This is a repository copy of *Comparing EQ-5D-3L and EQ-5D-5L in measuring the HRQoL burden of 4 health conditions in China*.

White Rose Research Online URL for this paper:
https://eprints.whiterose.ac.uk/189263/

Version: Accepted Version

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

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

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 analogue 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 Appendix Table 1. 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.

-----------------------Table 1----------------------

2

47 In this study, we aimed to compare the measurement properties of the two versions of EQ-5D in quantifying the HRQoL burden

48 associated with 4 chronic conditions including chronic hepatitis B (CHB), depression, generalized anxiety disorder (GAD), and

49 AIDS/HIV in China. The selection of disease groups covered two physical conditions, two mental conditions and used a healthy

50 group as the reference group. We hypothesize that the HRQoL burden of these conditions are mainly in the mental or psychological

51 domain which may be difficult for 3L to detect and necessitates a more sensitive measure like the 5L.

52

53 **Methods**

54 **Participants**

55 This study utilized the data collected in a cross-sectional online survey in China. The survey was part of an international study called

56 'extending the QALY (E-QALY) project' [35]. The E-QALY project aims to develop a new quality of life measure. As for the sample

57 size, Yfantopoulos [36] et al used the sample size of 396 for the study of the psychometric properties of the EQ-5D-3L and EQ-5D-

58 5L instruments in psoriasis, and Bhadhuri [37] et al included 224 patients in the psychometric analyses. Considering that many

59 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

60 study, and this number allows for robust analysis within the groups of interest. Therefore, in the online survey of China, 500

61 respondents with and without a selected health condition were recruited to complete the E-QALY items, EQ-5D-3L, EQ-5D-5L and

62 the Short Warwick-Edinburgh Mental Well-being Scale (SWEMWS). This data was collected between April and July 2019 online

63 by Accent, a U.K online survey company. Quotas and inclusion criteria were applied to recruit a sample of 500 individuals in which

64 there were similar numbers of individuals with GAD, HIV/AIDS, CHB, or depression, or without any of those 4 chronic conditions.

65 The sample was broadly representative to the country in terms of geography, ethnicity and gender. The study was approved by the

66 Ethic Committee of University of Sheffield, United Kingdom (Approval letter number 025524) and the IRB of Jinan University,

67 China (Approval letter number JNUKY-2020-001), and all methods were performed in accordance with the relevant guidelines and

68 regulations. Informed consent was obtained from all participants prior to the online survey through panel, the type was electronic.

69

70 The online survey began by giving an outline of the research purpose. Participants were then asked to report their disease history.

71 Eligible respondents reported their background information including education level, gender and age etc. Next, respondents were

72 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

73 responded to 5L and the other half 3L), some more E-QALY items, the SWEMWS, the 5L/3L, and the EQ-VAS.

74 **Instruments**

75 The EQ-5D-3L and EQ-5D-5L were both preference based HRQoL instruments developed by the EuroQol Group. Both instruments

76 have the same five health dimensions, i.e., mobility, self-care, usual activities, pain/discomfort, anxiety/depression. The difference

77 is the 3L has three response levels (no problems, some problems, extreme problems) while the 5L has five response levels (no

78 problems, slight problems, moderate problems, severe problems, and unable/extreme problems) for each dimension. To calculate

79 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].

80 **Statistical analyses**

81 We first described the characteristics of our sample and we reported, by condition groups 1) the median response distributions;2)

82 sensitivity to mild health problems as measured by the percentages of reporting 'no problems'   ; 3) the distributions of the 3L and

83 5L utility scores4) the known-group validities when compared with the healthy group and the relative efficiency between the 5L

84 and 3L. Data was analyzed using Stata for Windows, Version 14.0 MP, and IBM SPSS Statistics for Windows, Version 22.0. Armonk,

85 NY: IBM Corp (2013).

86

87 Given the skewed distribution of EQ-5D responses [42], the median responses were reported to understand the overall health state

88 of each group. Then, the percentages of reporting 'no problems' for each dimension if often being referred as 'ceiling effects' in

89 published studies. This is not accurate given it is unknown the reasons of this large proportions of reporting 'no problems'. For this

90 reason, we refer this phenomenon as sensitivity to mild health problems, which was defined as the proportion of respondents

91 indicating 'no problems' in each dimension and all five dimensions taken together [43]. Previous studies have shown that the use of

92 5L could reduce the responses of reporting 'no problems' and was considered as a more sensitive measure [16,18] . For this reason,

93 we hypothesized that the 5L has better sensitivity and calculated the reduction of reporting 'no problems' from 3L to 5L. We

94 evaluated the reduction separately for each dimension and all five dimensions taken together. Next, the index values of each group

95 were calculated using 3L and 5L value sets respectively and the distributions were plotted.

