals, age, and subjective well-being. *J Res Pers* **37**, 487–503 (2003). [https://doi.org/10.1016/S0092-6566\(03\)00033-3](https://doi.org/10.1016/S0092-6566(03)00033-3)
43. Rowen, D., Azzabi Zouraq, I., Chevrou-Severac, H., et al.: International regulations and recommendations for utility data for health technology assessment. *Pharmacoeconomics* **35**, 11–19 (2017). <https://doi.org/10.1007/s40273-017-0544-y>
44. van Dongen, J.M., Jornada Ben, Á., Finch, A.P., et al.: Assessing the impact of EQ-5D country-specific value sets on cost-utility outcomes. *Med Care* **59**, 82–90 (2021)
45. Pati, S., Swain, S., Knottnerus, J.A., et al.: Health related quality of life in multimorbidity: a primary-care based study from Odisha, India. *Health Qual Life Outcomes* (2019). <https://doi.org/10.1186/s12955-019-1180-3>
46. Klein, J., Hofreuter-Gätgens, K., Lüdecke, D., et al.: Socioeconomic status and health-related quality of life among patients with prostate cancer 6 months after radical prostatectomy: a longitudinal analysis. *BMJ Open* **6**, e010968–e010968 (2016)
47. Xu, R.H., Wong, E.L., Jin, J., et al.: Health-related quality of life measured using EQ-5D in patients with lymphomas. *Support Care Cancer* (2020). <https://doi.org/10.1007/s00520-020-05774-6>
48. Canaway, A., Al-Janabi, H., Kinghorn, P., et al.: Development of a measure (ICECAP-Close Person Measure) through qualitative methods to capture the benefits of end-of-life care to those close to the dying for use in economic evaluation. *Palliat Med* **31**, 53–62 (2017). <https://doi.org/10.1177/0269216316650616>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.