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, 2,186 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 2,186 respondents, 2,182 completed the questionnaire, and 2,178 (18–82 years old, mean 46.09 ± 17.49) met the data quality requirement. The mean scores for the EQ-5D-3L\(^a\)CN, EQ-5D-3L\(^b\)UK, and SF-6D\(^c\)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\(^a\)CN and SF-6D\(^c\)UK and between the EQ-5D-3L\(^b\)UK and SF-6D\(^c\)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. The EQ-5D-3L\(^a\)CN, EQ-5D-3L\(^b\)UK, and SF-6D have good discriminant validity in different sociodemographic and health condition groups. The SF-6D has a higher ceiling on 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\(^a\)CN = EQ-5D-3L utility calculated by Chinese TTO preference value set, EQ-5D-3L\(^b\)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\(^a\)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

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. 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, 9] five-level EQ-5D (EQ-5D-5L) and the Short-Form Six-Dimension (SF-6D) [10, 11] are widely used 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] TheEQ-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. [11] 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, [10] Hong Kong, [24] Japan, [20] 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, 13–27] 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 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-6D.

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 (3⁵) 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. 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_CNS, 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 (Sd).

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. [13] 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 111111 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-3LCN score was 0.95 (Std: 0.11) (median 1.0; interquartile range: 0.13), and that of the EQ-5D-3LUK score was 0.93 (Std: 0.15) (median 1.0; interquartile range: 0.15). The mean of the SF-6DUK 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 pain/discomfort, 0.24 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-3LCN and SF-6D UK (Fig. 1) and between the EQ-5D-3LUK and SF-6D UK (Fig. 2)
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, 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 and SF-6D 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 and SF-6D is 0.137, with 62 (2.85%) observations out of 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 and SF-6D was greater than that between the EQ-5D-3L and SF-6D.

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 and EQ-5D-3L tend to have very high ceiling effects (n = 1625, 74.60%) and a low floor effects (0.05%). The EQ-5D-3L utility scores are skewed toward high scores and are more skewed than the SF-6D utility scores. The SF-6D has low floor (0.05%) and ceiling effects (n = 55, 2.53%), and the distribution of the SF-6D 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, EQ-5D-3L and SF-6D 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, EQ-5D-3L and SF-6D 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 < 0.20), except for the SF-6D 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 and EQ-5D-3L 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 and EQ-5D-3L are greater than the SF-6D, and the EQ-5D-3L is greater than the EQ-5D-3L (Tables 4 and 5). The result also shows that the standard deviation of the EQ-5D-3L is lower than that of the EQ-5D-3L and SF-6D. Although the REs of the EQ-5D-3L and EQ-5D-3L were greater than 1.00 in single indicators of sociodemographic characteristics (age, gender, education, marriage status, employment, annual household income, chronic conditions, and...
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 (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-6DUK 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-6DUK 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-6DUK utility than people with poor or worse conditions. The effect sizes of the SF-6DUK 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. [31–35,33–58]
First, the scores on both the EQ-5D-3LCN and EQ-5D-3LUK are higher than those on the SF-6DUK 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-3LCN and SF-6DUK and 0.137 (Std: 0.141) (P < .001) between the EQ-5D-3LUK and SF-6DUK. 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-6DUK scoring algorithm was derived by the SG method, and the EQ-5D-3LCN and EQ-5D-3LUK 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-3LUK and SF-6DUK 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 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

