ScienceDirect

Contents lists available at sciencedirect.com
Journal homepage: www.elsevier.com/locate/jval

Preference-Based Assessments

Testing the Valuation of the EQ-5D-Y-5L in Adults and Adolescents: Results From a 5-Country Study and Implications for the Descriptive System

Tianxin Pan, PhD, Juan Manuel Ramos-Goni, PhD, Bram Roudijk, PhD, Shitong Xie, PhD, Feng Xie, PhD, Zhihao Yang, PhD, Brendan Mulhern, PhD, Richard Norman, PhD, Nancy Devlin, PhD

ABSTRACT

Objectives: The EQ-5D-Y-5L (Y-5L) is a new health-related quality-of-life instrument for children and adolescents. Value sets for the Y-5L are planned. This article aimed to test the ability of adult and adolescent respondents to differentiate the ordinal levels of the Y-5L in valuation tasks and to explore the characteristics of stated preferences for the Y-5L between adults and adolescents.

Methods: We collected latent-scale discrete choice experiment data via an online survey of adults (≥18 years) and adolescents (12-17 years) in Australia, Canada, China, The Netherlands, and Spain. A D-Efficient design consisting of 192 choice pairs was grouped into 16 blocks of 12 choice tasks per respondent. We used mixed-logit models to analyze the data and incremental dummies to represent movements from a less-severe level to its consecutive more-severe level.

Results: We did not observe preference inversions in adults or adolescents (ie, no statistically significant positive coefficients on the incremental dummies). Adults showed similar preferences for the Y-5L in terms of dimension importance: Pain/Discomfort was considered the most important dimension in all countries except for China; Looking After Myself and Usual Activity were the least important dimensions. In contrast, Mobility was considered the most important dimensions by adolescents in Canada, Spain, and China.

Conclusions: Adults could differentiate between the Y-5L level labels in valuation tasks, whereas more randomness was observed in adolescents' choices. Observed differences between adult and adolescent stated preferences for the Y-5L raise questions about how these preferences should be reflected in cost-effectiveness analysis.

Keywords: adolescent, DCE, EQ-5D-Y-5L, HRQoL, pediatric, pediatric PROs, preference weights, utilities, values.

VALUE HEALTH. 2025; 28(12):1900–1910

Highlight

- The EQ-5D-Y-5L (Y-5L) is a new health-related quality-of-life instrument for children and adolescents. However, it is unclear whether adults and adolescents can differentiate between its levels in valuation studies and how stated preferences for the Y-5L compare between adults and adolescents.
- Adults could interpret the Y-5L level labels in their intended order of severity when completing latent-scale discrete choice experiment tasks. Adolescents can largely differentiate between level labels; however, their choices exhibited more randomness. Adolescents and adults had different preferences regarding Y-5L dimensions.
- The Y-5L is feasible for use in health state valuation tasks. Differences between adult and adolescent preferences for the Y-5L pose a challenge for generating value sets (ie, can these be combined in a single value set?).

Introduction

The EQ-5D-Y-5L (Y-5L) is a new instrument for measuring health-related quality of life (HRQoL) in children and adolescents, comprising 5 dimensions with 5 levels of problems in each.[1] Value sets for the Y-5L are planned to support its use in economic evaluation. Emerging evidence shows that the Y-5L has improved psychometric performance as a descriptive system compared with the 3-level version of the instrument, the EQ-5D-Y-3L (Y-3L).[2-4] However, the expanded number of levels in the Y-5L inevitably introduces more subtle differences between the labels used to describe the levels and, potentially, more difficulty for respondents in stated preference studies to understand the intended ordinal relationship between them.

In addition to having more levels, there are several other noteworthy differences between the Y-5L and the Y-3L descriptive systems (see Appendix Fig. 1 in Supplemental Materials found at https://doi.org/10.1016/j.jval.2025.07.016). For 4 of the dimensions (Mobility [MO], Looking After Myself [LAM], Usual Activity [UA] and Pain/Discomfort [PD]), the Y-5L adds an intermediate level ("a little bit") between "no problem" and "some problems" on the Y-3L. However, the worst problem (level 5) on the Y-5L on these dimensions ("I cannot" on MO, LAM, and UA and "extreme" for PD) is descriptively worse than the worst problem (level 3) on these dimensions in the Y-3L ("a lot" of problems).[1] For these 4 dimensions, the worst level on the Y-3L directly corresponds to level 4 on the Y-5L. Second, for the last dimension, "feeling Worried, Sad, or Unhappy" (WSU), the level labels for all levels (apart from no problems) are descriptively different between the Y-3L and Y-5L. For example, level 3 on the Y-3L is "very," whereas level 5 on the Y-5L is "extremely"; unlike the other 4 dimensions, the Y-3L level-3 descriptor is different from the level-4 descriptor on the Y-5L ("really").

Although there is extensive evidence on values for the Y-3L,[5,6] valuing the Y-5L could be associated with greater challenges. Although there is evidence that Y-5L is comprehensible and feasible for children to complete,[4] its use in health state valuation exercises has not been investigated. If respondents in valuation studies find the labels for adjacent levels difficult to differentiate between, this could result in modeled Y-5L values showing no difference in utility between levels or result in apparent preference reversals. It is important to test the ability of respondents to complete stated preference tasks where states defined by the Y-5L are presented "out of the context" of the Y-5L questionnaire.

