

Comparison of the psychometric properties of the EQ-5D-3L and SF-6D in the general population of Chengdu city in China

Longchao Zhao, PhD^a, Xiang Liu, PhD^a, Danping Liu, PhD^a, Yan He, PhD^b, Zhijun Liu, PhD^c, Ningxiu Li, MD^{a,*}

Abstract

The EQ-5D-3L and SF-6D are the most commonly used economic evaluation instruments. Data comparing the psychometric properties of the instruments are scarce in the Chinese population. This study compared the psychometric properties of these measures in the Chinese general population in Chengdu.

From October to December 2012, 2186 respondents (age ≥ 18) were selected from urban and rural areas of Chengdu, China, via multistage stratified cluster sampling. Correlations, scatter plots and Bland-Altman plots were used to explore the relationships between the 2 measures. Ceiling and floor effects were used to analyze the score distribution. The known-groups method was used to evaluate discriminant validity.

Among 2186 respondents, 2182 completed the questionnaire, and 2178 (18–82 years old, mean 46.09 ± 17.49) met the data quality requirement. The mean scores for the EQ-5D-3L_{CN}, EQ-5D-3L_{UK}, and SF-6D_{UK} were 0.95 (Std: 0.11), 0.93 (Std: 0.15), and 0.79 (Std: 0.12), respectively. The correlations between domains ranged from 0.16 to 0.51. The correlation between the EQ-5D-3L_{CN} and SF-6D_{UK} and between the EQ-5D-3L_{UK} and SF-6D_{UK} was 0.46. The scatter plots and Bland-Altman plots demonstrated poor agreement between the EQ-5D-3L and SF-6D. The floor and ceiling effects were respectively 0.05% and 74.60% for the EQ-5D-3L and 0.05% and 2.53% for the SF-6D_{UK}. The EQ-5D-3L_{CN}, EQ-5D-3L_{UK} and SF-6D have good discriminant validity in different sociodemographic and health condition groups. The SF-6D has higher level of discriminant validity in moderately healthy groups in the EQ-5D-3L full-health population.

Both the EQ-5D-3L and SF-6D are valid economic evaluation instruments in the Chinese general population in Chengdu but do not seem to be interchangeable. The EQ-5D-3L has a higher ceiling effect and higher level of discriminant validity among different sociodemographic groups, and the SF-6D has a lower ceiling effect and higher level of discriminant validity in health condition groups. Users may consider the evidence in the choice of these instruments.

Abbreviations: 95%CI = 95% confidence interval, AD = anxiety/depression, ANOVA = analysis of variance, CUA = cost-utility analysis, EQ-5D-3L = three-level EQ-5D, EQ-5D-3L_{CN} = EQ-5D-3L utility calculated by Chinese TTO preference value set, EQ-5D-3L_{UK} = EQ-5D-3L utility calculated by UK TTO preference value set, EQ-5D-5L = five-level EQ-5D-5L, ES = effect size, HRQOL = health-related quality of life, HUI = the Health Utility Index, MH = mental, MO = mobility, PA = pain, PD = pain/discomfort, PF = physical functioning, QALY = quality-adjusted life year, QOL = quality of life, RE = relative efficiency, RL = role limitations, SC = self-care, SDH = social determinant of health, SF = social functioning, SF-6D = Short-Form Six-Dimension, SF-6D_{UK} = SF-6D utility calculated by UK SG preference value set, SG = standard gamble, Std = standard deviation, TTO = time trade-off, UA = usual activities, VAS = visual analogue scale, VT = vitality.

Keywords: agreement, China, discriminant validity, EQ-5D-3L, general population, SF-6D

Editor: Chunxiao Li.

The research is funded by the Ministry of Education of the People's Republic of China. (Grant No. 20110181110038).

The authors have no conflicts of interest to disclose.

^a Department of Health Related Social and Behavioral Science, West China School of Public Health, Sichuan University, Chengdu, ^b Department of Social Medicine and Health Management, Public Health College, Zhengzhou University, Zhengzhou, ^c Applied Psychology Department, School of Management, Zunyi Medical University, Zunyi, China.

* Correspondence: Ningxiu Li, Department of Health Related Social and Behavioral Science, West China School of Public Health, Sichuan University, No. 16 of South Renmin Road, Wuhou District, Chengdu 610041, China (e-mail: liningxiu@163.com).

Copyright © 2019 the Author(s). Published by Wolters Kluwer Health, Inc. This is an open access article distributed under the terms of the Creative Commons Attribution-Non Commercial License 4.0 (CCBY-NC), where it is permissible to download, share, remix, transform, and buildup the work provided it is properly cited. The work cannot be used commercially without permission from the journal.

Medicine (2019) 98:11(e14719)

Received: 12 April 2018 / Received in final form: 1 February 2019 / Accepted: 6 February 2019

<http://dx.doi.org/10.1097/MD.0000000000014719>

1. Introduction

As the pressure to contain the costs of medical care escalates, there is an increasing use of cost-utility analysis (CUA) to perform economic evaluation. CUA allows decision makers to compare the economic value of different health care interventions.^[1] In CUA, the quality-adjusted life year (QALY) is the widely applied health indicator, which combines the attributes of the length and quality of life (QOL) into a single health utility, whereby 1.0 corresponds to full health and 0.0 corresponds to death.^[2]

Health utility can be estimated by 2 methods:

1. direct preference elicitation,
2. preference-based health state classification systems.^[3]

Direct methods, such as standard gamble (SG), time trade-off (TTO), and visual analogue scale (VAS), are time-consuming and resource-intensive in calculating health utility. In contrast, the preference-based health state classification systems are increasingly used in CUA and are more convenient in assessing health utility.^[4] These instruments can define the health state based on a health status classification system and then assign a utility score to each health state by using a scoring algorithm that

incorporates population preferences.^[5] Several general health-related quality of life (HRQOL) instruments have been developed to estimate health utility. For example, the Health Utilities Index (HUI),^[6,7] three-level EQ-5D (EQ-5D-3L),^[8] five-level EQ-5D (EQ-5D-5L),^[9] and the Short-Form Six-Dimension (SF-6D)^[5,10] are widely used health utility index instruments.

Among the health utility index instruments, the EQ-5D and SF-6D are 2 of the most commonly used preference-based measurements in the world.^[11] The Chinese pharmaceutical economic research guide suggests that utility measures should use the country-specific value set.^[12] The EQ-5D and SF-6D utility value sets are developed by different methods in different countries and regions.^[5,13–15] The EQ-5D-3L value sets have been produced for many countries or regions using TTO,^[16] such as the UK,^[17] US,^[18] Australia,^[19] and Japan.^[20] Recently, there has been increased use of the EQ-5D in China after the preference-based EQ-5D-3L value sets for the mainland China population were developed by the TTO method.^[21] Previous studies of the EQ-5D-3L in China either use other countries' value sets or are restricted to use as an instrument to report HRQOL problems.^[22,23] In contrast, the SF-6D value sets were first developed using SG in the UK,^[5] Hong Kong,^[24] Japan,^[10] Portugal^[25] and Brazil.^[26] However, a preference-based value set for the mainland China population still has not been developed. The EQ-5D-3L was recommended for use in health technology assessment by the China Guidelines for Pharmacy Economic Evaluations in 2015.^[12] Derived from the SF-36 and SF-12, the SF-6D is also a widely used instrument in economic evaluations,^[5,15,27] and previous studies have validated the SF-6D in several population groups.^[24,28–30] However, the application of the SF-6D in mainland China is limited. To date, several studies have compared the EQ-5D and SF-6D in various general populations and patient groups and suggest that they are interchangeable in different target populations.^[31–39]

The CUA is one of the most important indicators used by decision makers in health technology assessment. Different instruments may lead to different economic evaluation outcomes, which may influence healthcare decisions.^[40–42] However, little is known about the performance of the EQ-5D and SF-6D in mainland China's general population. The aim of this study is to compare the performance of the EQ-5D-3L and SF-6D in the general population of Chengdu city in mainland China.

2. Methods

2.1. Study design

The survey was conducted in Chengdu, a city in southwestern China, from October to December 2012. A multistage stratified cluster sampling method was used to select respondents. Respondents were recruited if they were 18 years old and above. In the study, 5 districts (towns) were selected from urban areas (counties) according to economic level. Within each district or town, 5 communities or villages were selected according to the geographic location and economic level. Within each selected community or village, 60 households were randomly selected. Subsequently, in each household, all residents over 18 years old were chosen for the survey. A total of 2182 respondents were recruited, consented to participate, and completed questionnaires. All the respondents provided informal consent and were interviewed by trained interviewers using the standard questionnaire.

2.2. Instruments and measures

The questionnaire contained questions regarding demographics (age and sex), socioeconomic status (marriage, education, employment, annual household income and health insurance), and health status (emotions, chronic disease, recent health status and self-reported health status). The questionnaire also includes the Chinese versions of the EQ-5D-3L and SF-36v2.