4

96

97   Known-group validity of the two EQ-5D index scores was evaluated using e the analysis of variance (ANOVA) tests. Relative

98   efficiency was calculated as the ratio of F statistics derived from the ANOVA analysis. The F-statistic was widely used to assess RE

99   of measurement scales [19,44,45]. The index score with a higher F statistic would be deemed to be more efficient than its comparator

100   since a higher value of F statistic is more likely to result in statistical significance. To understand the RE of the index scores, we

101   compared the distributions of the responses to the EQ-5D dimensions between the healthy group and each of the condition group.

102   Mann-Whitney test was used. For reference, we listed the median values of each dimension reported by the healthy group and 4

103   condition groups.

104

105   **Results**

106   In total, 500 individuals completed the online survey, including 140 healthy individuals, 122 individuals with CHB, 107 with

107   depression, 90 individuals with GAD and 101 with HIV/AIDS. Some respondents reported multiple conditions, e.g. 68 individuals

108   reported both depression and GAD. In general, the whole study sample was young (mean age: 35.8, SD: 8.64) and well educated.

109   The gender proportions of the five groups were generally balanced except for the group of HIV/AIDS, in which, about 87.1% of

110   individuals were female. In terms of the age distribution, the healthy group was mostly young; the CHB group has more participants

111   aged between 40 and 49; the depression and GAD groups had individuals from all four age groups and the HIV/AIDS group aged

112   mainly from 30 to 49. Individuals with tertiary education accounted for over 80% for all four disease groups and the healthy group

113   had more individuals with secondary education. Table 2 shows the demographic information by condition.

114   ------------------------------Table 2--------------------------------------

115

116   There were 13, 34, 28, 26 and 26 unique states reported for the healthy, CHB, depression, GAD and HIV/AIDS group respectively

117   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

118   healthy and HIV/AIDS groups were 'no problems' across the five dimensions. In comparison, when measured by the 5L, while the

119   median responses for the healthy group remained to be 'no problems', the median responses of HIV/AIDS group were all 'slight

120   problems'. Similarly, the median responses for the CHB group were 'slight problems' for the last three dimensions when measured

121   by the 5L, while the only pain/discomfort dimension had 'moderate problem' measured by the 3L. The median responses of the last

5

122 two dimensions for the GAD and depression group were both on the second level for the 3L (moderate problems) and 5L (slight

123 problems). . The Mann-Whitney results were all significant at 0.01 level suggesting all 4 condition groups had a different distribution

124 of responses against the healthy group, for both 3L and 5L.

125

126 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

127 all condition groups. When all dimensions are considered, the number of health profiles 11111 decreased by almost 40% when

128 reporting using the 5L. The reduction of reporting 'no problems' is more salient in the pain/discomfort and anxiety/depression

129 dimensions. The most prominent difference was observed in the HIV/AIDS group, i.e., all dimensions had a relative reduction of

130 over 30%.

131 -----------------------------Table 3------------------------------------

132

133

134 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

135 B of 0.781 (SD: 0.166), depression of 0.721 (SD: 0.180) and GAD of 0.711 (SD: 0.178). In comparison, the mean utilities of the 5L

136 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,

137 0.746 (SD: 0.202) for the depression group and 0.718 (SD: 0.217) for the GAD group. As shown in Fig. 1-6, the 5L index value had

138 a smoother distribution, less clustering and wider range (-0.305 to 1) compared with the 3L (ranged from -0.03 to 1). An exception

139 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

140 reported most frequently in the HIV/AIDS group.

141

142 **Fig. 1** Distributions of utility score for the whole sample

143 **Fig. 2** Distributions of utility score for healthy group

144 **Fig. 3** Distributions of utility score for chronic hepatitis B group

145 **Fig. 4** Distributions of utility score for depression group

6

146   **Fig. 5** Distributions of utility score for GAD group

147   **Fig. 6** Distributions of utility score for HIV/AIDS group

148

149   Table 4 shows the mean index values of by groups and the ANOVA tests between the healthy group and the four condition groups.

