Comparing EQ-5D-3L and EQ-5D-5L in measuring the HRQoL burden of 4 health conditions in China

Guizhi Weng1, Yanming Hong1, Nan Luo2, Clara Mukuria3, Jie Jiang1,4, Zhihao Yang1,5, Sha Li1

Received: 4 August 2021 / Accepted: 5 April 2022 / Published online: 10 May 2022
© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2022

Abstract

Background EQ-5D-3L has been used in the National Health Services Survey of China since 2008 to monitor population health. The five-level version of EQ-5D was developed, but there lacks evidence to support the use of five-level version of EQ-5D in China. This study was conducted to compare the measurement properties of both the EQ-5D-3L and EQ-5D-5L in quantifying health-related quality of life (HRQoL) burden for 4 different health conditions in China.

Methods Participants from China were recruited to complete the 3L and 5L questionnaire via Internet. Quota was set to recruit five groups of individuals, with one group of individuals without any health condition and one group of generalized anxiety disorder (GAD), HIV/AIDS, chronic Hepatitis B (CHB), or depression, respectively. The 3L and 5L were compared in terms of response distributions, percentages of reporting ‘no problems’, index value distributions, known-group validity and their relative efficiency.

Results In total, 500 individuals completed the online survey, including 140 healthy individuals, 122 individuals with hepatitis B, 107 with depression, 90 individuals with GAD and 101 with HIV/AIDS. 5L also had smoother and less clustered index value distributions. Healthy group showed different response distributions to the four condition groups. The percentage of reporting ‘no problems’ decreased significantly in the 5L in all domains (P < 0.01), especially in the pain/discomfort dimension (relative difference: 43.10%). Relative efficiency suggested that 5L had a higher absolute discriminatory power than the 3L version between healthy participant and the other 4 condition groups, especially for the HIV/AIDS group when the 3L results was not significant.

Conclusions The 5L version may be preferable to the 3L, as it demonstrated superior performance with respect to higher sensitivity to mild health problems, better relative efficiency and responses and index value distributions.

Keywords EQ-5D-3L · EQ-5D-5L · Health-related quality of life · Disease burden · China

JEL Classification I15

Zhihao Yang
Zhihao_yang_cn@126.com

Sha Li
tlisha@jnu.edu.cn

1 College of Pharmacy, Jinan University, 601 West Huangpu avenue, 510632 Guangzhou, People’s Republic of China
2 Saw Swee Hock School of Public Health, National University of Singapore, Singapore, Singapore
3 School of Health and Related Research, University of Sheffield, Sheffield, UK
4 Dongguan Institute of Jinan University, Dongguan, People’s Republic of China
5 Health Services Management Department, Guizhou Medical University, Guiyang, People’s Republic of China

Abbreviations

HRQoL Health-related quality of life
EQ-VAS EQ visual analog scale
3L EQ-5D-3L
5L EQ-5D-5L
E-QALY Extending the QALY
SWEMWS Short Warwick–Edinburgh Mental Well-being Scale
GAD Generalized anxiety disorder
RE Relative efficiency
QALYs Quality-adjusted life years
DALYs Disability-adjusted life years
PRO Patient-reported outcome
Introduction

EQ-5D has been used to measure health-related quality of life (HRQoL) across the globe [1]. The EQ-5D questionnaire for data collection consists of two essential parts: a multidimensional health descriptive system and the EQ visual analog scale (EQ-VAS). The EQ-5D descriptive system comprises five dimensions: mobility, self-care, usual activities, pain/discomfort and anxiety/depression [2]. It has two versions, a three-level (EQ-5D-3L) and a five-level version (EQ-5D-5L). EQ-5D-3L (hereinafter 3L) was developed in 1987 and has been the most popular preference-based instrument. It is reported that the 3L has suboptimal sensitivity and suffers from ceiling effects [3]. Therefore, EQ-5D-5L (hereinafter 5L) was introduced in 2009 [4]. In total, 3L defines a total of 243 unique health states, while 5L defines 3,125 health states. The higher number of health states described by the 5L version is aimed at improving sensitivity to small differences or changes in HRQoL [5].

Table 1 summarized the findings of previous validation and 3L/5L comparison studies in China. Overall, both EQ-5D version have been validated in different disease groups and general population in China and most psychometric properties were either good or satisfactory, except for the ceiling effects [7–21]. The detailed results of these studies can be found in Table 5 in Appendix. Ten studies have compared the performance of two versions of EQ-5D in terms of their face validity [14], acceptability [14], ceiling effects [13–21], responsiveness [14], informativity [13–16, 18, 20, 21, 25], test–retest reliability [14, 15, 17], known-group validity [14, 15, 17, 18] and convergent validity [13–15, 17]. Findings of these studies supported the use of 5L. Despite these evidence in different health condition groups, there lacks evidence of how these two versions performed in mental conditions and most published studies did not report the distribution of responses and did not compare the relative efficiency between these two versions. To date, only one study found that the 3L had higher relative efficiency in individuals with hypertension [13]. Given these limitations, we conceptualized this study to further compare the measurement properties of two EQ-5D versions in China.