The EQ-5D-3L was developed by the EuroQol Group and consists of 5 health dimensions (mobility, self-care, usual activities, pain/discomfort and anxiety/depression), and each dimension has 3 levels (no problems, some problems, and major problems).^[16] Thus, it can describe 243 (3^5) health statuses. Using the scoring algorithm, each health status can be assigned a utility score.

The SF-6D is a preference-based instrument derived from the SF-36 and SF-12.^[5,15] This study used the Chinese version of the SF-36v2 for data collection. The SF-6D consists of 6 dimensions, and each dimension has 4 to 6 levels: physical functioning (6 levels), role limitations (4 levels), social functioning (5 levels), pain (6 levels), mental health (5 levels), and vitality (5 levels). Thus, the instrument can describe 18,000 possible health statuses, which can also be assigned a utility score by using the population-based preference algorithm.

2.3. Statistical analysis

Currently, there is no SF-6D algorithm based on the preference value set for mainland China's population. Therefore, we use the UK population-based scoring algorithm to calculate the SF-6D utility, and the utility score ranges from 0.29 to 1.00.^[5] The EQ-5D-3L scoring algorithm for mainland China was recently developed by TTO.^[21] To compare utility scores calculated by the same population preference, we also used the UK scoring algorithm for the EQ-5D-3L.^[17] The EQ-5D-3L China TTO preference value ranged from -0.149 to 1.00.^[21] The EQ-5D-3L UK preference value ranged from -0.114 to 1.00. A utility score under 0 represents a health status that is considered worse than being dead. The utility of the EQ-5D-3L calculated by the China TTO value set was represented as EQ-5D-3L_{CN}, and the utility calculated by the UK TTO value set was represented as EQ-5D-3L_{UK}. The utility of the SF-6D calculated by the UK value set was represented as SF-6D_{UK}.

The respondents' demographic characteristics and item distributions were described in numbers and percentages of the sample size. Continuous variables, including utility scores, and EQ-VAS scores, are presented as the mean and standard deviation (Std).

Health status was measured by emotions, chronic conditions, visits to the doctor in the past 2 weeks, and self-reported health. The chronic conditions include diabetes, hyperlipidemia, hypertension, heart disease, stroke, respiratory disease, liver disease, gastrointestinal disease, bone and joint disease, and cancer. The chronic conditions were diagnosed by a doctor in a hospital or community health service center.

The Spearman correlation was used to evaluate the correlations between domains and index scores as follows: negligible <0.20 ; poor 0.20 to 0.30, moderate 0.31 to 0.50; and strong >0.50 .^[43] We also used a scatter plot and Bland-Altman plot to evaluate the relationship between the EQ-5D-3L and SF-6D utility scores. The level of agreement between the EQ-5D-3L and SF-6D was analyzed by the Bland-Altman plot,^[44,45] which is an informative analytic method that allows for the identification of the relationship between measurement error and the best estimate of the true value. The average of 2 measures was plotted on the x-axis, and the mean difference between the 2 measures was plotted on the y-axis to check the systematic error. Good agreement between the 2 measures would

indicate a mean difference close to 0 and 95% of the differences falling within 2 standard deviations of the mean difference.

The ceiling and floor effects were used to compare the sensitivity of the EQ-5D-3L and SF-6D. The ceiling effect was assessed by computing the percentage of respondents reporting no problems (11111 for the EQ-5D-3L and 11111 for the SF-6D). The floor effect was assessed by computing the percentage of respondents reporting worse levels (33333 for the EQ-5D-3L and 645655 for the SF-6D).

Discriminant validity was used to assess the instruments' ability to distinguish groups with different demographic characteristics and health statuses.^[46] In terms of social determinants of health (SDH),^[47] we categorized 3 levels of external indicators: demographic characteristics, family indicators, and health conditions. The first level includes age, sex, education, and employment. The second level includes annual household income per member, marriage, and place of residence. The third level includes the quality of life score (QOL score), self-reported health status, the number of chronic diseases, and doctor visits in the last 2 weeks, as follows:

1. We used the median QOL score (QOL score=80) as a cutoff point to divide the respondents into 2 groups. The QOL score was obtained by a self-reported QOL item that asked the respondents to rate their physical health, mental health, social relationships, and living environment on a scale of 0 (worst) to 100 (best);
2. The self-reported health status item asked the respondents to rate their overall level of health as excellent, very good, good, fair, or poor;
3. We categorized the respondents' chronic disease status into 3 subgroups: 0 = no chronic disease, 1= one chronic disease, 2+ = 2 or more diseases;
4. The respondents were asked whether they had visited a doctor in the past 2 weeks, and their responses were recorded as "yes" or "no". The known group validity was evaluated by *t* tests and analysis of variance (ANOVA).

The effect size (ES) was used to detect health differences.^[48] The ES calculated by the mean difference found in utility divided by the standard deviation of utility and Cohen's moderate ES of 0.2 to 0.5 was adopted as the minimally important difference (MID) in this study.^[49] We calculate the ES between each characteristic subgroup to estimate the discriminant validity of the index score. The relative efficiency (RE) was also used to evaluate whether 1 instrument is more efficient or sensitive than another or more likely to result in a statistically significant difference between groups of respondents known to differ.^[50] RE can be calculated by the F statistic ratio or the square of the *t* ratio

between 2 measurements: $RE > 1$ indicates that the comparator measure has greater discriminating power or responsiveness than the reference measure and vice versa.^[51,52]

P-values less than .05 were considered statistically significant. All statistical analyses were two-sided and performed using R software (version 3.4.2; R Foundation for Statistical Computing, Vienna, Austria).

3. Results

3.1. Sample demographic characteristics

A total of 2182 respondents were randomly selected by the multistage sampling method from the rural and urban areas of Chengdu, Sichuan Province, China, and 2178 respondents completed the questionnaires. The respondents' ages ranged from 18 to 82, and the mean age was 46.09 (Std 17.49). Female respondents comprised 55.28% of the sample size. Those who were married accounted for 76.26% of the sample. A total of 11.85% of the sample had graduated from university. A total of 30.9% of the respondents had chronic diseases, and 26.95% had experienced discomfort or consulted a doctor 2 weeks before the survey.

The mean of the EQ-5D-3L_{CN} score was 0.95 (Std: 0.11) (median 1.0; interquartile range 0.13), and that of the EQ-5D-3L_{UK} score was 0.93 (Std: 0.15) (median 1.0; interquartile range: 0.15). The mean of the SF-6D_{UK} score was 0.79 (Std: 0.12) (median: 0.81; interquartile range: 0.19).

3.2. Relationship between the EQ-5D-3L and SF-6D

Correlations between the EQ-5D-3L and SF-6D domains are presented in Table 1. The SF-6D domains have a higher correlation with related domains on the EQ-5D-3L. The correlations between domains were as follows: 0.32 between physical functioning and mobility, 0.21 between physical functioning and self-care, 0.26 between physical functioning and usual activities, 0.36 between physical functioning and pain/discomfort, 0.31 between role limitation and pain/discomfort, 0.25 between social functioning and anxiety/depression, 0.51 between pain and pain/discomfort, and 0.20 between mental health and anxiety/depression. The vitality domain of the SF-6D has no counterpart domain on the EQ-5D-3L and was moderately correlated with the EQ-5D-3L pain/discomfort domain ($r = 0.30$).

3.3. Level of agreement between utility scores

The Spearman correlation between the EQ-5D-3L_{CN} and SF-6D_{UK} (Fig. 1) and between the EQ-5D-3L_{UK} and SF-6D_{UK} (Fig. 2)

Table 1
Correlations among dimensions of the EQ-5D-3L and SF-6D (n=2,178).

SF-6D	EQ-5D-3L				
	MO	SC	UA	PD	AD
PF	0.32	0.21	0.26	0.36	<i>0.18</i>
RL	<i>0.19</i>	<i>0.12</i>	<i>0.16</i>	0.31	0.27
SF	<i>0.18</i>	<i>0.14</i>	<i>0.19</i>	0.25	0.24
PA	<i>0.27</i>	<i>0.17</i>	<i>0.20</i>	0.51	<i>0.26</i>
MH	<i>0.05</i>	<i>0.05</i>	<i>0.04</i>	<i>0.09</i>	0.20
VT	0.22	0.10	0.17	0.30	0.20

The bold indicate the theoretically related domains coefficients, and the italic and underline indicates the non-theoretical related the domains' coefficients.

AD = anxiety/depression, MH = mental, MO = mobility, PA = pain, PD = pain/discomfort, PF = physical functioning, RL = role limitations, SC = self-care, SF = social functioning, UA = usual activities, VT = vitality.

