

RESEARCH

Open Access



Comparing the performance of the EQ-5D-3 L and the EQ-5D-5 L in an elderly Chinese population

Ruxu You¹, Jinyu Liu², Zhihao Yang³, Chenwei Pan⁴, Qinghua Ma^{5*} and Nan Luo^{6*}

Abstract

Background: This study was conducted to compare the validity and discriminative power of both the EQ-5D-3 L and EQ-5D-5 L in an elderly Chinese population with multiple chronic and acute conditions.

Methods: A total of 648 retired people from China (mean \pm standard deviation: 73.3 \pm 6.4 years; male: 55.7%) were recruited and randomized to complete the 3 L or 5 L questionnaire. The 3 L and 5 L were compared in terms of distribution properties, ceiling effects, informativity, validity and discriminatory performance. Convergent validity between the 3 L and 5 L was tested by spearman's rank-order correlation. Discriminatory power was conducted by relative efficiency as assessed by the F statistics.

Results: Most participants answered to "no problems" on both versions of EQ-5D. The 5 L trended towards a slightly lower ceiling compared with the 3 L. The Shannon index improved with the 5 L while the Shannon's Evenness index tended to be similar. Convergent validity was confirmed by the moderate to strong correlation for both 3 L and 5 L. Relative efficiency suggested that 5 L had a higher absolute discriminatory power than the 3 L version in terms of the presence conditions, especially for osteoporosis and metabolic syndrome.

Conclusions: Both the 3 L and 5 L are demonstrated to be valid based HRQoL instruments in Chinese elderly population. The 5 L system may be preferable to the 3 L, as it demonstrated superior performance with respect to lower ceiling effect and better discriminatory power. Further research is needed to examine the responsiveness of the two EQ-5D instruments in this population.

Keywords: EQ-5D-3 L, EQ-5D-5 L, Elderly population, Measurement properties, Quality of life

Background

The EQ-5D is a generic preference-based measure of health related quality of life (HRQOL) that can be used for clinical and economic evaluation of health care as well as for the assessment of population health. It describes a respondent's health status in five dimensions (mobility, self-care, usual activities, pain/discomfort and anxiety/depression) and

produces a utility score anchored by 0 (death) and 1 (full health). The original version of the EQ-5D uses 3 severity levels (no problems, some problems, and extreme problems) to describe each dimension (EQ-5D-3 L). In 2009, EQ-5D-3 L was modified to have 5 response levels for each dimension (EQ-5D-5 L) in order to improve the instrument's sensitivity and reduce ceiling effects [1]. Generally speaking, using measurement instruments with better measurement properties in comparative or evaluative studies is preferred. For example, more sensitive instruments are less likely to miss small but important between-group difference or with-group change. Regardless of the expected

* Correspondence: weitangfangbao@163.com; ephln@nus.edu.sg

⁵The 3rd People's Hospital of Xiangcheng District, 568 Weizhong Road, Suzhou 215134, China

⁶Saw Swee Hock School of Public Health, National University of Singapore, 16 Medical Drive, Block MD3, Singapore 117597, Singapore

Full list of author information is available at the end of the article



© The Author(s). 2020 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

size of the difference or change, a smaller study sample would be needed in hypothesis-testing studies if a more sensitive instrument is used to measure the outcomes. Therefore, effort to improving the measurement properties of the EQ-5D instrument is practically meaningful.