In this study, we aimed to compare the measurement properties of the two versions of EQ-5D in quantifying the HRQoL burden associated with 4 chronic conditions including chronic hepatitis B (CHB), depression, generalized anxiety disorder (GAD), and AIDS/HIV in China. The selection of disease groups covered two physical conditions, two mental conditions and used a healthy group as the reference group. We hypothesize that the HRQoL

| Measurement properties | EQ-5D-3L | EQ-5D-5L | 3L vs. 5L |
|------------------------|----------|----------|----------|
| Sensitivity            | ○ [6]    | √ [7]    |          |
| Ceiling effect         | × [8, 9] | × / ○ [7, 10–12] | 5L [13–21] |
| Responsiveness         | √ [22]   | √ [10, 23, 24] | 3L & 5L [14] |
| Informativity          | √ [12]   |          | 5L [13–16, 18, 20, 21, 25] |
| Acceptability          |          |          | 5L [14] |
| Construct validity     | √ / ○ [8, 22, 26–28] | √ [29, 30] |          |
| Discriminant validity  | √ / ○ [6, 8, 31] | √ [29, 32] |          |
| Convergent validity    | ○ [6, 31] | √ [10, 12, 29, 32] | 5L [13–15, 17] |
| Face validity          |          |          | 5L [14] |
| Known-group validity   | √ [31]   | √ [12, 32] | 5L / 3L [14, 15, 17, 18] |
| Content validity       | √ [27, 28] | ○ [33] |          |
| Test–retest reliability| ○ [26, 28, 31, 34] | √ / ○ [7, 10, 11, 32] | 5L / 3L [14, 15, 17] |
| Internal consistency   | √ / ○ [9, 26–28] | √ / ○ [7, 11, 29, 30] |          |

○: The measurement properties of EQ-5D-3L or EQ-5D-5L were acceptable in Chinese population in some studies
√: The measurement properties of EQ-5D-3L or EQ-5D-5L had good to excellent performance in Chinese population in some studies
×: The measurement properties of EQ-5D-3L or EQ-5D-5L were not satisfactory in Chinese population in some studies
× / ○: Some studies found serious ceiling effect of EQ-5D-5L, while other studies proved that it was acceptable
5L: 5L was better than 3L
3L & 5L: Both 3L and 5L were satisfactory and no significant difference between 3L and 5L was observed
5L / 3L: 5L performed better than 3L in some studies while 3L was better in other studies
burden of these conditions is mainly in the mental or psychological domain which may be difficult for 3L to detect and necessitates a more sensitive measure like the 5L.

**Methods**

**Participants**

This study utilized the data collected in a cross-sectional online survey in China. The survey was part of an international study called ‘extending the QALY (E-QALY) project’ [35]. The E-QALY project aims to develop a new quality of life measure. As for the sample size, Yfantopoulos et al. [36] used the sample size of 396 for the study of the psychometric properties of the EQ-5D-3L and EQ-5D-5L instruments in psoriasis, and Bhadhuri et al. [37] included 224 patients in the psychometric analyses. Considering that many studies on comparing EQ-5D-3L and EQ-5D-5L used a sample size of 500 or less [15, 36–39], we used a sample size of 500 in our study, and this number allows for robust analysis within the groups of interest. Therefore, in the online survey of China, 500 respondents with and without a selected health condition were recruited to complete the E-QALY items, EQ-5D-3L, EQ-5D-5L and the Short Warwick–Edinburgh Mental Well-being Scale (SWEMWS). This data was collected between April and July 2019 online by Accent, a U.K online survey company. Quotas and inclusion criteria were applied to recruit a sample of 500 individuals in which there were similar numbers of individuals with GAD, HIV/AIDS, CHB, or depression, or without any of those 4 chronic conditions. The sample was broadly representative to the country in terms of geography, ethnicity and gender. The study was approved by the Ethic Committee of University of Sheffield, United Kingdom (Approval letter number 025524) and the IRB of Jinan University, China (Approval letter number JNUKY-2020-001), and all methods were performed in accordance with the relevant guidelines and regulations. Informed consent was obtained from all participants prior to the online survey through panel, the type was electronic.

The online survey began by giving an outline of the research purpose. Participants were then asked to report their disease history. Eligible respondents reported their background information including education level, gender and age etc. Next, respondents were asked to respond to a battery of questionnaires including (in the order of) a subset of E-QALY items, the 3L/5L (half of the sample responded to 5L and the other half 3L), some more E-QALY items, the SWEMWS, the 5L/3L, and the EQ-VAS.

**Instruments**

The EQ-5D-3L and EQ-5D-5L were both preference-based HRQoL instruments developed by the EuroQol Group. Both instruments have the same five health dimensions, i.e., mobility, self-care, usual activities, pain/discomfort, and anxiety/depression. The difference is the 3L has three response levels (no problems, some problems, extreme problems), while the 5L has five response levels (no problems, slight problems, moderate problems, severe problems, and unable/extreme problems) for each dimension. To calculate the utility, we used the value set developed by Liu et al. for the 3L [40] and the value set developed by Luo et al. for the 5L [41].

**Statistical analyses**

We first described the characteristics of our sample and we reported, by condition groups (1) the median response distributions; (2) sensitivity to mild health problems as measured by the percentages of reporting ‘no problems’; (3) the distributions of the 3L and 5L utility scores; and (4) the known-group validities when compared with the healthy group and the relative efficiency between the 5L and 3L. Data were analyzed using Stata for Windows, Version 14.0 MP, and IBM SPSS Statistics for Windows, Version 22.0. Armonk, NY: IBM Corp (2013).

Given the skewed distribution of EQ-5D responses [42], the median responses were reported to understand the overall health state of each group. Then, the percentages of reporting ‘no problems’ for each dimension if often being referred as ‘ceiling effects’ in published studies. This is not accurate given it is unknown the reasons of this large proportions of reporting ‘no problems’. For this reason, we refer this phenomenon as sensitivity to mild health problems, which was defined as the proportion of respondents indicating ‘no problems’ in each dimension and all five dimensions taken together [43]. Previous studies have shown that the use of 5L could reduce the responses of reporting ‘no problems’ and was considered as a more sensitive measure [16, 18]. For this reason, we hypothesized that the 5L has better sensitivity and calculated the reduction of reporting ‘no problems’ from 3 to 5L. We evaluated the reduction separately for each dimension and all five dimensions taken together. Next, the index values of each group were calculated using 3L and 5L value sets respectively and the distributions were plotted.