150   The last column shows the relative efficiency between the 5L and 3L among these four comparisons. The index values of the 5L

151   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

152   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

153   sub-group and both versions demonstrated good known-group validity, except that the 3L did not show a statistically significant

154   result in the comparison of the healthy and HIV/AIDS groups. The relative efficiency of the 5L index was higher in all four

155   comparisons.

156   -----------------------------Table 4------------------------------------

157

158   **Discussion**

159   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

160   and focused on comparing the measurement sensitivity of these two EQ-5D versions. In general, we found both versions of EQ-5D

161   to be sensitive tools to quantify the HRQoL loss caused by the 4 chronic conditions, but the 5L showed an improved sensitivity to

162   pick up mild health problems. When combined with the life expectancy data, quality-adjusted life years (QALYs) can be calculated

163   and can be used as a standard measure to reflect how a condition could affect the length and the quality of an individual, which

164   provided a single metric to reflect disease burden. Disease burden is typically measured using disability adjusted life years (DALYs),

165   but QALYs can also be used and may be a better measure as it could provide individual level HRQoL based data and it is

166   recommended measure for economic evaluations [46].

167

168   For the measurement properties, our results generally agree with the findings reported before, that is, although both versions

7

169 produced highly agreed responses, index values and both had good known-group validity, the 5L performed better in terms of

170 response distribution, sensitivity to mild health problems, index value distribution and had higher relative efficiency [16,47,48]. The

171 less clustering of reporting 'no problems' for the 5L is most evident for the pain/discomfort and anxiety/depression dimensions.

172 Previous studies referred this issue of clustering at 'no problems' as ceiling effects, but it should be noted that there are two kinds

173 of responses in play, first, the respondents that do not have any problems and second, the respondents do have problems but report

174 'no problems' anyway. Theoretically, ceiling effects exist when the latter kind of responses contributed the most to the clustering at

175 'no problems'. Our results showed that the relative reduction is smaller for the healthy group, which suggests the large proportion

176 of reporting 'no problems' is not a ceiling effects, but a genuine reflection of the health state of the healthy group. In overall, around

177 40% of relative reduction of reporting 'no problems' was observed when five dimensions were taken together. This was larger than

178 the numbers reported in other studies, which ranged between 6.9% and 33.7% [16,47]. Overall, this shows the limitation of less

179 cutting-off points provided by the 3L descriptive system, which was first reported by Mathieu F et al [48]. Hence, the 5L is more

180 sensitive in measuring the HRQoL of individuals with mild health conditions.

181

182 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

183 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,

184 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,

185 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

186 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].

187 In comparison, there were still some clusterings in the 5L distributions, but it is not due to the clusterings of profiles, instead, it is

188 because more profiles were reported and some profiles have similar index values. In overall, the increased levels of the 5L defined

189 more health states and provided more subtle index values.

190

191 The relative efficiency results favored the 5L and is in line with the study of You et al and [16]. Based on the F-ratios in Table 4, it

8

192 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

193 condition group (i.e., GAD and depression), the profiles were more similar, that is, the median levels for both 3L and 5L were both

194 11122 for the mental condition groups. In contrast, the median levels of the HIV/AIDS group were 22222 when used the 5L and

195 were 11111 when used the 3L. This was also observed in the CHB group, where the median level of usual activities was slight

196 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,

197 respondents tent to report no problems. The minimal advantage of 5L compared to 3L in differentiating individuals with and without

198 a mental disorder could be due to the weakness of EQ-5D in measuring mental health. There is only one item in the instrument

199 targeting mental problems.

200

201 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

202 people and less educated people are less active on Internet. Therefore, findings of this study may not be generalizable to older

203 populations. It has been reported the elderly with less education attainment may have more difficulty to use the five-level descriptive

204 system and in such occasions, the 3L may be a more feasible alternative [33]. Second, some respondents reported more than one

205 condition, but we did not provide deep analysis about the possible effect of multi-conditions. It should be noted that our sample,

206 which was recruited online and the health condition was self-reported. Ideally, clinical data is used to verify the presence and absence

207 of diagnoses reported by the study subjects. Due to these limitations, the superiority of the 5L warrants further studies.