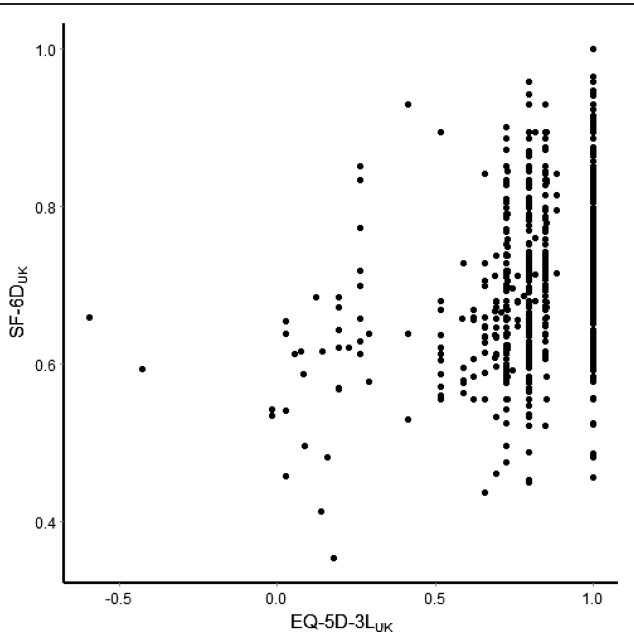


Figure 1. Scatter plot between the EQ-5D-3L_{UK} and SF-6D_{UK}.

was 0.46. A notable disagreement can be observed on both ends of the plot. The lowest EQ-5D-3L utility scores tended to have higher SF-6D utility scores. The highest scores on the EQ-5D-3L (EQ-5D-3L utility = 1.00) were associated with a very wide score range on the SF-6D_{UK}, from 0.46 to 1.00, which displays the high ceiling effect of the EQ-5D-3L.

The Bland-Altman plots show patterns similar to those of the scatter plots between the EQ-5D-3L and SF-6D utilities (Figs. 3 and 4). The mean difference between the EQ-5D-3L_{CN} and SF-6D_{UK} is 0.156, with a 95% limit of agreement of -0.067 to 0.378 . A total of 82 (3.76%) observations were out of the 95% limit of agreement. The mean difference between the EQ-5D-3L_{UK} and SF-6D_{UK} is 0.137, with 62 (2.85%) observations out of

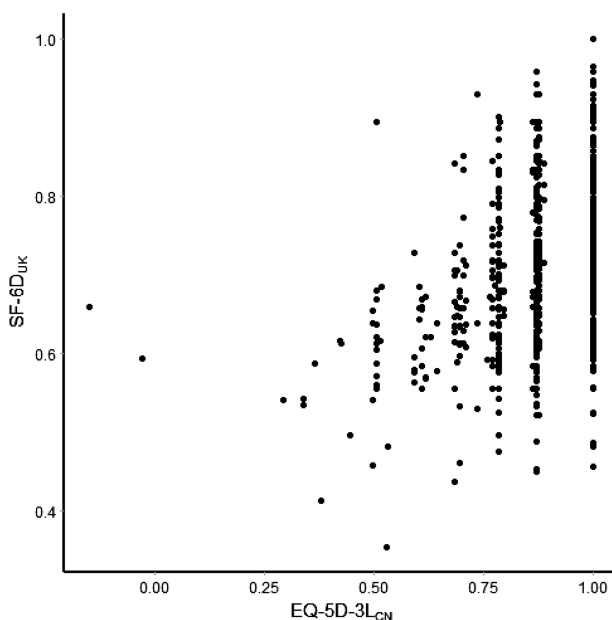


Figure 2. Scatter plot between the EQ-5D-3L_{CN} and SF-6D_{UK}.

the 95% limit of agreement of -0.139 to 0.414 . Bland-Altman plots indicate an acceptable agreement between 2 instruments. Notably, the figures also show a nonrandom mean difference between the EQ-5D-3L and SF-6D, and the EQ-5D-3L utilities were larger than the SF-6D utilities at the upper end and smaller at the lower end. The limit of agreement in the second figure is wider than that in the first, indicating that the difference in variation between the EQ-5D-3L_{UK} and SF-6D_{UK} was greater than that between the EQ-5D-3L_{CN} and SF-6D_{UK}.

3.4. Ceiling and floor effects

Tables 2 and 3 show the score distribution of both the EQ-5D and SF-6D dimensions. A large proportion of patients reported no problems in either the EQ-5D-3L or SF-6D dimensions except in the vitality dimension of the SF-6D. All domains of the EQ-5D-3L have higher ceiling effects (>80%) than those of the SF-6D. Floor effects can be negligible for 2 measurements' dimensions except for the role limitation domain of the SF-6D (21.03%).

The EQ-5D-3L_{CN} and EQ-5D-3L_{UK} tend to have very high ceiling effects ($n = 1625$, 74.60%) and a low floor effects (0.05%) (Fig. 5). The EQ-5D-3L utility scores are skewed toward high scores and are more skewed than the SF-6D_{UK} utility scores. The SF-6D_{UK} has low floor (0.05%) and ceiling effects ($n = 55$, 2.53%), and the distribution of the SF-6D_{UK} utilities is more normal than that of the EQ-5D-3L utilities.

3.5. Descriptive statistics and discriminant validity

Table 4 shows the discriminant ability of the EQ-5D-3L_{CN}, EQ-5D-3L_{UK} and SF-6D_{UK} in different sociodemographic groups. All of the utility scores were significantly different according to age, gender, marriage, education, employment and household incomes, but not according to health insurance. Specifically, respondents who were male, were younger, were more educated, were married, were employed, had higher household incomes, and had health insurance reported higher utility scores. In the adjacent sociodemographic characteristic groups, the effect sizes of 3 utilities show higher level of discriminant validity ($ES > 0.20$).

Table 5 shows the discrimination of the EQ-5D-3L_{CN}, EQ-5D-3L_{UK} and SF-6D_{UK} utility in different health groups. All the utility scores can discriminate the groups into the following different health indicator groups: self-reported health status, number of chronic diseases, outpatients in the recent 2 weeks, emotions and QOL score groups ($P < .01$). The utility scores were lower in the poor health groups than in the better health groups ($ES < 0$), except for the SF-6D_{UK} score in the emotion group of "better" and "as usual". The absolute value of $ES > 0.20$ indicates that the utility scores can discriminate among subgroups of health indicators. The RE shows that both the EQ-5D-3L_{CN} and EQ-5D-3L_{UK} were less discriminating than the SF-6D in self-reported health groups, outpatient groups, and QOL score groups ($RE < 1.00$) but not in chronic disease groups ($RE > 1.00$).

In each sociodemographic group and health group, both the EQ-5D-3L_{CN} and EQ-5D-3L_{UK} are greater than the SF-6D_{UK}, and the EQ-5D-3L_{CN} is greater than the EQ-5D-3L_{UK} (Tables 4 and 5). The result also shows that the standard deviation of the EQ-5D-3L_{CN} is lower than that of the EQ-5D-3L_{UK} and SF-6D_{UK}. Although the REs of the EQ-5D-3L_{CN} and EQ-5D-3L_{UK} were greater than 1.00 in single indicators of sociodemographic characteristics (age, gender, education, marriage status, employment, annual household income, chronic conditions, and

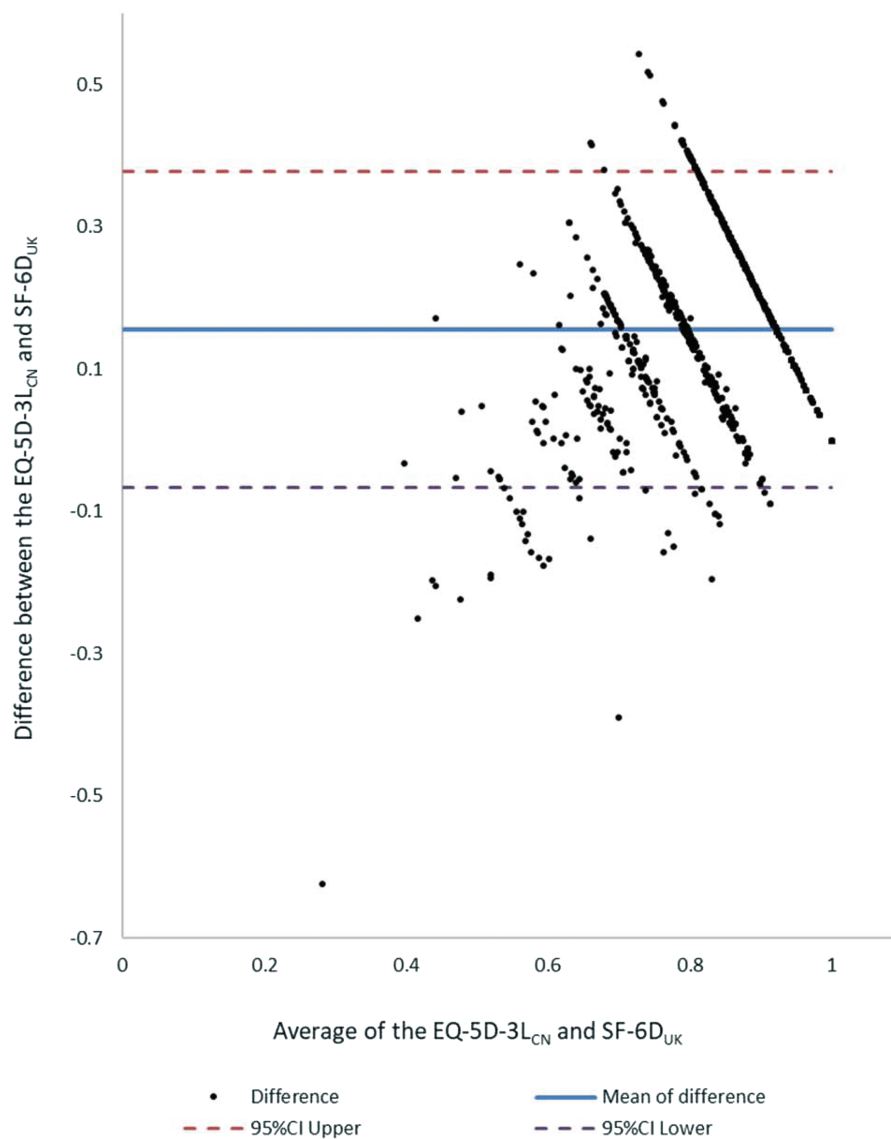


Figure 3. Bland-Altman plot between the EQ-5D-3L_{CN} and SF-6D_{UK}.