Known-group validity of the two EQ-5D index scores was evaluated using the analysis of variance (ANOVA) tests. Relative efficiency was calculated as the ratio of F statistics derived from the ANOVA analysis. The F statistic was widely used to assess RE of measurement scales [19, 44, 45]. The index score with a higher F statistic would be deemed to be more efficient than its comparator since a higher value of F statistic is more likely to result in statistical
significance. To understand the RE of the index scores, we compared the distributions of the responses to the EQ-5D dimensions between the healthy group and each of the condition group. Mann–Whitney test was used. For reference, we listed the median values of each dimension reported by the healthy group and 4 condition groups.

**Results**

In total, 500 individuals completed the online survey, including 140 healthy individuals, 122 individuals with CHB, 107 with depression, 90 individuals with GAD and 101 with HIV/AIDS. Some respondents reported multiple conditions, e.g., 68 individuals reported both depression and GAD. In general, the whole study sample was young (mean age: 35.8, SD: 8.64) and well educated. The gender proportions of the five groups were generally balanced except for the group of HIV/AIDS, in which, about 87.1% of individuals were female. In terms of the age distribution, the healthy group was mostly young; the CHB group has more participants aged between 40 and 49; the depression and GAD groups had individuals from all four age groups and the HIV/AIDS group aged mainly from 30 to 49. Individuals with tertiary education accounted for over 80% for all four disease groups and the healthy group had more individuals with secondary education. Table 2 shows the demographic information by condition.

There were 13, 34, 28, 26 and 26 unique states reported for the healthy, CHB, depression, GAD and HIV/AIDS group respectively for the 3L. The corresponding numbers were 18, 43, 46, 42 and 35 for the 5L. When measured by the 3L, the median responses of healthy and HIV/AIDS groups were ‘no problems’ across the five dimensions. In comparison, when measured by the 5L, while the median responses for the healthy group remained to be ‘no problems’, the median responses of HIV/AIDS group were all ‘slight problems’. Similarly, the median responses for the CHB group were ‘slight problems’ for the last three dimensions when measured by the 5L, while the only pain/discomfort dimension had ‘moderate problem’ measured by the 3L. The median responses of the last two dimensions for the GAD and depression group were both on the second level for the 3L (moderate problems) and 5L (slight problems). The Mann–Whitney results were all significant at 0.01 level suggesting all 4 condition groups had a different distribution of responses against the healthy group, for both 3L and 5L.

It is evident from Table 3 that the percentage of reporting ‘no problems’ of the 5L was smaller than the 3L for all dimensions and for all condition groups. When all dimensions are considered, the number of health profiles 11111 decreased by almost 40% when reporting using the 5L. The reduction of reporting ‘no problems’ is more salient in the pain/discomfort and anxiety/depression dimensions. The most prominent difference was observed in the HIV/AIDS group, i.e., all dimensions had a relative reduction of over 30%.

For the 3L, the healthy group had a mean utility of 0.948 (SD: 0.104), followed by HIV/AIDS of 0.791 (SD: 0.116), chronic hepatitis B of 0.781 (SD: 0.166), depression of 0.721 (SD: 0.180) and GAD of 0.711 (SD: 0.202). In comparison, the mean utilities of the 5L were 0.947 (SD: 0.080) for the healthy group, 0.785 (SD: 0.180) for the CHB group, 0.779 (SD: 0.115) for the HIV/AIDS group, 0.746 (SD: 0.202) for the depression group and 0.718 (SD: 0.217) for the GAD group.

| Table 2  | Sample characteristics | Conditions |
|----------|------------------------|------------|
| Characteristics | Whole sample | Healthy | Chronic Hepatitis B | Depression | GAD | HIV/AIDS |
| N (%) | N (%) | N (%) | N (%) | N (%) | N (%) | N (%) |
| Subgroup sample size | 500 | 140 | 122 | 107 | 90 | 101 |
| Gender | | | | | | | |
| Male | 200 (40) | 68 (48.6) | 53 (43.4) | 56 (52.3) | 42 (46.7) | 13 (12.9) |
| Female | 300 (60) | 72 (51.4) | 69 (56.6) | 51 (47.7) | 48 (53.3) | 88 (87.1) |
| Age group (years) | | | | | | | |
| 18–29 | 121 (24.2) | 61 (43.6) | 17 (13.9) | 40 (37.4) | 21 (23.3) | 1 (1.0) |
| 30–39 | 196 (39.2) | 59 (42.1) | 39 (32.0) | 37 (34.6) | 42 (46.7) | 50 (49.5) |
| 40–49 | 157 (31.4) | 15 (10.7) | 62 (50.8) | 18 (16.8) | 19 (21.1) | 50 (49.5) |
| ≥50 | 26 (5.2) | 5 (3.6) | 4 (3.3) | 12 (11.2) | 8 (8.9) | 0 (0) |
| Education level | | | | | | | |
| Secondary Education | 44 (8.8) | 20 (14.3) | 6 (4.9) | 5 (4.7) | 1 (1.1) | 4 (4.0) |
| Undergraduate education | 415 (83.0) | 104 (74.3) | 107 (87.7) | 87 (81.3) | 74 (82.2) | 97 (96.0) |
| Postgraduate education | 41 (8.2) | 16 (11.4) | 9 (7.4) | 15 (14.0) | 15 (16.7) | 0 (0) |

GAD Generalized anxiety disorder
Table 3  Percentages of reporting no problems between 3L and 5L across 5 groups.