Numerous studies have compared the psychometric performance of the EQ-5D-3 L (hereafter 3 L) and EQ-5D-5 L (hereafter 5 L). A recently published systematic review by Buchholz and colleagues found that compared to 3 L, 5 L has better ceiling effects and descriptive efficiency, as well as similar or better test-retest reliability [2]. It is not clear, however, whether 5 L is more sensitive than 3 L. Sensitivity can be assessed as sensitivity to change (or responsiveness) in longitudinal studies and sensitivity to difference (or discriminatory power) in cross-sectional studies. Buchholz and colleagues found that 5 L was more responsive than 3 L in 2 of 3 studies comparing the two EQ-5D instruments. Similarly, whether 5 L has higher discriminatory power compared to 3 L remains unclear. Although two recent studies in Greece demonstrated marginal to moderate better discriminatory power of 5 L compared to 3 L [3, 4]. Other studies found that 3 L performed better especially in discriminating between healthy and unhealthy populations [5–9]. One possible reason for the mixed results could be that the relative performance of 3 L and 5 L is population specific. Population characteristics such as age, race/culture, education level, language spoken, and health status may affect the performance of HRQOL instruments. For example, it is possible that 5 L is more sensitive than 3 L in patient populations but it is not more sensitive than 3 L in general elderly population because this population is generally healthy and tends to be inaccurate in describing their health due to lower literacy. Therefore, the relative performance of the two instruments in one population may not be generalizable to other populations and should be assessed for different populations individually.

Current research comparing the relative merits of 3 L and 5 L has two issues in study design. First, majority of the studies asked subjects to complete the 3 L and 5 L in one single survey, usually 5 L followed by 3 L. While such a within-subjects design is statistically efficient, it may suffer from the order effect, namely, responses to 3 L may be biased by the experience of responding to 5 L. Second, most of those comparative studies used a ‘crosswalk’ method to calculate the 5 L utility score [10, 11]. The ‘crosswalk’ method is an interim solution. The preferred method is to use a country-specific 5 L value set which was only made available in recent years for some countries. Studies have showed that utility scores derived from the ‘crosswalk’ method and value sets are not entirely equivalent and those may lead to different conclusions in cost-utility analysis [12, 13].

The purpose of the present study was to compare the performance of the 3 L and 5 L in health surveys of the elderly living in the community. The primary study aim was to assess the discriminatory power of the utility scores generated by the two EQ-5D instruments using a between-subjects design.

Methods

This was a questionnaire-based, cross-sectional study of a general elderly population. The study was approved by the Ethics Committee, Medical College of Soochow University and followed the tenets of the Declaration of Helsinki (No20170720).

Recruitment and data collection procedures

A consecutive sample of elderly residents who went to a community health center in a town near Suzhou city for routine health checkup was recruited from September to November 2017. Inclusion criteria were: 60 years or older, 2) ability to understand survey questions, and 3) informed consent.

All consenting participants were invited to a one-on-one, face-to-face interview in an office of the community health center after their health checkup. One community doctor conducted the interview after participating in a training workshop on the study design, interview protocol and recruitment and interview skills. After a brief introduction to the purpose of the study, participants were instructed to complete the following tasks: 1) describe their own health using the 3 L or 5 L questionnaire that was randomly assigned to them, 2) describe their own health using the SF-6D questionnaire, and 3) answer questions assessing socio-demographics and presence or absence of hypertension, hyperlipidemia, arthritis, osteoporosis, other chronic medical condition, and acute medical conditions. The average interview time per respondent was about 25 min.

Instruments

The EQ-5D 3 L and 5 L self-reporting questionnaires describe health in five dimensions including mobility, self-care, usual activities, pain/discomfort, and anxiety/depression.

The 3 L version uses three response options (no problems, moderate problems, and extreme problems), and the 5 L version uses five response options (no problems, slight problems, moderate problems, severe problems, and extreme problems) [2]. The 3 L defines a total of 243 different health states, whereas the 5 L defines 3125 different health states. The Chinese 3 L and 5 L questionnaires have demonstrated good validity in multiple Chinese patient populations [7, 14]. The 3 L and 5 L utility scores were calculated using the 3 L and 5 L value sets for China [15, 16].

Data analysis

The characteristics of participants who completed 3 L and 5 L questionnaires were compared using the chi-square test for categorical variables and the two-sample t test for continuous variables. The ceiling effects, descriptive efficiency, convergent validity, know-groups validity, and discriminatory power of the 3 L and 5 L measures were assessed separately.