emotions), the SF-6D shows higher level of discriminant validity in comprehensive health status (self-reported health status and QOL score groups).

3.6. Discriminating among respondents with better health

Fifty five (2.53%) respondents reported the best health condition on the SF-6D (the SF-6D health status was “111111”). However, 1,625 (72.4%) respondents reported the best health condition on the EQ-5D-3L (the EQ-5D-3L health status was “11111”). The mean EQ-5D-3L for those with the best SF-6D health status (“111111”) was 1.00. Conversely, the SF-6D_{UK} utility shows a normal distribution on the EQ-5D-3L full-health respondents (“11111”), and the mean was 0.82 (Std: 0.10), with scores ranging from 0.46 to 1.00.

Table 6 shows the SF-6D discriminant validity in the EQ-5D-3L full-health groups. Among the EQ-5D-3L full-health respondents, the SF-6D_{UK} index scores were significantly different in subgroups by age, employment, annual household income, self-reported health, emotions, outpatients, and QOL

($P < .05$). Respondents with better self-reported health, better emotional status, no chronic disease and no outpatient status have a higher SF-6D_{UK} utility than people with poor or worse conditions. The effect sizes of the SF-6D_{UK} in these groups also show higher level of discriminant validity ($ES > 0.20$), except for in the education and household income groups ($ES < 0.20$).

4. Discussion

Evidence regarding the performance of the EQ-5D-3L and SF-6D in the Chinese general population was provided in this study. The results show that the 2 measurements demonstrated good discriminant validity in the general population. Both displayed high ceiling effects, the domains showed moderate correlations between theoretically related pairs, and the level of agreement between the 2 measurement utilities was poor. However, there are some notable differences between the EQ-5D-3L and SF-6D, which is consistent with the results in the general population and in patient groups.^[31–35,53–58]

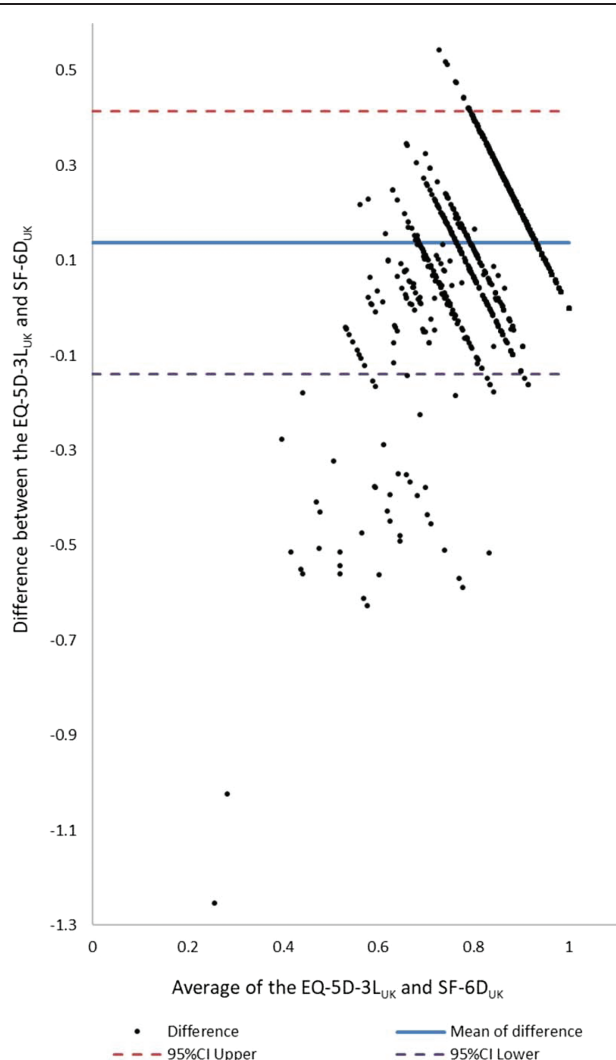


Figure 4. Bland-Altman plot between the EQ-5D-3L_{UK} and SF-6D_{UK}.

First, the scores on both the EQ-5D-3L_{CN} and EQ-5D-3L_{UK} are higher than those on the SF-6D_{UK} in the overall sample, which is consistent with previous studies.^[33,35] The absolute difference is 0.156 (Std: 0.113) ($P < .001$) between the EQ-5D-3L_{CN} and SF-6D_{UK} and 0.137 (Std: 0.141) ($P < .001$) between the EQ-5D-3L_{UK} and SF-6D_{UK}. Previous studies have suggested that there are several reasons for the discrepancy. The first reason is the method used to derive the value sets and scoring algorithms.^[59] The SF-6D_{UK} scoring algorithm was derived by the SG method, and the EQ-5D-3L_{CN} and EQ-5D-3L_{UK} scoring algorithms were derived by the TTO method. The SG method usually derives higher scores than the TTO method in patients with severe health states and lower scores than the TTO method in patients with mild health states.^[60–62] Second, the population resource may be another reason for the absolute difference. The UK population’s preferential value was set to calculate the EQ-5D-3L_{UK} and SF-6D_{UK} scores in the Chinese population. In this study, the EQ-5D-3L means exceeded the SF-6D mean across the whole sample, which was inconsistent with some studies.^[60] Furthermore, Whithurst et al^[63] used the same method to derive the EQ-5D-3L and SF-6D preference value set in the same population and found that the SF-6D is still lower than the EQ-5D-3L (mean difference 0.253). Thus, the mean discrepancy may result from characteristics of the EQ-5D-3L and SF-6D and not only from the method and population difference.^[31,63] The most possible explanation may be the high ceiling effect of the EQ-5D-3L. The EQ-5D-3L had a much higher ceiling effect (1625 full-health respondents, 74.6%) than the SF-6D (55 full-health respondents, 2.53%), as shown in Fig. 4. The high ceiling effect will elevate the mean score of the EQ-5D-3L in all samples, which is consistent with a study in the general population.^[31]

Second, the correlation between the EQ-5D-3L and SF-6D domains (0.20–0.51) and between utilities (0.46) was acceptable, but the scatter plot and the Bland-Altman plot revealed a lack of agreement between the EQ-5D-3L and SF-6D. The lowest EQ-5D-3L utility scores tend to have a high SF-6D score, and the highest EQ-5D-3L utility (1.00) tends to have a wide range of SF-6D utility (0.456–1.00). The high-end discrepancy between the EQ-5D-3L and SF-6D also revealed the high ceiling effect of the EQ-5D-3L.

Table 2

Frequency distribution of EQ-5D-3L scores by dimensions (%) (n=2,178).

Level	EQ-5D-3L				
	MO	SC	UA	PD	AD
1	2071 (95.08)	2142 (98.35)	2105 (96.65)	1756 (80.62)	1918 (88.06)
2	101 (4.64)	32 (1.47)	70 (3.21)	395 (18.14)	249 (11.43)
3	6 (0.28)	4 (0.18)	3 (0.14)	27 (1.24)	11 (0.51)

AD = anxiety/depression, MO = mobility, PD = pain/discomfort, SC = self-care, UA = usual activities.

Table 3

Frequency distribution of SF-6D scores by dimensions (%) (n=2,178).

Level	SF-6D					
	PF	RL	SF	PA	MH	VT
1	1163 (53.40)	1203 (55.23)	1234 (56.66)	1259 (57.81)	1251 (57.44)	165 (7.58)
2	428 (19.65)	270 (12.40)	562 (25.8)	181 (8.31)	581 (26.68)	123 (56.61)
3	330 (15.15)	247 (11.34)	274 (12.58)	466 (21.40)	258 (11.85)	450 (20.66)
4	197 (9.05)	458 (21.03)	95 (4.36)	191 (8.77)	76 (3.49)	177 (8.13)
5	48 (2.20)		13 (0.60)	68 (3.12)	12 (0.55)	153 (7.02)
6	12 (0.55)			13 (0.60)		

MH = mental, PA = pain, PF = physical functioning, RL = role limitations, SF = social functioning, VT = vitality.