| Dimension | Healthy (N=140) | Chronic Hepatitis B (N=122) | Depression (N=140) | GAD (N=90) | HIV/AIDS (N=101) |
|-----------|-----------------|-----------------------------|-------------------|------------|-----------------|
| Mobility | 89.3 (91.4) (%) | 55.7 (2.3) (%) | 61.7 (16.5) (%) | 66.7 (54.7) (%) | 44.3 (30.3) (%) |
| Self-care | 90.7 (92.2) (%) | 58.2 (2.3) (%) | 62.4 (16.5) (%) | 67.8 (54.7) (%) | 43.4 (30.3) (%) |
| Usual activities | 82.9 (84.3) (%) | 45.9 (1.7) (%) | 66.4 (54.7) (%) | 62.2 (54.7) (%) | 37.4 (30.3) (%) |
| Pain/discomfort | 60.7 (81.4) (%) | 19.7 (25.4) (%) | 20.2 (54.7) (%) | 37.4 (54.7) (%) | 22.4 (30.3) (%) |
| Anxiety/depression | 55.0 (70.7) (%) | 22.2 (24.6) (%) | 26.2 (54.7) (%) | 37.4 (54.7) (%) | 22.4 (30.3) (%) |
| Overall (1111) | 44.3 (62.6) (%) | 30.3 (14.3) (%) | 37.4 (14.3) (%) | 62.6 (14.3) (%) | 43.3 (14.3) (%) |

Fig. 1  Distributions of utility score for the whole sample

Fig. 2  Distributions of utility score for healthy group

Fig. 3  Distributions of utility score for chronic hepatitis B group
group. As shown in Figs. 1, 2, 3, 4, 5 and 6, the 5L index value had a smoother distribution, less clustering and wider range (−0.305 to 1) compared with the 3L (ranged from −0.03 to 1). An exception is the 5L utility score clustered at 0.734 for the HIV/AIDS group. This is the utility score of health state 22222, which had been reported most frequently in the HIV/AIDS group.

Table 4 shows the mean index values of by groups and the ANOVA tests between the healthy group and the four condition groups. The last column shows the relative efficiency between the 5L and 3L among these four comparisons. The index values of the 5L ranged from 0.711 of the GAD group to 0.948 of the healthy group. In comparison, the index values of the 3L ranged from 0.718 of the GAD group to 0.947 of the healthy group. It is clear that two versions of EQ-5D produced comparable index values for each sub-group and both versions demonstrated good known-group validity, except that the 3L did not show a statistically significant result in the comparison of the healthy and HIV/AIDS groups. The relative efficiency of the 5L index was higher in all four comparisons.

**Discussion**

Our study used both the 3L version and the 5L version of EQ-5D to measure the HRQoL burden of 4 chronic conditions in China and focused on comparing the measurement sensitivity of these two EQ-5D versions. In general, we found both versions of EQ-5D to be sensitive tools to quantify the HRQoL loss caused by the 4 chronic conditions, but the 5L showed an improved sensitivity to pick up mild health problems. When combined with the life expectancy data, quality-adjusted life years (QALYs) can be calculated and can be used as a standard measure to reflect how a condition could affect the length and the quality of an individual, which provided a single metric to reflect disease burden. Disease burden is typically measured using disability-adjusted life years (DALYs), but QALYs can also be used and may be a better measure as it could provide individual-level HRQoL-based data and it is a recommended measure for economic evaluations [46].

For the measurement properties, our results generally agree with the findings reported before, that is, although both versions produced highly agreed responses, index values and both had good known-group validity, the 5L performed better in terms of response distribution, sensitivity to mild health problems, index value distribution and had higher relative efficiency [16, 47, 48]. The less clustering of reporting ‘no problems’ for the 5L is most evident for the pain/discomfort and anxiety/depression dimensions. Previous studies referred this issue of clustering at ‘no problems’ as ceiling effects, but it should be noted that there are two kinds of responses in play, first, the respondents that do not have any problems and second, the respondents do have problems but report ‘no problems’ anyway. Theoretically, ceiling effects
exist when the latter kind of responses contributed the most to the clustering at ‘no problems’. Our results showed that the relative reduction is smaller for the healthy group, which suggests the large proportion of reporting ‘no problems’ is not a ceiling effects, but a genuine reflection of the health state of the healthy group. In overall, around 40% of relative reduction of reporting ‘no problems’ was observed when five dimensions were taken together. This was larger than the numbers reported in other studies, which ranged between 6.9 and 33.7% [16, 47]. Overall, this shows the limitation of less cutting-off points provided by the 3L descriptive system, which was first reported by Mathieu F et al. [48]. Hence, the 5L is more sensitive in measuring the HRQoL of individuals with mild health conditions.

For the index value distributions, it was observed that the 3L had more gaps and clustering’s than the 5L. Two major factors are in play to decide the distribution of the index value, i.e., the health profiles and characteristics of the value set [49]. Notably in the 3L, there was a large gap between the index value 1 (profile 11111) and the second highest index value 0.887 (profile 11211). In the 5L, there are 5 different profiles worse than 11111 but having an index value higher than 0.9. In terms of the clusterings, 3L resulted in more clusterings than the 5L and the reason for the clusterings of the 3L index values is due to the clusterings of the profiles [49]. In comparison, there were still some clusterings in the 5L distributions, but it is not due to the clusterings of profiles, instead, it is because more profiles were reported and some profiles have similar index values. In overall, the increased levels of the 5L defined more health states and provided more subtle index values.

The relative efficiency results favored the 5L and is in line with the study of You et al. [16]. Based on the F-ratios in Table 4, it appears that 5L is more sensitive than 3L in physical diseases (i.e., CHB and HIV/AIDS). This is mainly because in the two mental condition group (i.e., GAD and depression), the profiles were more similar, that is, the median levels for both 3L and 5L were both 11122 for the mental condition groups. In contrast, the median levels of the HIV/AIDS group were 22222 when used the 5L and were 11111 when used the 3L. This was also observed in the CHB group, where the median level of usual activities was slight problems in the 5L, but no problems in the 3L. The 5L could pick up more subtle problems than the 3L, and when used the 3L, respondents tent to report no problems. The minimal advantage of 5L compared to 3L in differentiating individuals with and without a mental disorder could be due to the weakness of EQ-5D in measuring mental health. There is only one item in the instrument targeting mental problems.