Ceiling effects were assessed according to the proportion of participants reporting no problems in each of the EQ-5D dimensions. To facilitate interpretation, we estimated the absolute and relative reduction when going from 3 L to 5 L. The relative reduction (%) was computed by $(\text{Ceiling}_{3\text{L}} - \text{Ceiling}_{5\text{L}}) / \text{Ceiling}_{3\text{L}} \times 100$.

Descriptive efficiency was assessed using the Shannon index (H') and the evenness index (J'). H' is defined as follows:

$$H' = - \sum_{i=1}^L p_i \log_2 p_i$$

Where H' represents the absolute amount of informativity captured, L is the number of levels, and p_i is the proportion of observations in the i^{th} level ($i = 1, \dots, L$). In the case of an even (rectangular) distribution (i.e., all levels are equally filled), H' reaches its maximum that equals $\log_2 L$, which amounts to 1.58 to the 3 L (i.e., $\log_2 3$) and 2.32 to the 5 L (i.e., $\log_2 5$). J' is a complementary index that reflects the extent to which the evenness distribution is achieved, and it is defined as: $J' = H' / H'_{\text{max}}$.

Convergent validity was tested by examining the correlation of the 3 L/5 L utility score with the SF-6D score using the Pearson's correlation coefficient (r). The strength of the correlation was interpreted as strong ($r > 0.50$), moderate ($0.35 \leq r \leq 0.50$), weak ($0.20 \leq r < 0.35$), or absent ($r < 0.20$). Known-groups construct validity was tested by comparing the 3 L/5 L utility scores of participants with and without metabolic syndrome (defined as having one or more of the five medical conditions: central obesity, high blood pressure, high blood sugar, high serum triglycerides, and low serum high-density lipoprotein), osteoporosis, arthritis, other chronic condition (defined as having heart disease, stroke, chronic kidney disease, chronic obstructive pulmonary disease, dementia, cancer, diabetes, stone, tracheitis, and/or cataract), and acute medical condition (defined as having cold, fever, diarrhea, fall, and/or injury). We hypothesized that participants with a chronic or acute condition would have lower utility scores than those without the condition [3].

Discriminatory power of the two EQ-5D utility scores was assessed in terms of their ability to distinguish participants with and without health conditions. The ratio of F statistics derived from the analysis of variance

(ANOVA) tests of the two EQ-5D scores between participants with and without a condition was used. We computed the F statistic ratio in such a way that a ratio higher than 1.0 indicates that 5 L is more discriminative than 3 L.

Data were analyzed by using IBM SPSS Statistics for Windows, Version 22.0. Armonk, NY: IBM Corp (2013). All the statistics were two sided, and $p < 0.05$ was considered statistically significant.

Results

Characteristics of participants

A total of 648 participants were enrolled. The mean (\pm standard deviation) age of the participants was 73.3 ± 6.4 years, with male being 55.7%. The majority of them were not formally educated (94.9%) and low level of income (68.4%). Table 1 shows the full characteristics of the entire sample and participants who completed the 3 L and 5 L questionnaires. There were no significant differences in demographic characteristics between two groups.

Characteristics of responses to the EQ-5D questionnaires

For all of the dimensions, majority of the participants reported no problems (level 1) to both 3 L and 5 L (Table 2). For both 3 L and 5 L, the highest proportion of participants reported no anxiety/depression (3 L: 95.7%; 5 L: 84.3%), whereas the lowest proportion of participants reported no pain/discomfort (3 L: 56.2%; 5 L: 66.7%). Compared to 5 L, 3 L exhibited higher ceiling effects in the anxiety/depression dimension (absolute difference: 11.4%; relative difference: 11.9%). Ceiling effects between 3 L and 5 L were similar in the mobility, self-care and usual activities dimensions (range of absolute difference: -3.8 to 4.0% ; range of relative difference: -5.2 to 5.0%), while lower ceiling effects of 3 L were found in the pain/discomfort dimension (absolute difference: -10.5% ; relative difference: -18.7%). All dimensions considered, the proportion of participants reporting full health (11111) was 52.4 and 41.3% for 3 L and 5 L respectively, representing a reduction of 11.1 and 21.2% from 5 L to 3 L in absolute and relative term.