| Level | 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

| Level | PF | RL | SF | PA | MH | VT |
|-------|----|----|----|----|----|----|
| 1     | 1163 (53.43) | 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.
Third, the EQ-5D-3L CN, EQ-5D-3LUK and SF-6DUK 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-3LCN and EQ-5D-3LUK 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-3LCN and EQ-5D-3LUK. Nevertheless, the results still show that the SF-6DUK 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-6DUK 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 groups 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-3LUK and SF-6DUK 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-3LUK are lower than those on the EQ-5D-3LCN 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                                      | EQ-5D-3L China | 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)    | 0.99 (0.04) | 0.98 (0.06) | 0.81 (0.10) |
| 25–34                                                     | 352 (16.16)    | 0.98 (0.07) | 0.97 (0.09) | 0.81 (0.11) | 0.02       |
| 35–44                                                     | 351 (16.12)    | 0.96 (0.11) | 0.94 (0.16) | 0.81 (0.11) | 0.00       |
| 45–54                                                     | 355 (16.30)    | 0.96 (0.08) | 0.94 (0.12) | 0.81 (0.11) | 0.03       |
| 55–64                                                     | 435 (19.97)    | 0.92 (0.11) | 0.89 (0.17) | 0.77 (0.12) | 0.36       |
| >65                                                       | 365 (16.78)    | 0.90 (0.15) | 0.87 (0.20) | 0.75 (0.12) | 0.20       |
| F                                                         | 36.32**        | 31.93**     | 20.29**    |
| RE                                                        | 1.79           | 1.55        | 1.00      |
| Gender                                                    |                |             |           |             |
| Male                                                      | 974 (44.72)    | 0.96 (0.10) | 0.94 (0.14) | 0.80 (0.11) |
| Female                                                    | 1204 (55.28)   | 0.94 (0.11) | 0.92 (0.16) | 0.79 (0.12) | 0.13       |
| t                                                         | 3.21**         | 3.26**      | 3.06**    |
| RE                                                        | 1.09           | 1.13        | 1.00      |
| Marital status                                            |                |             |           |             |
| Single                                                    | 358 (16.44)    | 0.98 (0.06) | 0.96 (0.09) | 0.81 (0.10) |
| Married                                                   | 1661 (76.26)   | 0.95 (0.11) | 0.93 (0.15) | 0.79 (0.12) | 0.13       |
| Widowed                                                   | 126 (5.78)     | 0.94 (0.11) | 0.93 (0.12) | 0.78 (0.11) | 0.15       |
| Divorced                                                  | 33 (1.52)      | 0.90 (0.14) | 0.87 (0.21) | 0.75 (0.12) | 0.26       |
| F                                                         | 22.86**        | 20.23**     | 8.94**    |
| RE                                                        | 2.56           | 2.27        | 1.00      |
| Education                                                 |                |             |           |             |
| Elementary school or illiterate                           | 678 (31.13)    | 0.92 (0.14) | 0.89 (0.19) | 0.77 (0.12) |
| Middle school                                             | 604 (27.73)    | 0.95 (0.09) | 0.94 (0.14) | 0.81 (0.11) | 0.30       |
| High school                                               | 402 (18.46)    | 0.97 (0.08) | 0.96 (0.10) | 0.80 (0.11) | 0.05       |
| Technical college                                         | 236 (10.84)    | 0.97 (0.08) | 0.96 (0.13) | 0.81 (0.11) | 0.05       |
| University or graduate education                          | 258 (11.65)    | 0.98 (0.07) | 0.97 (0.10) | 0.81 (0.11) | 0.00       |
| F                                                         | 25.08**        | 23.03**     | 11.40**   |
| RE                                                        | 2.20           | 2.02        | 1.00      |
| Employment                                                |                |             |           |             |
| Employed                                                  | 1378 (63.27)   | 0.96 (0.08) | 0.95 (0.13) | 0.81 (0.11) |
| Retired                                                   | 666 (30.58)    | 0.92 (0.13) | 0.90 (0.17) | 0.77 (0.12) | 0.33       |
| Other                                                     | 134 (6.15)     | 0.94 (0.15) | 0.91 (0.20) | 0.80 (0.13) | 0.28       |
| F                                                         | 38.15**        | 29.48**     | 23.84**   |
| RE                                                        | 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) | 0.93 (0.15) | 0.79 (0.12) | 0.03       |
| t                                                         | 1.30           | 0.42        | 0.75      |
| RE                                                        | 3.00           | 0.31        | 1.00      |
| Annual household income (in 1000-yuan increments)         |                |             |           |             |
| <15                                                       | 331 (15.27)    | 0.91 (0.15) | 0.88 (0.21) | 0.77 (0.12) |
| 15–30                                                     | 540 (24.91)    | 0.95 (0.10) | 0.93 (0.13) | 0.79 (0.12) | 0.15       |
| 30–50                                                     | 526 (24.26)    | 0.96 (0.09) | 0.94 (0.13) | 0.80 (0.11) | 0.12       |
| 50–100                                                    | 522 (24.08)    | 0.96 (0.09) | 0.94 (0.15) | 0.80 (0.11) | 0.01       |
| >100                                                      | 249 (11.49)    | 0.97 (0.08) | 0.96 (0.12) | 0.82 (0.11) | 0.15       |
| F                                                         | 14.09**        | 12.27**     | 8.77**    |
| RE                                                        | 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) | 0.93 (0.15) | 0.79 (0.12) | 0.12       |
| t                                                         | 0.93           | 1.25        | 1.50      |
| RE                                                        | 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 of t square ratio between the EQ-5D utility and SF-6D utility. We set the F of 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 China
developed based on the local population health preferences and are usually affected by cultural differences.\(^\text{[60]}\) 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.\(^\text{[17,18,67,68]}\) 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.\(^\text{[13]}\) 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.\(^\text{[31,63]}\) 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.\(^\text{[69]}\) 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 are interchangeable. And recently, the mainland China-specific EQ-5D-5L value sets have been developed by TTO method.\(^\text{[69]}\) 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 are interchangeable.