There are some limitations in this study. First, the study sample was young and highly educated. It should be due to the fact that old people and less educated people are less active on Internet. Therefore, findings of this study may not be generalizable to older populations. It has been reported the elderly with less education attainment may have more difficulty to use the five-level descriptive system and in such occasions, the 3L may be a more feasible alternative [33]. Second, some respondents reported more than one condition, but we did not provide deep analysis about the possible effect of multi-conditions. It should be noted that our sample which was recruited online and the health condition were self-reported. Ideally, clinical data are used to verify the presence and absence of diagnoses reported by the study subjects. Due to these limitations, the superiority of the 5L warrants further studies.

**Table 4** Relative efficiency of the EQ-5D-3L and the EQ-5D-5L index scores

| Comparison | n  | EQ-5D-3L | | | | | EQ-5D-5L | | | | | | | | F5L/F3L |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Healthy | 140 | 0.948 | 1.000 | 0.364 | 0.104 | – | – | 0.947 | 1.000 | 0.385 | 0.080 | – | – | – |
| Healthy vs Chronic Hepatitis B | 122 | 0.781 | 1.000 | 0.344 | 0.166 | 26.070 | 0.000 | 0.785 | 1.000 | 0.155 | 0.180 | 59.925 | 0.000 | 2.299 |
| Healthy vs Depression | 107 | 0.721 | 1.000 | 0.245 | 0.180 | 46.346 | 0.000 | 0.746 | 1.000 | 0.085 | 0.202 | 51.847 | 0.000 | 1.119 |
| Healthy vs GAD | 90 | 0.711 | 1.000 | 0.245 | 0.178 | 36.612 | 0.000 | 0.718 | 1.000 | 0.085 | 0.217 | 58.762 | 0.000 | 1.605 |
| Healthy vs HIV/AIDS | 101 | 0.791 | 1.000 | 0.364 | 0.116 | 0.998 | 0.319 | 0.779 | 1.000 | 0.249 | 0.115 | 5.653 | 0.018 | 5.664 |

GAD Generalized anxiety disorder

**Conclusions**

In this study, both EQ-5D versions could quantify the disease burden of both physical and mental diseases in terms of self-reported HRQoL. Overall, our findings favor the 5L version of EQ-5D for measuring health-related quality of life burden caused by chronic disease in China. As patient-reported outcome (PRO) is increasingly being used all over the world to measure disease burden, EQ-5D can be a useful tool in this context.

**Appendix**

See Table 5.
### Table 5 Measurement properties of EQ-5D-3L and EQ-5D-5L from published studies

| Measurement properties | EQ-5D-3L | EQ-5D-5L | 3L vs. 5L |
|------------------------|----------|----------|-----------|
| Sensitivity            | Acceptable in patients with stable angina pectoris (SAP) [6] | Satisfactory because it was able to detect statistical differences in treatment modalities in patients with adolescent idiopathic scoliosis (AIS) [7] | Compared to the 3L version, the ceiling effect decreased in 5L in patients with common cancers, hepatitis B, elderly Chinese population, patients with acute myeloid leukemia (AML), hypertensive, diabetes, and general population [13–21] |
| Ceiling effect         | Serious ceiling effects were observed in patients with systemic lupus erythematosus (SLE) [8] and chronic diseases [9] | Ceiling effects of all dimensions were significant in patients with chronic hepatitis B (CHB) [10], AIS [7] and spondyloarthriti (SpA) [11] but acceptable in patients with haemophilia [12] | Both 3L and 5L had good responsiveness because of their ability to detect changes of patients’ health status and no significant difference between 3 and 5L was observed [14] |
| Responsiveness         | Good responsiveness was observed in differentiating patients with heart failure and without heart failure [22] | The responsiveness to changes of patients’ health condition was satisfactory in patients with SpA [23], AIS [24] and CHB [10] | The 5L version had better discriminatory power than 3L [13–16, 18, 20, 21, 25] |
| Informativity          | Strong informativity of EQ-5D-5L has been confirmed since it could differentiate haemophilia patients with different health condition [12] | The 5L version had better acceptability because most patients chose 5L as the easier system to answer [14] |
| Acceptability          | Good in patients with SLE [8], chronic kidney disease (CKD) [26], chronic heart failure (CHF) [22], and Uygur cervical cancer patients [27], but moderate in knee osteoarthritis (KOA) [28] | The health descriptive system showed good construct validity in southwest residents [29] and health population in China [30] |
| Construct validity     | Satisfactory in patients with SAP [6] and SLE [8] but only acceptable in Kashin-beck disease (KBD) [31] | Good discriminant validity has been confirmed in patients with southwest residents [29] and family caregivers of leukemia patients [32] |
| Discriminant validity  | Moderate correlation coefficient indicated that 3L had an acceptable convergent validity in patients with SAP [6] and KBD [31] | 5L had a good convergent validity in various populations because moderate to strong correlation coefficient was found in patients with chronic hepatitis disease [10], haemophilia [12], southwest residents [29], family caregivers of leukemia patients [32] |
| Convergent validity    | Good known-group validity was confirmed because lower scores were gained in patients with worse health status [31] | Both 3L and 5L had good convergent validity in patients with common cancer, hepatitis B and AML. Statistics suggested that 5L is slightly stronger than 3L [13–15, 17] |
| Face validity          | Good known-group validity was confirmed because lower scores were gained in patients with worse health status [31] | Both 3L and 5L had good known group validity. Three studies show that 5L performed slightly better than 3L [14, 15, 17], while higher efficiency was observed in 3L in discriminating between known groups in rural hypertensive patients [18] |
| Known-group validity   | Desirable in patients with KOA [28] and Uygur cervical cancer patients because strong correlation was found between all domains and total score [27] | Barely satisfactory in rural residents due to their poor comprehensibility of some dimensions [33] |
| Content validity       | Desirable in patients with KOA [28] and Uygur cervical cancer patients because strong correlation was found between all domains and total score [27] | Barely satisfactory in rural residents due to their poor comprehensibility of some dimensions [33] |
Comparing EQ-5D-3L and EQ-5D-5L in measuring the HRQoL burden of 4 health conditions in China