208 **Conclusions**

209 In this study, both EQ-5D versions could quantify the disease burden of both physical and mental diseases in terms of self-reported

210 HRQoL. Overall, our findings favor the 5L version of EQ-5D for measuring health related quality of life burden caused by chronic

211 disease in China. As patient-reported outcome (PRO) is increasingly being used all over the world to measure disease burden, EQ-

212 5D can be a useful tool in this context.

213 **List of abbreviations**

214 **HRQoL:** Health related quality of life

9

215   **EQ-VAS:** EQ visual analogue scale

216   **3L:** EQ-5D-3L

217   **5L:** EQ-5D-5L

218   **E-QALY :** Extending the QALY

219   **SWEMWS:** Short Warwick-Edinburgh Mental Well-being Scale

220   **GAD:** Generalized anxiety disorder

221   **RE:** Relative efficiency

222   **QALYs:** Quality-adjusted life years

223   **DALYs:** Disability adjusted life years

224   **PRO:** Patient-reported outcome

225

226   **References**

227   1. Devlin, N.J. and Brooks, R.: EQ-5D and the EuroQol Group: Past, Present and Future. Appl. Health Econ. Health Policy.
228   **15**(2), 127-137 (2017). https://doi.org/10.1007/s40258-017-0310-5
229   2. Szende, A. and Janssen, B.: Socio-demographic Indicators Based on EQ-5D. In: A. Szende, B. Janssen, and J. Cabases (eds.)
230   Self-Reported Population Health: An International Perspective based on EQ-5D, pp. 37-46. Dordrecht (NL) (2014)
231   3. Devlin, N.J. and Krabbe, P.F.: The development of new research methods for the valuation of EQ-5D-5L. Eur. J. Health
232   Econ. **14 Suppl 1**, S1-3 (2013). https://doi.org/10.1007/s10198-013-0502-3
233   4. Janssen, M.F., Pickard, A.S., Golicki, D., Gudex, C., Niewada, M., Scalone, L., et al.: Measurement properties of the EQ-5D-
234   5L compared to the EQ-5D-3L across eight patient groups: a multi-country study. Qual. Life Res. **22**(7), 1717-27 (2013).
235   https://doi.org/10.1007/s11136-012-0322-4
236   5. Herdman, M., Gudex, C., Lloyd, A., Janssen, M.F., Kind, P., Parkin, D., et al.: Development and preliminary testing of the
237   new five-level version of EQ-5D (EQ-5D-5L). Qual. Life Res. **20**(10), 1727-36 (2011). https://doi.org/10.1007/s11136-011-
238   9903-x
239   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
240   Angina Pectoris. Chinese Journal of Health Statistics. **30**(06), 829-832 (2013). (in Chinese)
241   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
242   validation of the EuroQoL 5-Dimension 5-Level (EQ-5D-5L) in Chinese patients with adolescent idiopathic scoliosis.
243   Scoliosis Spinal Disord. **11**, 19 (2016). https://doi.org/10.1186/s13013-016-0083-x
244   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
245   patients with systemic lupus erythematosus. PLoS One. **9**(6), e98883 (2014). https://doi.org/10.1371/journal.pone.0098883
246   9. Yao, J., Liang, Y., Xu, Y., Guo, L., Yang, L., Pan, J., et al.: Analysis of the Measurement Characteristics of Chinese Version
247   of EQ-5D-3L Applied in Population with Chronic Disease: Rasch Model. Chinese Journal of Health Statistics. **35**(05), 681-
248   683+688 (2018). (in Chinese)
249   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

10

250   Patients with Chronic Hepatitis B. Chinese Journal of Vaccines and Immunization. **19**(06), 492-497+544 (2013). (in Chinese)

251 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

252   the EuroQoL 5-dimension (EQ-5D) questionnaire in patients with spondyloarthritis. Arthritis Res. Ther. **21**(1), 41 (2019).

253   https://doi.org/10.1186/s13075-019-1826-x

254 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-

255   6D among patients with haemophilia. Eur. J. Health Econ. **22**(4), 547-557 (2021). https://doi.org/10.1007/s10198-021-01273-

256   5

257 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

258   cancers: suggestions for instrument choosing. Qual. Life Res. **30**(3), 841-854 (2021). https://doi.org/10.1007/s11136-020-

259   02636-w

260 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

261   patients with hepatitis B. Qual. Life Res. **23**(8), 2355-2363 (2014). https://doi.org/10.1007/s11136-014-0670-3