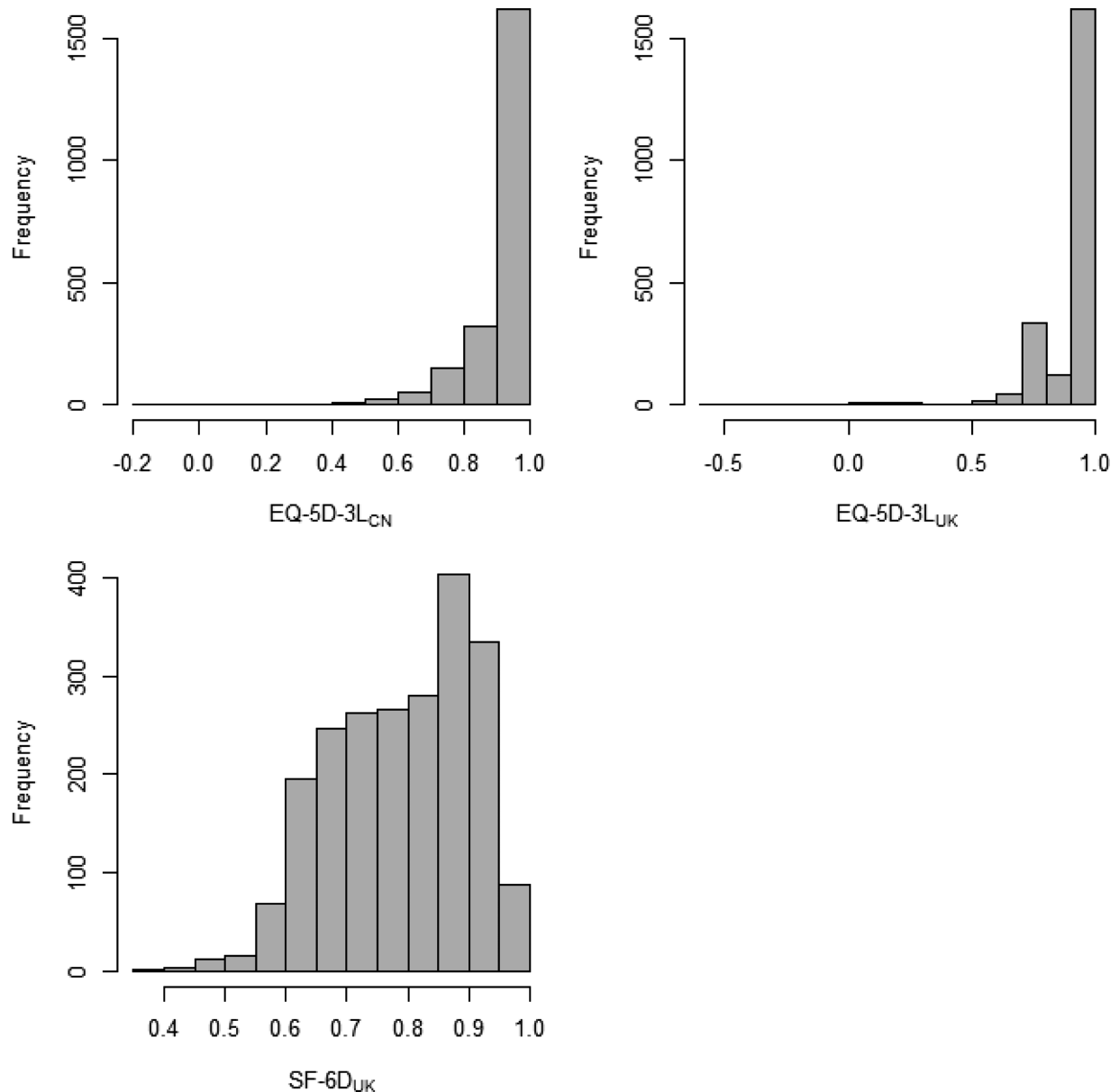


Figure 5. Health utility histogram of the EQ-5D-3L and SF-6D (n=2,178).

Third, the EQ-5D-3L_{CN}, EQ-5D-3L_{UK} and SF-6D_{UK} performed well in discriminating among different sociodemographic and health groups. All 3 utility scores were lower among the groups with poor health than among those with good health (Table 5). This result is consistent with previous studies.^[35,57,64] The EQ-5D-3L_{CN} and EQ-5D-3L_{UK} seem more sensitive than the SF-6D in discriminating among sociodemographic subgroups based on age, gender, marriage, education, employment, household income, and health insurance (RE > 1.00, Table 4). This higher consistency may be caused by the larger standard deviation of the SF-6D. The larger standard deviation will lead to smaller F statistic values in ANOVA and smaller REs than in the EQ-5D-3L_{CN} and EQ-5D-3L_{UK}. Nevertheless, the results still show that the SF-6D_{UK} significantly discriminates among all sociodemographic groups, including those based on age, gender, marriage, education, employment, household income, and health insurance ($P < .01$, Table 4). Furthermore, the

SF-6D_{UK} is more sensitive than the EQ-5D-3L in detecting smaller health differences. Among EQ-5D-3L full-health respondents, there are about 28.94% of respondents who self-reported their health as “good” and 20% of respondents with chronic conditions whom the EQ-5D-3L failed to discriminate with regard to health differences. The SF-6D can discriminate among different health differences on the ceiling of the EQ-5D-3L ($P < .01$, Table 6), although the results are inconsistent with those of the US population.^[65]

Finally, the EQ-5D-3L_{UK} and SF-6D_{UK} utility scores calculated by the UK population preference value set and algorithm tended to have higher standard deviations, and the scores on the EQ-5D-3L_{UK} are lower than those on the EQ-5D-3L_{CN} in each sociodemographic group (Tables 4 and 5). These differences may be caused by the population source of the value set not representing the Chinese population. Values for health status may vary across countries because country-specific value sets are

Table 4

Discriminant validity of the EQ-5D-3L and SF-6D in different demographic populations.