Sample size and type of each reference:
1. $n = 411$, stable angina pectoris
2. $n = 227$, adolescent idiopathic scoliosis
3. $n = 214$, systemic lupus erythematosus
4. $n = not applicable$, Population with Chronic Disease
5. $n = 645$, chronic hepatitis B
6. $n = 220$, spondyloarthritis
7. $n = 875$, haemophilia
8. $n = 1802$, breast cancer ($n = 601$), colorectal cancer ($n = 601$), lung cancer ($n = 600$)
9. $n = 645$, hepatitis B
10. $n = 416$, lung cancer ($n = 93$), breast cancer ($n = 88$), liver cancer ($n = 67$), colorectal cancer ($n = 60$), thyroid cancer ($n = 57$), Gastric cancer ($n = 51$)
11. $n = 648$, elderly Chinese population
12. $n = 168$, acute myeloid leukemia
13. $n = 11412$, hypertension
14. $n = 289$, diabetes
15. $n = 1934$, general population
16. $n = 1017$, hypertension ($n = 241$), diabetes ($n = 241$), chronic hepatitis ($n = 240$), general population ($n = 295$)
17. $n = 180$, chronic heart failure
18. $n = 113$, spondyloarthritis
19. $n = 227$, adolescent idiopathic scoliosis
20. $n = 324$, general population
21. $n = 441$, chronic kidney disease
22. $n = 60$, cervical cancer
23. $n = 141$, knee osteoarthritis
24. $n = 12950$, residents in Southwest China
25. $n = 324$, Chinese health population
26. $n = 362$, kashin–Beck Disease
27. $n = 298$, family caregivers of leukemia patients
28. $n = 62$, rural residents
29. $n = 205$, stroke

Acknowledgements Not applicable.

Author contributions Guizhi Weng contributed to acquisition of data. Zhihao Yang and Yanming Hong analyzed and interpreted the data. Guizhi Weng draft and revise the manuscript. Sha Li, Jie Jiang and Nan Luo contributed to conception, design and the interpretation of the results. Clara Mukuria contributed to the study design and data analysis. All authors reviewed and approved the final version.

Funding We thank Natural Science Foundation of Guangdong Province (No. 2020A1515010990) to SL and Science and Technology Program of Guangzhou, China (No. 201704020198) for funding this study.

Availability of data and materials Not applicable.

Code availability Not applicable.
Declarations

Conflict of interest The authors have no relevant financial or non-financial interests to disclose.

Ethics approval The study was approved by the Ethic Committee of University of Sheffield, United Kingdom. And the committee’s reference number was 025524.

Consent to participate Written, informed consent was obtained from all participants included in the study.

Consent for publication Manuscript is approved by all authors for publication.