262 15. Zeng, X., Sui, M., Liu, B., Yang, H., Liu, R., Tan, R., et al.: Measurement Properties of the EQ-5D-5L and EQ-5D-3L in Six

263   Commonly Diagnosed Cancers. Patient. **14**(2), 209-222 (2021). https://doi.org/10.1007/s40271-020-00466-z

264 16. You, R., Liu, J., Yang, Z., Pan, C., Ma, Q., and Luo, N.: Comparing the performance of the EQ-5D-3 L and the EQ-5D-5 L in

265   an elderly Chinese population. Health Qual. Life Outcomes. **18**(1), 97 (2020). https://doi.org/10.1186/s12955-020-01324-0

266 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-

267   5D-3L and EQ-5D-5L in acute myeloid leukemia patients. Qual. Life Res. **30**(3), 855-866 (2021)

268   https://doi.org/10.1007/s11136-020-02644-w

269 18. Jiang, J., Hong, Y., Zhang, T., Yang, Z., Lin, T., Liang, Z., et al.: Comparing the measurement properties of the EQ-5D-5L

270   and the EQ-5D-3L in hypertensive patients living in rural China. Qual. Life Res. **30**(7), 2045-2060 (2021).

271   https://doi.org/10.1007/s11136-021-02786-5

272 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-

273   5D-3L index score in diabetes patients. Qual. Life Res. **24**(7), 1767-1774 (2015). https://doi.org/10.1007/s11136-014-0902-6

274 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

275   China. Chinese Health Economics. **35**(03), 17-20 (2016). (in Chinese)

276 21. Cai, Y., Wu, H., Yang, Z.: Comparison analysis of European quality of life-5 dimensions with three-level and five-level

277   based on four types of people in China. Chin J Dis Control Prev. **24**(10), 1190-1194 (2020). (in Chinese)

278 22. Zhao, Z., Xu, S., Wang, X., Zhang, P., Bi, Y., Li, D., et al.: Scientific evaluation of the European Five-dimensional Health

279   Scale inpatients with chronic heart failure. Chin. J. Tradit. Chin. Med. Pharm. **34**(04), 1715-1718 (2019). (in Chinese)

280 23. Tsang, H. H. L., Wong, C. K. H., Cheung, P. W. H., Lau, C. S., Chung, H. Y., Cheung, J. P. Y. Responsiveness of the

281   EuroQoL 5-Dimension (EQ-5D) questionnaire in patients with spondyloarthritis. BMC Musculoskelet Disord. **22**(1), 439

282   (2021). https://doi.org/10.1186/s12891-021-04315-4

283 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

284   adolescent idiopathic scoliosis. Eur. Spine J. **27**(2), 278-285 (2018). https://doi.org/10.1007/s00586-017-5330-1

285 25. Xing, Y., Ma, A.: An empirical study comparing the Chinese versions of EQ-5D-3L and EQ-5D-5L. Shanghai Medical &

286   Pharmaceutical Journal. **34**(07), 27-31 (2013). (in Chinese)

287 26. Hu, H., Shen, W., Jia, M., Yu, G., Mei, X., Bian, Q., et al.: Reliability and Validity of the Chinese Version of the EQ-5D-3L

288   in Patients of Chronic Kidney Disease. Chinese Journal of Integrated Traditional and Western Nephrology. **13**(08), 690-693

289   (2012). (in Chinese)

290 27. Hu, X., Jing, M., Zhang, M., Yan, X., Huang, Y.: The reliability and validity of EQ-5D-3L in Uygur cervical cancer patients.

291   **37**(06), 780-785 (2019). (in Chinese)

292 28. Qiu, Y.: Reliability and Validity of the Chinese Version of EQ-5D in the knee osteoarthritis [Master, Capital Institute of

293   Physical Education]. (2015). (in Chinese)

294 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

295   China. J. Sichuan Univ. (Med. Sci. Edi). **51**(05), 691-694 (2020). (in Chinese)

296 30. Xing, Y., Ma, A.: Study on reliability and validity of Chinese version of EQ-5D-5L. Shanghai Medical & Pharmaceutical

297 Journal. **34**(09), 40-43 (2013). (in Chinese)

298 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
299   disease in China. Springerplus. **5**(1), 1924 (2016) https://doi.org/10.1186/s40064-016-3613-3

300 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
301   leukemia patients. BMC Cancer. **19**(1), 522 (2019). https://doi.org/10.1186/s12885-019-5721-2