Social demographics	N (%)	EQ-5D-3L _{CN}		EQ-5D-3L _{UK}		SF-6D _{UK}	
		Mean (Std)	ES	Mean (Std)	ES	Mean (Std)	ES
Entire sample	2,178	0.95 (0.11)		0.93 (0.15)		0.79 (0.12)	
Age group (years)							
18–24	320 (14.69)	<u>.99 (.04)</u>		<u>.98 (.06)</u>		<u>.81 (.10)</u>	
25–34	352 (16.16)	<u>.98 (.07)</u>	−0.18	<u>.97 (.09)</u>	−0.19	<u>.81 (.11)</u>	0.02
35–44	351 (16.12)	<u>.96 (.11)</u>	−0.28	<u>.94 (.16)</u>	−0.33	<u>.81 (.11)</u>	0.00
45–54	355 (16.30)	<u>.96 (.08)</u>	0.01	<u>.94 (.12)</u>	0.00	<u>.81 (.11)</u>	0.03
55–64	435 (19.97)	<u>.92 (.11)</u>	−0.46	<u>.89 (.17)</u>	−0.38	<u>.77 (.12)</u>	−0.36
>=65	365 (16.76)	<u>.90 (.15)</u>	−0.19	<u>.87 (.20)</u>	−0.12	<u>.75 (.12)</u>	−0.20
<i>F</i>		36.32**		31.39**		20.29**	
<i>RE</i>		1.79		1.55		1.00	
Gender							
Male	974 (44.72)	<u>.96 (.10)</u>		<u>.94 (.14)</u>		<u>.80 (.11)</u>	
Female	1204 (55.28)	<u>.94 (.11)</u>	−0.15	<u>.92 (.16)</u>	−0.15	<u>.79 (.12)</u>	−0.13
<i>t</i>		3.21**		3.28**		3.08**	
<i>RE</i>		1.09		1.13		1.00	
Marital status							
Single	358 (16.44)	<u>.98 (.06)</u>		<u>.98 (.09)</u>		<u>.81 (.10)</u>	
Married	1661 (76.26)	<u>.95 (.11)</u>	−0.65	<u>.93 (.15)</u>	−0.55	<u>.79 (.12)</u>	−0.13
Widowed	126 (5.78)	<u>.94 (.11)</u>	−0.01	<u>.93 (.12)</u>	0.05	<u>.78 (.11)</u>	−0.15
Divorced	33 (1.52)	<u>.90 (.14)</u>	−0.43	<u>.87 (.21)</u>	−0.59	<u>.75 (.12)</u>	−0.26
<i>F</i>		22.86**		20.29**		8.94**	
<i>RE</i>		2.56		2.27		1.00	
Education							
Elementary school or illiterate	678 (31.13)	<u>.92 (.14)</u>		<u>.89 (.19)</u>		<u>.77 (.12)</u>	
Middle school	604 (27.73)	<u>.95 (.09)</u>	0.27	<u>.94 (.14)</u>	0.25	<u>.81 (.11)</u>	0.30
High school	402 (18.46)	<u>.97 (.08)</u>	0.12	<u>.96 (.10)</u>	0.14	<u>.80 (.11)</u>	−0.05
Technical college	236 (10.84)	<u>.97 (.08)</u>	0.06	<u>.96 (.13)</u>	0.02	<u>.81 (.11)</u>	0.05
University or graduate education	258 (11.85)	<u>.98 (.07)</u>	0.06	<u>.97 (.10)</u>	0.08	<u>.81 (.11)</u>	0.00
<i>F</i>		25.08**		23.03**		11.40**	
<i>RE</i>		2.20		2.02		1.00	
Employment							
Employed	1378 (63.27)	<u>.96 (.08)</u>		<u>.95 (.13)</u>		<u>.81 (.11)</u>	
Retired	666 (30.58)	<u>.92 (.13)</u>	−0.50	<u>.90 (.17)</u>	−0.42	<u>.77 (.12)</u>	−0.33
Other	134 (6.15)	<u>.94 (.15)</u>	0.11	<u>.91 (.20)</u>	0.10	<u>.80 (.13)</u>	0.28
<i>F</i>		38.15**		29.48**		23.84**	
<i>RE</i>		1.60		1.24		1.00	
Residence region							
Urban	1034 (47.47)	0.95 (0.11)		0.93 (0.14)		0.80 (0.11)	
Rural	1144 (52.53)	0.95 (0.10)	0.06	0.93 (0.15)	0.02	0.79 (0.12)	−0.03
<i>t</i>		1.30		0.42		0.75	
<i>RE</i>		3.00		0.31		1.00	
Annual household income (in 1000-yuan increments)							
<=15	331 (15.27)	<u>.91 (.15)</u>		<u>.88 (.21)</u>		<u>.77 (.12)</u>	
15–30	540 (24.91)	<u>.95 (.10)</u>	0.23	<u>.93 (.13)</u>	0.24	<u>.79 (.12)</u>	0.15
30–50	526 (24.26)	<u>.96 (.09)</u>	0.10	<u>.94 (.13)</u>	0.08	<u>.80 (.11)</u>	0.12
50–100	522 (24.08)	<u>.96 (.09)</u>	0.01	<u>.94 (.15)</u>	−0.02	<u>.80 (.11)</u>	0.01
>100	249 (11.49)	<u>.97 (.09)</u>	0.11	<u>.96 (.12)</u>	0.11	<u>.82 (.11)</u>	0.15
<i>F</i>		14.09**		12.27**		8.77**	
<i>RE</i>		1.61		1.40		1.00	
Health insurance							
Yes	2001 (91.87)	0.96 (0.10)		0.94 (0.13)		0.81 (0.11)	
No	177 (8.13)	0.95 (0.11)	−0.08	0.93 (0.15)	−0.10	0.79 (0.12)	−0.12
<i>t</i>		0.93		1.25		1.50	
<i>RE</i>		0.38		0.69		1.00	

Bold, italic and underline indicates $P < .01$.

Std: standard deviation.

ES: mean difference divided by standard deviation. For example, the ES of the EQ-5D-3L China utility scores between the 25 to 34 and 35 to 44 age groups was (0.96–0.98)/0.07.

RE: F or t -square ratio between the EQ-5D utility and SF-6D utility. We set the F or t -square of the SF-6D as the denominator. $RE > 1$ indicates that the EQ-5D has better discriminant validity than the SF-6D, $RE = 1$ indicates that the two measures have equal discriminant validity, and $RE < 1$ indicates that the EQ-5D has worse discriminant validity than the SF-6D.

EQ-5D-3L_{CN}

Table 5
Discriminant validity of the EQ-5D-3L and SF-6D in different health groups.(n=2178).

Social demographics		EQ-5D-3L _{CN}		EQ-5D-3L _{UK}		SF-6D _{UK}	
Health status (N)	N (%)	Mean (Std)	ES	Mean (Std)	ES	Mean (Std)	ES
Self-reported health							
Excellent	101 (4.64)	<u>.99 (.03)</u>		<u>.99 (.04)</u>		<u>.85 (.11)</u>	
Very good	500 (22.96)	<u>.98 (.05)</u>	-0.31	<u>.98 (.08)</u>	-0.37	<u>.84 (.10)</u>	-0.12
Good	755 (34.66)	<u>.97 (.08)</u>	-0.35	<u>.95 (.11)</u>	-0.35	<u>.82 (.10)</u>	-0.21
Fair	714 (32.78)	<u>.93 (.12)</u>	-0.51	<u>.90 (.17)</u>	-0.46	<u>.75 (.11)</u>	-0.61
Poor	108 (4.96)	<u>.78 (.18)</u>	-1.20	<u>.72 (.26)</u>	-1.06	<u>.63 (.10)</u>	-1.12
F		121.37**		99.28**		129.84**	
RE		0.93		0.76		1.00	
Chronic disease							
0	1505 (69.10)	<u>.98 (.07)</u>		<u>.97 (.09)</u>		<u>.82 (.11)</u>	
1	528 (24.24)	<u>.90 (.14)</u>	-1.04	<u>.87 (.20)</u>	-1.09	<u>.75 (.11)</u>	-0.68
2+	145 (6.66)	<u>.84 (.16)</u>	-0.44	<u>.79 (.23)</u>	-0.38	<u>.70 (.13)</u>	-0.37
F		192.1**		179.04**		137.83**	
RE		1.39		1.30		1.00	
Outpatients							
No	1590 (73.05)	<u>.97 (.08)</u>		<u>.96 (.11)</u>		<u>.82 (.11)</u>	
Yes	587 (26.95)	<u>.89 (.14)</u>	-1.04	<u>.85 (.21)</u>	-1.08	<u>.73 (.12)</u>	-0.83
t		13.39**		12.87**		16.26**	
RE		0.68		0.63		1.00	
Emotions							
Better	585 (26.87)	<u>.96 (.09)</u>		<u>.94 (.13)</u>		<u>.79 (.12)</u>	
Normal	1458 (66.93)	<u>.95 (.10)</u>	-0.01	<u>.94 (.15)</u>	0.00	<u>.80 (.11)</u>	0.06
Worse	135 (6.20)	<u>.88 (.14)</u>	-0.76	<u>.83 (.20)</u>	-0.73	<u>.72 (.11)</u>	-0.69
F		37.54**		33.01**		28.71**	
RE		1.31		1.15		1.00	
QOL score							
>=80	1572 (72.16)	<u>.97 (.08)</u>		<u>.96 (.11)</u>		<u>.81 (.11)</u>	
<80	606 (27.84)	<u>.90 (.15)</u>	-0.90	<u>.87 (.20)</u>	-0.79	<u>.74 (.12)</u>	-0.64
t		11.11**		10.26**		12.65**	
RE		0.77		0.66		1.00	

Bold, italic and underline indicates $P < .01$.

Std: standard deviation.

ES: mean difference divided by standard deviation. For example, the ES of the EQ-5D-3L China utility scores between the groups who self-reported their health as "fair" and "good" were (0.93-0.97)/0.08. RE: F or t-square ratio between the EQ-5D utility and SF-6D utility. We set the F or t-square of the SF-6D as the denominator. RE > 1 indicates that the EQ-5D has better discriminant validity than the SF-6D, RE = 1 indicates that the two measures have equal discriminant validity, and RE < 1 indicates that the EQ-5D has worse discriminant validity than SF-6D.

developed based on the local population health preferences and are usually affected by cultural differences.^[66] Previous studies comparing different countries' specific preference value sets suggest that there are some differences between health statuses in the value sets of the UK, the US, Spanish and Japan.^[17,18,67,68] These differences may lead to different outcomes of QALYs, which in turn may influence healthcare decisions when the QALYs are used for economic evaluations. Therefore, a country-specific preference value set should be applied when it is available. Further research is needed to develop the SF-6D Chinese general population preference value sets.

Our study must be interpreted in light of several study limitations. First, the Chinese pharmaceutical economic research guide suggests that the utility measures should use country-specific value sets.^[12] In this study, we used the UK population-based value set to calculate SF-6D scores because of the lack of China-specific SF-6D value sets. It is a limitation to compare the SF-6D and EQ-5D-3L using different country-specific values. Second, the country-specific value sets were developed based on different methods: TTO and SG. However, previous studies using the same method and population-based value sets have also displayed differences.^[31,63] Third, this is a cross-sectional study without any interventions. Therefore, it is not possible to

compare longitudinal responsiveness and discriminant validity. Therefore, new research on establishing the mainland China-specific SF-6D value sets may be an important future advancement. And recently, the mainland China-specific EQ-5D-5L value sets have been developed by TTO method.^[69] Future studies are needed to compare the psychometric properties between EQ-5D-5L and SF-6D to explore the responsiveness of them in studies which involved interventions that would lead to changes in health conditions and assess whether the choice of EQ-5D-5L or SF-6D have different impact on estimates cost-utility and decision making.