References

1. Devlin, N.J., Brooks, R.: EQ-5D and the EuroQol Group: past, present and future. Appl. Health Econ. Health Policy. 15(2), 127–137 (2017). https://doi.org/10.1007/s10051-017-0310-5
2. Szende, A., Janssen, B.: Socio-demographic indicators based on EQ-5D. In: Szende, A., Janssen, B., Cabases, J. (eds.) Self-Reported Population Health: An International Perspective Based on EQ-5D, pp. 37–46. Springer, Dordrecht (2014)
3. Devlin, N.J., Krabbe, P.F.: The development of new research methods for the valuation of EQ-5D-5L. Eur. J. Health Econ. 14(Suppl. 1), S1-3 (2013). https://doi.org/10.1007/s10051-013-0502-3
4. Janssen, M.F., Pickard, A.S., Golicki, D., Guex, C., Niewada, M., Scalone, L., et al.: Measurement properties of the EQ-5D-5L compared to the EQ-5D-3L across eight patient groups: a multicountry study. Qual. Life Res. 22(7), 1717–1727 (2013). https://doi.org/10.1007/s11136-011-0322-4
5. Herdman, M., Guex, C., Lloyd, A., Janssen, M.F., Kind, P., Parkin, D., et al.: Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L). Qual. Life Res. 20(10), 1727–1736 (2011). https://doi.org/10.1007/s11136-011-9903-x
6. Han, Y., Wu, J., Cong, H., Zhou, J., Xu, C.: Validity and sensitivity of the SF-6D and EQ-5D in Chinese patients with stable angina pectoris. Chin. J. Health Stat. 30(06), 829–832 (2013). (in Chinese)
7. Cheung, P.W.H., Wong, C.K.H., Samartzis, D., Luk, K.D.K., Lam, C.L.K., Cheung, K.M.C., et al.: Psychometric validation of the EuroQol 5-Dimension 5-Level (EQ-5D-5L) in Chinese patients with adolescent idiopathic scoliosis. Scoliosis Spinal Disorder. 11, 19 (2016). https://doi.org/10.1186/s13013-016-0083-x
8. Wang, S.L., Wu, B., Zhu, L.A., Leng, L., Bucala, R., Lu, L.J.: Construct and criterion validity of the Euro Qol-5D in patients with systemic lupus erythematosus. PLoS One 9(6), e98833 (2014). https://doi.org/10.1371/journal.pone.0098833
9. Yao, J., Liang, Y., Xu, Y., Guo, L., Yang, L., Pan, J., et al.: Analysis of the measurement characteristics of Chinese version of EQ-5D-3L applied in population with chronic disease: Rasch Model. Chin. J. Health Stat. 35(05), 681–683 (2018). (in Chinese)
10. Jia, Y., Cui, F., Zhang, D., Zhang, G., Gong, X., Wang, F., et al.: Investigation on the applicability of EQ-5D-5L in Chinese Patients with chronic hepatitis B. Chin. J. Vaccines Immunol. 19(06), 492–497 (2013). (in Chinese)
11. Tsang, H.H.L., Cheung, J.P.Y., Wong, C.K.H., Cheung, P.W.H., Lau, C.S., Chung, H.Y.: Psychometric validation of the EuroQol 5-dimension (EQ-5D) questionnaire in patients with spondyloarthropathy. Arthritis Res. Ther. 21(1), 41 (2019). https://doi.org/10.1186/s13075-019-1826-x
12. Xu, R., Dong, D., Luo, N., Wong, E., Wu, Y., Yu, S., et al.: Evaluating the psychometric properties of the EQ-5D-5L and SF-6D among patients with haemophilia. Eur. J. Health Econ. 22(4), 547–557 (2021). https://doi.org/10.1007/s10198-021-01273-5
13. Zhu, J., Yan, X., Liu, C., Wang, H., Wang, L., Cao, S., et al.: Comparing EQ-5D-3L and EQ-5D-5L performance in common cancers: suggestions for instrument choosing. Qual. Life Res. 30(3), 841–854 (2021). https://doi.org/10.1007/s11136-020-02636-w
14. Jia, Y., Cui, F., Li, L., Zhang, D., Zhang, G., Wang, F., et al.: Comparison between the EQ-5D-5L and the EQ-5D-3L in patients with hepatitis B. Qual. Life Res. 28(3), 2355–2363 (2014). https://doi.org/10.1007/s11136-013-0670-3
15. Zeng, X., Sui, M., Liu, Y., Yang, H., Liu, R., Tan, R., et al.: Measurement properties of the EQ-5D-5L and EQ-5D-3L in six commonly diagnosed cancers. Patient 14(2), 209–222 (2021). https://doi.org/10.40272/2014-00466-z
16. You, R., Liu, J., Yang, Z., Pan, C., Ma, Q., Luo, N.: Comparing the performance of the EQ-5D-3 L and the EQ-5D-5 L in an elderly Chinese population. Health Qual. Life Outcomes 18(1), 97 (2020). https://doi.org/10.1186/s12959-020-01324-0
17. Yu, H., Zeng, X., Sui, M., Liu, R., Tan, R., Yang, J., et al.: A head-to-head comparison of measurement properties of the EQ-5D-3L and EQ-5D-5L in acute myeloid leukemia patients. Qual. Life Res. 30(3), 855–866 (2021). https://doi.org/10.1007/s11136-020-02644-w
18. Jiang, J., Hong, Y., Zhang, T., Yang, Z., Lin, T., Liang, Z., et al.: Comparing the measurement properties of the EQ-5D-5L and the EQ-5D-3L in hypertensive patients living in rural China. Qual. Life Res. 30(7), 2045–2060 (2021). https://doi.org/10.1007/s11136-021-02786-5
19. Pan, C., Sun, H., Wang, X., Ma, Q., Xu, Y., Luo, N., et al.: The EQ-5D-5L index score is more discriminative than the EQ-5D-3L index score in diabetes patients. Qual. Life Res. 24(7), 1767–1774 (2015). https://doi.org/10.1007/s11136-014-0902-6
20. Zhou, T., Guan, H., Liu, G., Ma, A.: Comparison analysis on the EQ-5D-5L and the EQ-5D-3L for general population in China. Chin. Health Econ. 35(03), 17–20 (2016). (in Chinese)
21. Cai, Y., Wu, H., Yang, Z.: Comparison analysis of European quality of life-5 dimensions with three-level and five-level based on four types of people in China. Chin. J. Dis. Control Prev. 24(10), 1190–1194 (2020). (in Chinese)
22. Zhao, Z., Xu, S., Wang, X., Zhang, P., Bi, Y., Li, D., et al.: Scientific evaluation of the European Five-dimensional Health Scale inpatients with chronic heart failure. Chin. J. Tradit. Chin. Med. Pharm. 34(04), 1715–1718 (2019). (in Chinese)
23. Tsang, H.H.L., Cheung, P.W.H., Lau, C.S., Chung, H.Y., Cheung, J.P.Y.: Responsiveness of the EuroQol 5-Dimension (EQ-5D) questionnaire in patients with spondyloarthritis. BMC Musculoskelet. Disord. 22(1), 439 (2021). https://doi.org/10.1186/s12891-021-04315-4
24. Cheung, P.W.H., Wong, C.K.H., Lau, S.T., Cheung, J.P.Y.: Responsiveness of the EuroQol 5-dimension (EQ-5D) in adolescent idiopathic scoliosis. Eur. Spine J. 27(2), 278–285 (2018). https://doi.org/10.1007/s00586-017-5330-1
25. Xing, Y., Ma, A.: An empirical study comparing the Chinese versions of EQ-5D-3L and EQ-5D-5L. Shanghai Med. Pharma. J. 22(2), 278–285 (2018). https://doi.org/10.1007/s11136-014-0670-3
26. Yu, H., Shen, W., Jia, M., Yu, G., Mei, X., Bian, Q., et al.: Reliability and validity of EuroQol 5-dimension (EQ-5D) in patients with chronic kidney disease. Chin. J. Integr. Tradit. Western Nephrol. 13(08), 690–693 (2012). (in Chinese)
27. Hu, X., Ding, M., Zhang, M., Yan, X., Huang, Y.: The reliability and validity of EQ-5D-3L in Uygur cervical cancer patients. J. Shihezi Univ. (Nat. Sci.). 37(06), 780–785 (2019). (in Chinese)
28. Qiu, Y.: Reliability and validity of the Chinese Version of EQ-5D in the knee osteoarthritis [Master, Capital Institute of Physical Education]. (2015). (in Chinese)

29. Xia, J., Wu, N., Ma, T., Yu, C., Li, N.: Evaluation of reliability and validity of EQ-5D-5L based on residents in Southwest China. J. Sichuan Univ. (Med. Sci. Ed.) 51(05), 691–694 (2020). (in Chinese)