302 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?
303   Evidence From a Qualitative Study. Pharmacoecon. Open. **5**(1), 101-109 (2021). https://doi.org/10.1007/s41669-020-00212-z

304 34. Dai, L., Wen, Y., Lu, T., Cai, J., Wang, L., Qin, M., et al.: Measurement characteristics evaluation of SF-HSTCM and EQ-
305   5D-3L among patients with stroke. Journal of Capital Medical University. **40**(04), 627-634 (2019). (in Chinese)

306 35. Peasgood, T., Mukuria, C., Brazier, J., Marten, O., Kreimeier, S., Luo, N., et al.: Developing a New Generic Health and
307   Wellbeing Measure: Psychometric Survey Results for the EQ Health and Wellbeing. Value in Health. (2022).
308   https://doi.org/10.1016/j.jval.2021.11.1361

309 36. Yfantopoulos, J., Chantzaras, A., and Kontodimas, S.: Assessment of the psychometric properties of the EQ-5D-3L and EQ-
310   5D-5L instruments in psoriasis. Arch. Dermatol. Res. **309**(5), 357-370 (2017). https://doi.org/10.1007/s00403-017-1743-2

311 37. Bhadhuri, A., Kind, P., Salari, P., Jungo, K.T., Boland, B., Byrne, S., et al: Measurement properties of EQ-5D-3L and EQ-
312   5D-5L in recording self-reported health status in older patients with substantial multimorbidity and polypharmacy. Health
313   Qual. Life Outcomes. **18**(1), 317 (2020). https://doi.org/10.1186/s12955-020-01564-0

314 38. Poor, A.K., Rencz, F., Brodszky, V., Gulacsi, L., Beretzky, Z., Hidvegi, B., et al: Measurement properties of the EQ-5D-5L
315   compared to the EQ-5D-3L in psoriasis patients. Qual. Life Res. **26**(12), 3409-3419 (2017). https://doi.org/10.1007/s11136-
316   017-1699-x

317 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
318   to Treatment Benefit of Cataract Surgery. Patient-Patient Centered Outcomes Research. **12**(4), 383-392 (2019)

319 40. Liu, G.G., Wu, H., Li, M., Gao, C., and Luo, N.: Chinese time trade-off values for EQ-5D health states. Value Health. **17**(5),
320   597-604 (2014). https://doi.org/10.1016/j.jval.2014.05.007

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

323 42. Parkin, D., Devlin, N., and Feng, Y.: What Determines the Shape of an EQ-5D Index Distribution?. Med Decis Making.
324   **36**(8), 941–951 (2016). https://doi.org/10.1177/0272989X16645581

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

327 44. Vickrey, B.G., Hays, R.D., Genovese, B.J., Myers, L.W., and Ellison, G.W.: Comparison of a generic to disease-targeted
328   health-related quality-of-life measures for multiple sclerosis. J. Clin. Epidemiol. **50**(5), 557-69 (1997).
329   https://doi.org/10.1016/s0895-4356(97)00001-2

330 45. Luo, N., Johnson, J.A., Shaw, J.W., and Coons, S.J.: Relative Efficiency of the EQ-5D, HUI2, and HUI3 Index Scores in
331   Measuring Health Burden of Chronic Medical Conditions in a Population Health Survey in the United States. Medical Care.
332   **47**(1), 53-60 (2009)

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

336 47. Buchholz, I., Janssen, M.F., Kohlmann, T., and Feng, Y.S.: A Systematic Review of Studies Comparing the Measurement
337   Properties of the Three-Level and Five-Level Versions of the EQ-5D. Pharmacoeconomics. **36**(6), 645-661 (2018).
338   https://doi.org/10.1007/s40273-018-0642-5

339 48. Janssen, M.F., Bonsel, G.J., and Luo, N.: Is EQ-5D-5L Better Than EQ-5D-3L? A Head-to-Head Comparison of Descriptive
340   Systems and Value Sets from Seven Countries. Pharmacoeconomics. **36**(6), 675-697 (2018). https://doi.org/10.1007/s40273-
341   018-0623-8

342 49. Feng, Y., Devlin, N., Bateman, A., Zamora, B., and Parkin, D.: Distribution of the EQ-5D-5L Profiles and Values in Three
343   Patient Groups. Value Health. **22**(3), 355-361 (2019). https://doi.org/10.1016/j.jval.2018.08.012

344
345