Descriptive efficiency

The H' values for 5 L were higher than those for 3 L for all dimensions as five levels led to a larger amount of information. The J' values for 5 L were lower than those for 3 L for all but the anxiety/depression dimension (Table 3).

Construct validity

Pearson's correlation coefficient was 0.608 ($p < 0.001$) between 3 L and SF-6D and 0.433 ($p < 0.001$) between 5 L

Table 1 Characteristics of participants

Characteristic	All participants (n = 648)	Participants who completed EQ-5D-3 L (n = 324)	Participants who completed EQ-5D-5 L (n = 324)	P-value
Age				0.244
Mean (SD)	73.25 (6.41)	72.95 (5.78)	73.54 (6.97)	
Median (IQR)	72.0 (8.00)	72.50 (6.00)	72.0 (9.00)	
Range	60.0 to 95.0	61.0 to 93.0	60.0 to 95.0	
Gender				0.937
Male	361 (55.7)	181 (55.9)	180 (55.6)	
Female	287 (44.3)	143 (44.1)	144 (44.4)	
BMI level				0.519
< 25	493 (76.1)	250 (77.2)	243 (75.0)	
≥ 25	155 (23.9)	74 (22.8)	81 (25.0)	
Education				0.372
Low (no/primary/secondary)	615 (94.9)	310 (95.7)	305 (94.1)	
High (tertiary/above)	33 (5.1)	14 (4.3)	19 (5.9)	
Marital status				0.279
Married	603 (93.1)	305 (94.1)	298 (92.0)	
Other (single/widowed/separated/ divorced)	45 (6.9)	19 (5.9)	26 (8.0)	
Household income per month				0.151
< 2000 CNY	443 (68.4)	213 (65.7)	230 (71.0)	
≥ 2000 CNY	205 (31.6)	111 (34.3)	94 (29.0)	

Note EQ-5D EuroQol five-dimensional questionnaire, EQ-5D-3 L (3 L) three-level version of the EQ-5D, EQ-5D-5 L (5 L) five-level version of the EQ-5D

and SF-6D. Generally speaking, participants without a medical condition had higher L/5 L score than those with the medical condition (Table 4). For example, The mean 3 L and 5 L scores for participants without osteoporosis (0.900 and 0.955) were higher than the mean 3 L and 5 L scores for those reporting to have osteoporosis (0.831 and 0.884) ($p < 0.001$).

Discriminatory power

The F-statistics derived from the known-groups comparisons were higher for 5 L than 3 L in each comparison (Table 4), suggesting that 5 L was more efficient or discriminative than 3 L. The F-statistic ratio ranged from 1.160 for comparison of participants with and without acute medical condition to 2.453 for comparison of participants with and without osteoporosis.

Discussion

In this study, we compared the psychometric performance of the 3 L and 5 L EQ-5D instruments in a community-dwelling elderly Chinese population. Both instruments demonstrated validity. Furthermore, the 5 L may be preferable to the 3 L, as it exhibited lower ceiling effect and better discriminatory power. Our study design is unique in that we used an experiment design to compare the two EQ-5D instruments using two different

groups. This is in contrast to the previous one-group design in which all subjects completed both 3 L and 5 L in a cross-sectional survey. Moreover, to the best of our knowledge, this is the first head-to-head comparison of preference-based HRQOL instruments in the fast growing elderly population. Thus, our study provides evidence for the appropriateness of switching from the 3 L EQ-5D to the 5 L EQ-5D for measuring the health outcomes of community-dwelling old people in China. The 3 L EQ-5D has been used in many population health surveys in the country [17–19]. Since the old population is rapidly increasing, the 5 L EQ-5D should be used in health surveys covering the entire adult Chinese population.