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

| Social demographics | EQ-5D-3L<sub>CH</sub> | SF-6D<sub>CH</sub> | EQ-5D-3L<sub>UK</sub> | SF-6D<sub>UK</sub> |
|---------------------|----------------------|----------------------|----------------------|----------------------|
| Health status (N)   | Mean (Std) ES        | Mean (Std) ES        | Mean (Std) ES        | Mean (Std) ES        |
| Self-reported health|                      |                      |                      |                      |
| Excellent 101 (4.64)| .99 (.03) –0.31 .85 (.11) | .99 (.04) –0.37 .84 (.10) |                      |                      |
| Very good 500 (22.96)| .96 (.05)             | .98 (.08)             | .95 (.11)             |                      |
| Good 755 (34.66)    | .97 (.08) –0.35 .82 (.10) | .97 (.10) –0.35 .82 (.10) |                      |                      |
| Fair 714 (32.78)    | .93 (.12) –0.51 .75 (.11) | .90 (.17) –0.46 .75 (.11) |                      |                      |
| Poor 108 (4.96)     | .78 (.18) –1.20 .63 (.10) | .72 (.26) –1.06 .63 (.10) |                      |                      |
| F                   | 121.37**             | 99.28**              | 129.84**             |                      |
| RE                  | 0.03                 | 0.76                 | 1.00                 |                      |
| Chronic disease     |                      |                      |                      |                      |
| 0 1505 (69.10)      | .98 (.07)             | .97 (.09)             | .82 (.11)             |                      |
| 1 528 (24.24)       | .90 (.14) –1.04 .75 (.11) | .87 (.20) –1.09 .75 (.11) |                      |                      |
| 2+ 145 (6.66)       | .84 (.16) –0.44 .70 (.13) | .79 (.23) –0.38 .70 (.13) |                      |                      |
| F                   | 192.1**              | 179.04**             | 137.83**             |                      |
| RE                  | 1.39                 | 1.30                 | 1.00                 |                      |
| Outpatients         |                      |                      |                      |                      |
| No 1590 (73.05)     | .97 (.08)             | .96 (.11)             | .82 (.11)             |                      |
| Yes 587 (26.95)     | .89 (.14) –1.04 .73 (.12) | .85 (.21) –1.08 .73 (.12) |                      |                      |
| t                   | 13.39**              | 12.87**              | 16.26**              |                      |
| RE                  | 0.68                 | 0.63                 | 1.00                 |                      |
| Emotions            |                      |                      |                      |                      |
| Better 585 (26.87)  | .96 (.09)             | .94 (.13)             | .79 (.12)             |                      |
| Normal 1458 (66.93) | .95 (.10) –0.01 .80 (.11) | .94 (.15) 0.00 .80 (.11) |                      |                      |
| Worse 125 (5.69)    | .88 (.14) –0.76 .63 (.20) | .83 (.20) –0.73 .72 (.11) |                      |                      |
| F                   | 37.54**              | 33.01**              | 26.71**              |                      |
| RE                  | 1.31                 | 1.15                 | 1.00                 |                      |
| QOL score           |                      |                      |                      |                      |
| >=80 1572 (72.16)   | .97 (.08)             | .96 (.11)             | .81 (.11)             |                      |
| <80 606 (27.84)     | .90 (.15) –0.90 .87 (.20) | .90 (.15) –0.90 .87 (.20) |                      |                      |
| t                   | 11.11**              | 10.26**              | 12.65**              |                      |
| RE                  | 0.77                 | 0.69                 | 1.00                 |                      |