In conclusion, the study compared the construct validity, sensitivity and level of agreement between the EQ-5D-3L and SF-6D in the Chinese general population in Chengdu. Both are valid economic evaluation instruments in the Chinese general population. Country-specific value sets should be used when available. It seems that the 2 measurements are not interchangeable. The EQ-5D-3L has a higher ceiling effect and higher level of discriminant validity to discriminate among different socio-demographic groups, and the SF-6D has a lower ceiling effect and higher level of discriminant validity in moderately healthy groups. Users may consider this evidence in the choice of these instruments.

Table 6
Discriminant validity of the SF-6D in the EQ-5D full-health group (n = 1,625).

Social demographics	N (%)	SF-6D _{UK}	ES	Social demographics	N (%)	SF-6D _{UK}	ES
Age group (years)				Annual household income (in 1000-yuan increments)			
18–24	288 (17.73)	.81 (.10)		<=15	206 (12.75)	.82 (.10)	
25–34	305 (18.72)	.82 (.11)	0.09	15–30	392 (24.26)	.81 (.11)	–0.02
35–44	279 (17.18)	.83 (.09)	0.08	30–50	399 (24.69)	.83 (.10)	0.14
45–54	270 (16.63)	.84 (.10)	0.11	50–100	411 (25.43)	.83 (.10)	–0.02
55–64	271 (16.69)	.82 (.10)	–0.24	>100	208 (12.87)	.84 (.10)	0.09
>=65	212 (13.05)	.81 (.10)	–0.15				
Gender				Health insurance			
Male	760 (46.80)	0.83 (0.10)		No	137 (8.44)	0.82 (0.11)	
Female	864 (53.20)	0.82 (0.10)	–0.04	Yes	1487 (91.56)	0.83 (0.10)	–0.04
Marital status				Number of chronic diseases			
Single	320 (19.7)	0.81 (0.10)		0	1279 (78.76)	.83 (.10)	
Married/cohabiting	1209 (74.45)	0.83 (0.10)	0.14	1	292 (17.98)	.79 (.10)	–0.40
Widowed	23 (1.42)	0.81 (0.11)	–0.22	2+	53 (3.26)	.80 (.11)	0.08
Divorced	72 (4.43)	0.81 (0.11)	0.02	Outpatients in 2 weeks			
Education level				Yes	287 (17.67)	.78 (.10)	
Elementary school or illiterate	421 (25.92)	0.82 (0.10)		No	1337 (82.33)	.83 (.10)	0.49
Middle school	459 (28.20)	0.83 (0.10)	0.13	Self-reported health			
High school	323 (19.90)	0.82 (0.11)	–0.16	Excellent	97 (5.97)	.86 (.11)	
Technical college	198 (12.19)	0.82 (0.11)	0.04	Very good	454 (27.96)	.85 (.10)	–0.09
University/ graduate education	224 (13.79)	0.82 (0.10)	–0.05	Good	603 (37.13)	.84 (.09)	–0.11
Working status				Fair	447 (27.52)	.79 (.10)	–0.52
Working	1104 (67.98)	.83 (.10)		Poor	23 (1.42)	.71 (.10)	–0.81
Retired	423 (26.05)	.82 (.10)	–0.11	Emotions			
Other	98 (5.97)	.84 (.10)	0.26	Better	448 (27.57)	.82 (.11)	
Region				As usual	1120 (68.92)	.83 (.10)	0.09
Urban	763 (46.98)	0.83 (0.10)		Worse	57 (3.51)	.77 (.10)	–0.56
Rural	861 (53.02)	0.82 (0.11)	–0.09	QOL score			
				>=80	1292 (79.51)	.83 (.10)	
				<80	333 (20.49)	.79 (.10)	–0.34

Bold and underline indicates $P < .05$. **Bold, italic and underline** indicates $P < .01$. *Std* = standard deviation.

Acknowledgments

We would like to thank all the respondents who participated in the survey and all the investigators who helped conduct the survey in 2012.

Author contributions

LZ, NL and DL conceived the study and designed the study protocol; LZ, YH and ZL performed the data collection; LZ wrote the first draft of manuscript in addition to performing the literature search; and XL and NL provided statistical support and critically revised the manuscript for intellectual content. All authors read and approved the final manuscript. NL is the corresponding author of the paper.

Conceptualization: Longchao Zhao, Ningxiu Li.

Data curation: Longchao Zhao, Ningxiu Li.

Formal analysis: Longchao Zhao.

Funding acquisition: Ningxiu Li.

Investigation: Longchao Zhao, Danping Liu, Yan He, Zhijun Liu, Ningxiu Li.

Methodology: Longchao Zhao, Xiang Liu, Danping Liu, Yan He, Zhijun Liu, Ningxiu Li.

Project administration: Longchao Zhao, Danping Liu, Yan He, Ningxiu Li.

Supervision: Xiang Liu, Danping Liu, Ningxiu Li.

Writing – original draft: Longchao Zhao.

Writing – review & editing: Longchao Zhao, Xiang Liu, Ningxiu Li.

References

- [1] Neumann PJ, Thorat T, Shi J, et al. The changing face of the cost-utility literature, 1990–2012. *Value Health* 2015;18:271–7.
- [2] Joish VN, Oderda GM. Cost-utility analysis and quality adjusted life years. *J Pain Palliat Care Pharmacother* 2005;19:57–61.
- [3] Shandas M, Bassilious K, Murray PI. Health-related quality of life in patients with uveitis. *Br J Ophthalmol* 2018;bjophthalmol-2018-312882.
- [4] McDonough CM, Tosteson ANA. Measuring preferences for cost-utility analysis - how choice of method may influence decision-making. *Pharmacoeconomics* 2007;25:93–106.
- [5] Brazier J, Usherwood T, Harper R, et al. Deriving a preference-based single index from the UK SF-36 health survey. *J Clin Epidemiol* 1998;51:1115–28.
- [6] Feeny D, Furlong W, Boyle M, et al. Multi-attribute health status classification systems. Health utilities index. *Pharmacoeconomics* 1995;7:490–502.
- [7] Feeny D, Furlong W, Torrance GW, et al. Multiattribute and single-attribute utility functions for the health utilities index mark 3 system. *Med Care* 2002;40:113–28.
- [8] Brooks R. EuroQol: the current state of play. *Health Policy* 1996;37:53–72.
- [9] 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* 2011;17:27–36.
- [10] Brazier JE, Fukuhara S, Roberts J, et al. Estimating a preference-based index from the Japanese SF-36. *J Clin Epidemiol* 2009;62:1323–31.
- [11] Lamu AN, Olsen JA. Testing alternative regression models to predict utilities: mapping the QLQ-C30 onto the EQ-5D-5L and the SF-6D. *Qual Life Res* 2018;27:2823–39.
- [12] Liu G, Hu S, Wu J. China guidelines for pharmacoeconomic evaluations. *China J Pharm Econ* 2011;3:6–48.
- [13] Devlin N, Parkin D. Guidance to users of EQ-5D value sets EQ-5D Value Sets. 2007;Springer, 39–52.
- [14] Brooks R. The EuroQol Group After 25 Years. 2012;Springer, 13–35.