Consistent with previous studies [4, 20–23], on general populations, 5 L was found to have smaller ceiling effect as compared to 3 L. For general populations above 18 years old, magnitude of relative difference reported in this study was greater than those reported from Spain (1.62%) [20], England (15.3%) [21], USA (20.5%) [22] and South Korea (6.9%) [23] but less than that reported from Greece [4]. Being the only study conducted on middle-aged general population, the Greek study reported a lower percentage of respondents reporting full healthy on the 3 L (47.0%) and 5 L (31.2%) than our study did (52.4% for the 3 L and 41.3% for the 5 L), yielding a higher absolute (15.8% vs. 11.1%) and relative (33.7% vs.

Table 2 Distributions of responses to EQ-5D-3 L and EQ-5D-5 L items

Dimension	EQ-5D-3 L		EQ-5D-5 L		Absolute difference in ceiling effects	Relative difference in ceiling effects
	Level	n (%)	Level	n (%)		
Mobility	1	240 (74.1)	1	238 (73.5)	0.6	0.8
	2	84 (25.9)	2	74 (22.8)		
	3	0.0	3	9 (2.8)		
			4	0 (0)		
			5	3 (0.9)		
Self-care	1	258 (79.6)	1	245 (75.6)	4.0	5.0
	2	65 (20.1)	2	72 (22.2)		
	3	1 (0.3)	3	7 (2.2)		
			4	0 (0)		
			5	0 (0)		
Usual activities	1	237 (73.1)	1	249 (76.9)	-3.8	-5.2
	2	87 (26.9)	2	67 (20.7)		
	3	0 (0.0)	3	6 (1.9)		
			4	0 (0)		
			5	2 (0.6)		
Pain/discomfort	1	182 (56.2)	1	216 (66.7)	-10.5	-18.7
	2	142 (43.8)	2	103 (31.8)		
	3	0 (0.0)	3	5 (1.5)		
			4	0 (0)		
			5	0 (0)		
Anxiety/depression	1	310 (95.7)	1	273 (84.3)	11.4	11.9
	2	14 (4.0)	2	50 (15.4)		
	3	0 (0.0)	3	1 (0.3)		
			4	0 (0)		
			5	0 (0)		
Full health (all dimensions considered)		170 (52.4)		134 (41.3)	11.1	21.2

Note EQ-5D EuroQol five-dimensional questionnaire

* Relative reduction was defined as ((Ceiling3 L-Ceiling5 L)/Ceiling3 L) in ceiling effects from the EQ-5D-3 L to the EQ-5D-5 L.

Table 3 Classification efficiency of the EQ-5D-3 L and the EQ-5D-5 L classification systems measured by Shannon index (H')

Dimension	3 L		5 L	
	H' (95% CI)	J'(95% CI)	H'(95% CI)	J'(95% CI)
Mobility	0.83	0.52	1.02	0.44
Self-care	0.75	0.48	0.91	0.39
Usual activities	0.84	0.53	0.91	0.39
Pain/discomfort	0.99	0.63	1.01	0.43
Anxiety/depression	0.26	0.16	0.65	0.28

Note CI confidence interval, EQ-5D EuroQol five-dimensional questionnaire, EQ-5D-3 L (3 L) three-level version of the EQ-5D, EQ-5D-5 L (5 L) five level version of the EQ-5D

* Larger H' value indicates higher classification efficiency

21.2%) term respectively. Reported ceiling effect of present study being intermediate to previous studies could be due to factors such as true difference in the health status and cultural difference in answering health questions [24] between target populations.

As expected, extending the EQ-5D descriptive system from three to five levels led to a larger amount of information captured or higher absolute informativity. However, the relative informativity is slightly lower in some dimensions of 5 L, which have also been found in previous comparative studies [25–28]. The absolute and relative informativity of both EQ-5D versions in our study was lower than that in most of previous studies [4, 6, 29]. The relatively good health of our study sample, as evident by the very few endorsements of ‘severe’ or ‘extreme’ health problems with the EQ-5D questionnaires, could partly explain the relatively lower informativity observed in our study.