Bold, italic and underline indicates \( P<0.01 \).
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: For \( F \)-square ratio between the EQ-5D-3L utility and SF-6D utility. We set the \( F \)-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.
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) | 0.81 (1.0) |   | <=15                | 206 (12.75) | 0.82 (1.0) | -0.02 |
| 25–34               | 305 (18.72) | 0.82 (1.1) | 0.09 | 15–30               | 392 (24.26) | 0.81 (1.1) | -0.02 |
| 35–44               | 279 (17.18) | 0.83 (0.9) | 0.08 | 30–50               | 399 (24.69) | 0.83 (1.0) | 0.14 |
| 45–54               | 270 (16.63) | 0.84 (1.0) | 0.11 | 50–100              | 411 (25.43) | 0.83 (1.0) | -0.02 |
| 55–64               | 271 (16.69) | 0.82 (1.0) | -0.24 | >100                | 208 (12.87) | 0.84 (1.0) | 0.09 |
| >=65                | 212 (13.09) | 0.81 (1.0) | -0.15 |                     |        |          |    |
| **Gender**          |       |          |    |                     |        |          |    |
| Male                | 760 (46.80) | 0.83 (0.10) |   | Yes                 | 1487 (91.56) | 0.83 (0.10) | -0.04 |
| Female              | 864 (53.20) | 0.82 (0.10) | -0.04 |                     |        |          |    |
| **Marital status**  |       |          |    |                     |        |          |    |
| Single              | 320 (19.7) | 0.81 (0.10) |   | 0                    | 1279 (78.76) | 0.83 (1.0) | -0.40 |
| Married/cohabiting | 1209 (74.45) | 0.83 (0.10) | 0.14 | 1                   | 292 (17.98) | 0.79 (1.0) | -0.1 |
| Widowed             | 23 (1.42) | 0.81 (0.11) | -0.22 | 2+                  | 53 (3.28) | 0.81 (1.0) | 0.08 |
| Divorced            | 72 (4.43) | 0.81 (0.11) | 0.02 |                     |        |          |    |
| **Education level** |       |          |    |                     |        |          |    |
| Elementary school/illiterate | 421 (25.92) | 0.82 (0.10) |   | Yes                 | 287 (17.67) | 0.78 (1.0) | 0.49 |
| Middle school       | 459 (28.20) | 0.83 (0.10) | 0.13 | No                  | 1337 (82.33) | 0.83 (1.0) | 0.49 |
| High school         | 323 (19.90) | 0.82 (0.11) | -0.16 | 0                   | 1337 (82.33) | 0.83 (1.0) | 0.49 |
| Technical college   | 198 (12.19) | 0.82 (0.11) | 0.04 | 0                   | 1337 (82.33) | 0.83 (1.0) | 0.49 |
| University/graduate education | 224 (13.79) | 0.82 (0.10) | -0.05 | 2                   | 1337 (82.33) | 0.83 (1.0) | 0.49 |
| **Working status**  |       |          |    |                     |        |          |    |
| Working             | 1104 (67.98) | 0.83 (1.0) |    | Poor                | 21 (1.42) | 0.74 (1.0) | -0.81 |
| Retired             | 423 (26.05) | 0.82 (1.0) | -0.11 | Fair                | 447 (27.52) | 0.79 (1.0) | -0.52 |
| Other               | 98 (5.97) | 0.84 (1.0) | 0.26 | As usual            | 1120 (68.92) | 0.83 (1.0) | 0.09 |
| Region              |       |          |    |                     |        |          |    |
| Urban               | 763 (46.98) | 0.83 (0.10) |   | 0                   | 1337 (82.33) | 0.83 (1.0) | 0.49 |
| Rural               | 861 (53.02) | 0.82 (0.11) | -0.09 | 0                   | 1337 (82.33) | 0.83 (1.0) | 0.49 |

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] Joshi VN, Oderda GM. Cost-utility analysis and quality adjusted life years. J Pain Palliat Care Pharmacoeuther 2015;19:57–61.
[3] Zhao et al. Medicine (2019) 98:11