- [15] Brazier JE, Roberts J. The estimation of a preference-based measure of health from the SF-12. *Med Care* 2004;42:851–9.
- [16] Oppe M, Devlin NJ, Szende A. EQ-5D Value Sets: Inventory, Comparative Review and User Guide. Springer; 2007.
- [17] Dolan P. Modeling valuations for EuroQol health states. *Med Care* 1997;35:1095–108.
- [18] Shaw JW, Johnson JA, Coons SJ. US valuation of the EQ-5D health states: development and testing of the D1 valuation model. *Med Care* 2005;43:203–20.
- [19] Viney R, Norman R, King MT, et al. Time trade-off derived EQ-5D weights for Australia. *Value Health* 2011;14:928–36.
- [20] Tsuchiya A, Ikeda S, Ikegami N, et al. Estimating an EQ-5D population value set: the case of Japan. *Health Econ* 2002;11:341–53.
- [21] Liu GG, Wu H, Li M, et al. Chinese time trade-off values for EQ-5D health states. *Value Health* 2014;17:597–604.
- [22] Wang HM, Patrick DL, Edwards TC, et al. Validation of the EQ-5D in a general population sample in urban China. *Qual Life Res* 2012;21:155–60.
- [23] Jin H, Wang B, Gao Q, et al. Comparison between EQ-5D and SF-6D utility in rural residents of Jiangsu Province, China. *PLoS One* 2012;7:e41550.
- [24] Lam CL, Brazier J, McGhee SM. Valuation of the SF-6D health states is feasible, acceptable, reliable, and valid in a Chinese population. *Value Health* 2008;11:295–303.
- [25] Ferreira LN, Ferreira PL, Pereira LN, et al. A Portuguese value set for the SF-6D. *Value Health* 2010;13:624–30.
- [26] Cruz LN, Camey SA, Hoffmann JF, et al. Estimating the SF-6D value set for a population-based sample of Brazilians. *Value Health* 2011;14:S108–14.
- [27] Brazier J, Roberts J, Deverill M. The estimation of a preference-based measure of health from the SF-36. *J Health Econ* 2002;21:271–92.
- [28] Ferreira PL, Ferreira LN, Pereira LN. SF-6D Portuguese population norms. *Eur J Health Econ* 2015;16:235–41.
- [29] Berg B. Sf-6d population norms. *Health Econ* 2012;21:1508–12.
- [30] Wong CKH, Mulhern B, Cheng GHL, et al. SF-6D population norms for the Hong Kong Chinese general population. *Qual Life Res* 2018;27:2349–59.
- [31] Shirowa T, Fukuda T, Ikeda S, et al. Japanese population norms for preference-based measures: EQ-5D-3L, EQ-5D-5L, and SF-6D. *Qual Life Res* 2016;25:707–19.
- [32] Yousefi M, Najafi S, Ghaffari S, et al. Comparison of SF-6D and EQ-5D scores in patients with breast cancer. *Iran Red Crescent Med J* 2016;18:e23556.
- [33] Huppertz-Hauss G, Aas E, Hoivik ML, et al. Comparison of the multiattribute utility instruments EQ-5D and SF-6D in a Europe-wide population-based cohort of patients with inflammatory bowel disease 10 years after diagnosis. *Gastroent Res Pract* 2016;1–9.
- [34] Joshi N, Khanna R, Bentley JP, et al. A comparison of Eq-5d-3l and Sf-6d among caregivers of individuals with multiple sclerosis. *Value Health* 2016;19:A66–166.
- [35] Kularatna S, Byrnes J, Chan YK, et al. Comparison of the EQ-5D-3L and the SF-6D (SF-12) contemporaneous utility scores in patients with cardiovascular disease. *Qual Life Res* 2017;26:3399–408.
- [36] Dritsaki M, Petrou S, Williams M, et al. An empirical evaluation of the SF-12, SF-6D, EQ-5D and michigan hand outcome questionnaire in patients with rheumatoid arthritis of the hand. *Health Qual Life Outcomes* 2017;15.
- [37] Zhang F, Yang Y, Huang T, et al. Is there a difference between EQ-5D and SF-6D in the clinical setting? A comparative study on the quality of life measured by AIMS2-SF, EQ-5D and SF-6D scales for osteoarthritis patients. *Int J Rheum Dis* 2018;21:1185–92.
- [38] Summerfield AQ, Barton GR, Group UKCIS. Sensitivity of EQ-5D-3L, HUI2, HUI3, and SF-6D to changes in speech reception and tinnitus associated with cochlear implantation. *Qual Life Res* 2018;1–0.
- [39] Harvie HS, Honeycutt AA, Neuwahl SJ, et al. Responsiveness and minimally important difference of SF-6D and EQ-5D utility scores for the treatment of pelvic organ prolapse. *Am J Obstet Gynecol* 2018.
- [40] Sach TH, Barton GR, Jenkinson C, et al. Comparing cost-utility estimates does the choice of EQ-5D or SF-6D matter? *Med Care* 2009;47:889–94.
- [41] Davis JC, Liu-Ambrose T, Khan KM, et al. SF-6D and EQ-5D result in widely divergent incremental cost-effectiveness ratios in a clinical trial of older women: implications for health policy decisions. *Osteoporos Int* 2012;23:1849–57.
- [42] Hanmer J, Cherepanov D, Palta M, et al. Health condition impacts in a nationally representative cross-sectional survey vary substantially by preference-based health index. *Med Decis Making* 2016;36:264–74.
- [43] Mukaka MM. A guide to appropriate use of correlation coefficient in medical research. *Malawi Med J* 2012;24:69–71.
- [44] Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1986;1:307–10.
- [45] Thaweethamcharoen T, Noparatayaporn P, Sritippayawan S, et al. Comparison of EQ-5D-5L, VAS, and SF-6D in Thai patients on peritoneal dialysis. *Value Health Reg Issues* 2018;18:59–64.
- [46] Palfreyman S, Mulhern B. The psychometric performance of generic preference-based measures for patients with pressure ulcers. *Health Qual Life Outcomes* 2015;13:117–26.
- [47] Kivits J, Erpelding ML, Guillemin F. Social determinants of health-related quality of life. *Revue d'Épidémiologie et de Santé Publique* 2013;61(Supplement 3):S189–94.
- [48] Kazis LE, Anderson JJ, Meenan RF. Effect sizes for interpreting changes in health status. *Med Care* 1989;27(3 Suppl):S178–189.
- [49] Cohen J. Statistical power analysis. *Curr Direct Psychol Sci* 1992;1:98–101.
- [50] Luo N, Johnson JA, Shaw JW, et al. 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* 2009;47:53–60.
- [51] Liang MH, Larson MG, Cullen KE, et al. Comparative measurement efficiency and sensitivity of five health status instruments for arthritis research. *Arthritis Rheum* 1985;28:542–7.
- [52] Lam ET, Lam CL, Fong DY, et al. Is the SF-12 version 2 Health Survey a valid and equivalent substitute for the SF-36 version 2 health survey for the Chinese? *J Eval Clin Pract* 2013;19:200–8.
- [53] Ferreira LN, Ferreira PL, Pereira LN. Comparing the performance of the SF-6D and the EQ-5D in different patient groups. *Acta Medica Port* 2014;27:236–45.
- [54] Torrance N, Lawson KD, Afolabi E, et al. Estimating the burden of disease in chronic pain with and without neuropathic characteristics: does the choice between the EQ-5D and SF-6D matter? *Pain* 2014;155:1996–2004.
- [55] Yang F, Lau T, Lee E, et al. Comparison of the preference-based EQ-5D-5L and SF-6D in patients with end-stage renal disease (ESRD). *Eur J Health Econ* 2015;16:1019–26.
- [56] Kanters TA, Redekop WK, Kruishaar ME, et al. Comparison of EQ-5D and SF-6D utilities in Pompe disease. *Qual Life Res* 2015;24:837–44.
- [57] Garcia-Gordillo MA, del Pozo-Cruz B, Adsuar JC, et al. Validation and comparison of EQ-5D-3L and SF-6D instruments in a Spanish Parkinson's disease population sample. *Nutr Hosp* 2015;32:2808–21.
- [58] Kontodimopoulos N, Stamatopoulou E, Brinia A, et al. Are condition-specific utilities more valid than generic preference-based ones in asthma? Evidence from a study comparing EQ-5D-3L and SF-6D with AQL-5D. *Expert Rev Pharmacoecon Outcomes Res* 2018;18:667–75.
- [59] Al Sayah F, Qiu WY, Xie F, et al. Comparative performance of the EQ-5D-5L and SF-6D index scores in adults with type 2 diabetes. *Qual Life Res* 2017;26:2057–66.
- [60] Brazier J, Roberts J, Tsuchiya A, et al. A comparison of the EQ-5D and SF-6D across seven patient groups. *Health Econ* 2004;13:873–84.
- [61] Barton GR, Sach TH, Avery AJ, et al. A comparison of the performance of the EQ-5D and SF-6D for individuals aged ≥ 45 years. *Health Econ* 2008;17:815–32.
- [62] Szende A, Leidy NK, Stahl E, et al. Estimating health utilities in patients with asthma and COPD: evidence on the performance of EQ-5D and SF-6D. *Qual Life Res* 2009;18:267–72.
- [63] Whitehurst DG, Norman R, Brazier JE, et al. Comparison of contemporaneous EQ-5D and SF-6D responses using scoring algorithms derived from similar valuation exercises. *Value Health* 2014;17:570–7.
- [64] Kontodimopoulos N, Pappa E, Papadopoulos A, et al. Comparing SF-6D and EQ-5D utilities across groups differing in health status. *Qual Life Res* 2009;18:87–97.
- [65] Bharmal M, Thomas J. Comparing the EQ-5D and the SF-6D descriptive systems to assess their ceiling effects in the US general population. *Value Health* 2006;9:262–71.
- [66] Norman R, Cronin P, Viney R, et al. International comparisons in valuing EQ-5D health states: a review and analysis. *Value Health* 2009;12:1194–200.
- [67] Badia X, Roset M, Herdman M, et al. A comparison of United Kingdom and Spanish general population time trade-off values for EQ-5D health states. *Med Decis Making* 2001;21:7–16.
- [68] Kiadaliri AA, Eliasson B, Gerdtham U-G. Does the choice of EQ-5D tariff matter? A comparison of the Swedish EQ-5D-3L index score with UK, US, Germany and Denmark among type 2 diabetes patients. *Health Qual Life Outcomes* 2015;13.
- [69] Luo N, Liu G, Li M, et al. Estimating an EQ-5D-5L value set for China. *Value Health* 2017;20:662–9.