Table 4 Known-groups validity and the discriminatory power of EQ-5D-3 L and EQ-5D-5 L

	EQ-5D-3 L			EQ-5D-5 L			F _{5L} / F _{3L}
	n (%)	Mean (SD)	F-statistic	n (%)	Mean (SD)	F-statistic	
Metabolic syndrome							
No	66 (20.4)	0.899 (0.144) *	2.742	67 (20.7)	0.952 (0.101) *	5.504	2.007
Yes	258 (79.6)	0.862 (0.169)		257 (79.3)	0.916 (0.113)		
Osteoporosis							
No	180 (55.6)	0.900 (0.152) *	14.635	181 (55.9)	0.955 (0.090) *	35.902	2.453
Yes	144 (44.4)	0.831 (0.172)		143 (44.1)	0.884 (0.124)		
Arthritis							
No	255 (78.7)	0.871 (0.174)	0.194	261 (80.5)	0.932 (0.115)	0.236	1.216
Yes	69 (21.3)	0.862 (0.126)		63 (19.5)	0.929 (0.090)		
Other chronic condition							
No	261 (80.6)	0.870 (0.161)	0.056	261 (80.5)	0.932 (0.108)	0.080	1.429
Yes	63 (19.4)	0.864 (0.179)		63 (19.5)	0.926 (0.125)		
Acute medical condition							
No	290 (89.5)	0.874 (0.163) *	2.411	306 (94.4)	0.926 (0.108) *	2.797	1.160
Yes	34 (10.5)	0.828 (0.176)		18 (5.6)	0.881 (0.158)		

Note EQ-5D EuroQol five-dimensional questionnaire, EQ-5D-3 L (3 L) three-level version of the EQ-5D, EQ-5D-5 L (5 L) five-level version of the EQ-5D, SD Standard deviation

* $p < 0.05$ (t tests of the difference between known groups)

5 L demonstrated better discriminatory power than 3 L in this study, which is consistent with previous studies [3]. For example, a study of diabetic patients showed that 5 L differentiated between people with and without a complication or comorbid condition more efficiently than 3 L [7]. The better discriminatory power suggests that the 5 L is able to detect true between-group difference in health outcomes with a smaller sample size, compared with the 3 L. Therefore, the 5 L is preferred to the 3 L for measuring the health-related quality of life in the elderly Chinese population. It should be noted that the advantage of the 5 L over the 3 L as measured by the F-statistic ratio in our study is smaller than that in studies of patient populations. This could be because the participants were generally in good and they rarely endorsed the more severe levels of health problems, especially those from the 5 L. This limited the full capacity of the EQ-5D questionnaires especially for the 5 L because it contains more levels of severe health problems. The smaller advantage of the 5 L exhibited in our study could also be partly due to the fact that previous studies used a crosswalk algorithm to calculate the 5 L index score. Index scores based on crosswalk are less precise or reliable than scores based on value sets which are estimated using preferences directly elicited from representative general population samples [30].

Our study has some limitations. First, we did not assess important psychometric properties such as responsiveness and reliability, due to the cross-sectional design of the survey. The second limitation is that participants'

chronic conditions were self-reported without verification by clinicians. It is possible that some participants had undiagnosed medical conditions. Third, all participants in this study were recruited from one community of a small town in east China. Moreover, all participants were recruited from a health center. As a result, our study sample was fairly homogeneous and healthy. Therefore, it might not be representative and our findings may not be generalizable to the entire general elderly population in China. Last but not least, the between-subjects approach is more difficult to implement than the within-subjects approach because of sampling errors and the requirement of randomization. Future studies using this approach should take this limitation into account and perform a form sample size estimation.

Conclusions

The study shows that both EQ-5D questionnaires are valid for measuring the health outcomes of the elderly in China. It appears that the EQ-5D-5 L is a more sensitive measure than the EQ-5D-3 L in this particular population. Therefore, the new EQ-5D questionnaire should be considered when there is a need to measure the health-related quality of life in the general elderly population in the country. Future research is needed to assess whether the EQ-5D-5 L is also more sensitive to change in this population.

Abbreviations

HRQOL: health related quality of life; EQ-5D: EuroQol five-dimensional questionnaire; EQ-5D-3 L (3 L): three-level version of the EQ-5D; EQ-5D-5 L (5 L)

L): five-level version of the EQ-5D; CI: confidence interval; SD: Standard deviation

Acknowledgements

Not applicable.