30. Xing, Y., Ma, A.: Study on reliability and validity of Chinese version of EQ-5D-5L. Shanghai Med. Pharma. J. 34(09), 40–43 (2013). (in Chinese)

31. Fang, H., Farooq, U., Wang, D., Yu, F., Younus, M.I., Guo, X.: Reliability and validity of the EQ-5D-3L for Kashin-Beck disease in China. Springerplus 5(1), 1924 (2016). https://doi.org/10.1186/s40064-016-3613-3

32. Li, L., Liu, C., Cai, X., Yu, H., Zeng, X., Sui, M., et al.: Validity and reliability of the EQ-5D-5L in family caregivers of leukemia patients. BMC Cancer 19(1), 522 (2019). https://doi.org/10.1186/s12885-019-5721-2

33. Yang, F., Jiang, S., He, X., Li, H., Wu, H., Zhang, T., et al.: Do rural residents in China understand EQ-5D-5L as intended? Evidence from a qualitative study. Pharmacoecocon. Open S1(1), 101–109 (2021). https://doi.org/10.1007/s41669-020-00212-z

34. Dai, L., Wen, Y., Lu, T., Cai, J., Wang, L., Qiu, M., et al.: Measurement characteristics evaluation of SF-HSTCM and EQ-5D-3L among patients with stroke. J. Cap. Med. Univ. 40(04), 627–634 (2019). (in Chinese)

35. Peasgood, T., Makuria, C., Brazier, J., Marten, O., Kreimeier, S., Luo, N., et al.: Developing a new generic health and wellbeing measure: psychometric survey results for the EQ health and wellbeing. Value Health. (2022). https://doi.org/10.1016/j.jval.2021.11.1361

36. Yfantopoulos, J., Chantzaras, A., Kontodimas, S.: Assessment of the psychometric properties of the EQ-5D-3L and EQ-5D-5L instruments in psoriasis. Arch. Dermatol. Res. 309(5), 357–370 (2017). https://doi.org/10.1007/s00403-017-1743-2

37. Bhadhuri, A., Kind, P., Salari, P., Jungo, K.T., Boland, B., Byrne, S., et al.: Measurement properties of EQ-5D-3L and EQ-5D-5L in recording self-reported health status in older patients with substantial multimorbidity and polypharmacy. Health Qual. Life Outcomes 18(1), 317 (2020). https://doi.org/10.1186/s12955-020-01564-0

38. Poor, A.K., Renz, F., Brodsky, V., Gulacsi, L., Beretzky, Z., Hidvegi, B., et al.: Measurement properties of the EQ-5D-5L compared to the EQ-5D-3L in psoriasis patients. Qual. Life Res. 26(12), 3409–3419 (2017). https://doi.org/10.1007/s11136-017-1699-x

39. Gandhi, M., Ang, M., Teo, K., Wong, C.W., Wei, Y.C.H., Tan, R.L.Y., et al.: EQ-5D-5L is more responsive than EQ-5D-3L to treatment benefit of cataract surgery. Patient-Patient Cent. Outcomes Res. 12(4), 383–392 (2019)

40. Liu, G.G., Wu, H., Li, M., Gao, C., Luo, N.: Chinese time trade-off values for EQ-5D health states. Value Health 17(5), 597–604 (2014). https://doi.org/10.1016/j.jval.2014.05.007

41. Luo, N., Liu, G., Li, M., Guan, H., Jin, X., Rand-Hendrikse, K.: Estimating an EQ-5D-5L value set for China. Value Health 20(4), 662–669 (2017). https://doi.org/10.1016/j.jval.2016.11.016

42. Parkin, D., Devlin, N., Feng, Y.: What determines the shape of an EQ-5D index distribution? Med. Decis. Making 36(8), 941–951 (2016). https://doi.org/10.1177/0272989X16645581

43. Terwee, C.B., Bot, S.D.M., de Boer, M.R., van der Windt, D.A.W.M., Knol, D.L., Dekker, J., Bouter, L.A., et al.: Quality criteria were proposed for measurement properties of health status questionnaires. J. Clin. Epidemiol. 60(1), 34–42 (2007)

44. Vickrey, B.G., Hays, R.D., Genoves, B.J., Myers, L.W., Ellison, G.W.: Comparison of a generic to disease-targeted health-related quality-of-life measures for multiple sclerosis. J. Clin. Epidemiol. 50(5), 557–569 (1997). https://doi.org/10.1016/s0895-4356(97)00001-2

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

46. Kennedy-Martin, M., Slaap, B., Herdman, M., van Reenen, M., Kennedy-Martin, T., Greiner, W., et al.: Which multi-attribute utility instruments are recommended for use in cost-utility analysis? A review of national health technology assessment (HTA) guidelines. Eur. J. Health Econ. 21(8), 1245–1257 (2020). https://doi.org/10.1007/s10198-020-01195-8

47. Buchholz, I., Janssen, M.F., Kohlmann, T., Feng, Y.S.: A systematic review of studies comparing the measurement properties of the three-level and five-level versions of the EQ-5D. Pharmacoeconomics 36(6), 645–661 (2018). https://doi.org/10.1007/s40273-018-0642-5

48. Janssen, M.F., Bonsel, G.J., Luo, N.: Is EQ-5D-5L better than EQ-5D-3L? A head-to-head comparison of descriptive systems and value sets from seven countries. Pharmacoeconomics 36(6), 675–697 (2018). https://doi.org/10.1007/s40273-018-0623-8

49. Feng, Y., Devlin, N., Bateman, A., Zamora, B., Parkin, D.: distribution of the EQ-5D-5L profiles and values in three patient groups. Value Health. 22(3), 355–361 (2019). https://doi.org/10.1016/j.jval.2018.08.012

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