Authors' contributions

Author RY drafted the manuscript and performed the data analysis, author JL contributed to the data analysis and edited the manuscript, author ZY edited the manuscript, author CP prepared the data collection, author QM supervised the data collection and edited the manuscript, author NL proposed this study and edited the manuscript. All authors read and approved the final manuscript.

Funding

This study was supported by the Science and Technology Bureau of Xiangcheng District in Suzhou, China under grant no. XJ201706 and the Health Commission of Suzhou under grant no. GSW2019090.

Availability of data and materials

All data generated or analysed during this study are included in this published article.

Ethics approval and consent to participate

All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

¹Department of Pharmacy, Union Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, Hubei, China. ²Department of Pharmacy, Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, Hubei, China. ³Health Services Management Department, Guizhou Medical University, Gui'an, China. ⁴School of Public Health, Medical College of Soochow University, Suzhou, China. ⁵The 3rd People's Hospital of Xiangcheng District, 568 Weizhong Road, Suzhou 215134, China. ⁶Saw Swee Hock School of Public Health, National University of Singapore, 16 Medical Drive, Block MD3, Singapore 117597, Singapore.

Received: 1 August 2019 Accepted: 10 March 2020

Published online: 09 April 2020

References

- EuroQol Group. EuroQol-a new facility for the measurement of health-related quality of life. *Health Policy*. 1990;16(3):199–208.
- Herdman M, Gudex C, Lloyd A, Janssen M, Kind P, Parkin D, Bonsel G, Badia X. Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L). *Qual Life Res*. 2011;20(10):1727–36.
- Buchholz I, Janssen M, F, Kohlmann T, Feng YS. A systematic review of studies comparing the measurement properties of the three-level and five-level versions of the EQ-5D. *Pharmacoeconomics*. 2018; 36(6): 645–661.
- Yfantopoulos JN, Chantzaras AE. Validation and comparison of the psychometric properties of the EQ-5D-3L and EQ-5D-5L instruments in Greece. *Eur J Health Econ*. 2017;18(4):519–31.
- 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*. 2017;309(5):357–70.
- Janssen MF, Bonsel GJ, 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*. 2018;36(6):675–97.
- Pan CW, Sun HP, Wang X, Ma Q, Xu Y, Luo N, Wang P. The EQ-5D-5L index score is more discriminative than the EQ-5D-3L index score in diabetes patients. *Qual Life Res*. 2015;24(7):1767–74.
- Wang P, Luo N, Tai E, Thumboo J. The EQ-5D-5L is more discriminative than the EQ-5D-3L in patients with diabetes in Singapore. *Value Health Reg Issues*. 2016;9:57–62.
- Ferreira LN, Ferreira PL, Ribeiro FP, Pereira LN. Comparing the performance of the EQ-5D-3L and the EQ-5D-5L in young Portuguese adults. *Health Qual Life Outcomes*. 2016;14(1):89.
- van Hout B, Janssen M, Feng YS, Kohlmann T, Busschbach J, Golicki D, Lloyd A, Scalone L, Kind P, Pickard A. Interim scoring for the EQ-5D-5L: mapping the EQ-5D-5L to EQ-5D-3L value sets. *Value Health*. 2012;15(5):708–15.
- Mulhern B, Feng Y, Shah K, Janssen MF, Herdman M, van Hout B, Devlin N. Comparing the UK EQ-5D-3L and English EQ-5D-5L value sets. *Pharmacoeconomics*. 2018;36(6):699–713.
- Luo N, Cheung YB, Ng R, Lee CF. Mapping and direct valuation: do they give equivalent EQ-5D-5L index scores? *Health Qual Life Outcomes*. 2015; 13(1):166.
- Yang F, Devlin N, Luo N. Impact of mapped EQ-5D utilities on cost-effectiveness analysis: in the case of dialysis treatments. *Eur J Health Econ*. 2019;20(1):99–105.
- Yang Z, Busschbach J, Liu G, Luo N. EQ-5D-5L norms for the urban Chinese population in China. *Health Qual Life Outcomes*. 2018;16(1):210.
- Luo N, Liu G, Li M, Guan H, Jin X, Rand-Hendriksen K. Estimating an EQ-5D-5L value set for China. *Value Health*. 2017;20(4):662–9.
- Liu GG, Wu H, Li M, Gao C, Luo N. Chinese time trade-off values for EQ-5D health states. *Value Health*. 2014;17(5):597–604.
- Sun S, Chen J, Kind P, Xu L, Zhang Y, Burström K. Experience-based VAS values for EQ-5D-3L health states in a national general population health survey in China. *Qual Life Res*. 2015;24(3):693–703.
- Zhao L, Liu X, Liu D, He Y, Liu Z, Li N. Comparison of the psychometric properties of the EQ-5D-3L and SF-6D in the general population of Chengdu city in China. *Medicine*. 2019;98(11):e14719.
- Gu Y, Zhang H, Ali SH, Huang M, Wei J, Gu S, Zhen X, Hu X, Sun X, Dong H. Social determinants of health-related quality of life among residents in Zhejiang and Qinghai, China. *Int J Environ Res Public Health*. 2019;16(8):1–14.
- Marti-Pastor M, Pont A, Ávila M, Garin O, Vilagut G, Forero CG, Pardo Y, Tresserras R, Medina-Bustos A, García-Codina O, Cabañes J, Rajmil L, Alonso J, Ferrer M. Head-to-head comparison between the EQ-5D-5L and the EQ-5D-3L in general population health surveys. *Popul Health Metrics*. 2018;16(1):14.
- Feng Y, Devlin N, Herdman M. Assessing the health of the general population in England: how do the three- and five-level versions of EQ-5D compare? *Health Qual Life Outcomes*. 2015;13:171.
- Craig BM, Pickard AS, Lubetkin EI. Health problems are more common, but less severe when measured using newer EQ-5D versions. *J Clin Epidemiol*. 2014;67(1):93–9.
- Kim TH, Jo MW, Lee SJ, Kim SH, Chung SM. Psychometric properties of the EQ-5D-5L in the general population of South Korea. *Qual Life Res*. 2013; 22(8):2245–53.
- Fu AZ, Kattan MW. Racial and ethnic differences in preference-based health status measure. *Curr Med Res Opin*. 2006;22(12):2439–48.
- Buchholz I, Thielker K, Feng YS, Kupatz P, Kohlmann T. Measuring changes in health over time using the EQ-5D 3L and 5L: a head-to-head comparison of measurement properties and sensitivity to change in a German inpatient rehabilitation sample. *Qual Life Res*. 2015;24(4):829–35.
- Poór AK, Rencz F, Brodzsky V, Gulácsi L, Beretzky Z, Hidvégi B, Holló P, Kárpáti S, Péntek M. Measurement properties of the EQ-5D-5L compared to the EQ-5D-3L in psoriasis patients. *Qual Life Res*. 2017;26(12):3409–19.
- Pattanasaporn J, Thavorncharoensap M. Measurement properties of the EQ-5D-5L compared to EQ-5D-3L in the Thai diabetes patients. *Health Qual Life Outcomes*. 2015;13(1):14.
- Janssen M, Pickard AS, Golicki D, Gudex C, Niewada M, Scalone L, Swinburn P, Busschbach J. Measurement properties of the EQ-5D-5L compared to the EQ-5D-3L across eight patient groups: a multi-country study. *Qual Life Res*. 2013;22(7):1717–27.
- Janssen MF, Birnie E, Haagsma JA, Bonsel GJ. Comparing the standard EQ-5D three-level system with a five-level version. *Value Health*. 2008;11(2):275–84.
- Golicki D, Niewada M, van Hout B, Janssen M, Pickard AS. Interim EQ-5D-5L value set for Poland: first crosswalk value set in central and Eastern Europe. *Value Health Reg Issues*. 2014;4:19–23.

Publisher's Note

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