Furthermore, despite translations of the Y-5L following a robust translation process to achieve semantic equivalence,[7] there can be subtle differences between language versions, which potentially adds to the complexity of valuing the descriptive system. For example (as shown in Appendix Table 1 in Supplemental Materials found at https://doi.org/10.1016/j.jval.2025.07.016), in simplified Chinese, the level labels for level 4 on PD and WSU are identical, ("非常"), whereas in English, these are "a lot of problems" and "really," respectively. In the Dutch version, level 5 on PD and WSU use different wording ("heel veel" and "heel erg," respectively), whereas "extreme(ly)" is used for level 5 on both of these dimensions in the English version. It is therefore also important to examine whether the Y-5L descriptive system is interpreted by respondents in a broadly consistent way across different language versions.

Another key issue in the valuation of the child HRQoL instruments is the feasibility of eliciting values from younger people.[8] Stakeholder consultation has indicated support for incorporating adolescents' preference in valuation of child HRQoL, rather than relying (exclusively) on adults' preferences.[8-10] However, there are several challenges, including both ethical and feasibility considerations, in eliciting preferences from adolescents, which may also have a bearing on adolescents' ability to value more complex HRQoL instruments. Thus, there is a need to test whether both adult and adolescent respondents can clearly interpret the expanded levels and labels in their intended order of severity, to inform the development of Y-5L value sets.

The primary aim of this study was to test the ability of adult and adolescent respondents to differentiate between the ordinal levels of the Y-5L in a stated preference study in 5 countries: Australia, Canada, China, The Netherlands, and Spain. The secondary aim was to explore the characteristics of stated preferences for the Y-5L across countries and between adults and adolescents.

Methods

Valuation Methods

We used latent-scale discrete choice experiments (DCE) as the principal valuation method for 2 main reasons: (1) We were interested in the ability of respondents to differentiate between levels (rather than construction of a value set); therefore, we did not need to anchor it on the 0-1 scale and (2) the latent-scale DCE is a central part of the current EQ-VT approach and in the Y-3L valuation protocol.[11] Respondents were asked to consider 2 Y-5L health states and indicate which one they prefer when considering the choice on behalf of a hypothetical 10-year-old child (for adult respondents)[11] or from their own perspective (for adolescent respondents) (examples provided in Appendix Fig. 2 in Supplemental Materials found at https://doi.org/10.1016/j.jval.2025.07.016).

To ensure comparability across data collection sites, we used the same D-Efficient design across countries, with −0.001 as the prior value for all model coefficients (we used incremental dummies in the model as explained in the Data Analysis section). The design consisted of 192 choice pairs, grouped into 16 blocks of 12 choice tasks per respondent. The order of the choice tasks within each block was randomized to avoid ordering effects. Given that we aimed to estimate separate models for adults and adolescents in each country, the study targeted 1000 adults (≥18 years) and 1000 adolescents (12-17 years) in each country. This means that we aimed for (1000 × 12)/192 = approx. 62 or 63 observations per choice set, which is in the range of much of the preexisting literature using DCE for health state valuation.[12,13]

The online survey was developed using an adapted version of the EuroQol Valuation Technology (EQ-VT) protocol used in Y-3L valuation.[11] The online survey included (1) introduction and consent; (2) background information including age, gender, experience with serious illness, parental status, and respondent's own health using Y-5L; and (3) 12 latent-scale DCE tasks. Additional tasks included in the survey (eg, DCE with duration or dead tasks among adults in 5 countries and adolescents in Canada) are reported separately.[14]

Selection of Countries

Data were collected in Australia (AU), Canada (CA), China (CHN), The Netherlands (NL), and Spain (SP). The rationale for selecting these countries was that (1) the countries span a number of regions and languages: 2 English-speaking countries (AU and CA) in North America and Australasia, 1 country in Asia (CHN), and 2 European countries (SP and NL); and (2) all 5 countries had an approved translation of the Y-5L available.

Sample Recruitment

This study was approved by The University of Melbourne Human Research Ethics Committee (2022-24583-31815-4) in Australia, Hamilton Integrated Research Ethics Board (2022-#15051) in Canada, the Ethics Committee of Guizhou Medical University (2022-249), Research Ethics Committee at the Canary Islands University Hospital (2022-11-24) in Spain, and received a waiver by the Medical Ethical Committee of Erasmus University Medical Center (MEC-2022-0555) in the Netherlands. Adult and adolescent samples were initially recruited from multiple panels via a market research technology company, Cint, in each country. Because of challenges in achieving the required sample size in adolescents in Australia and The Netherlands, additional samples were collected via local panels specifically targeted at adolescents (Student Edge in Australia, Bureau Fris BV in The Netherlands). A mixed recruitment strategy was used for the adolescent sample: self-recruiting and/or recruiting via parents depending on country-specific regulations and local ethics considerations. Data collection commenced in each country as soon as local ethics approval was obtained and took place in the following order: China (January-February 2023; additional sample in July 2023 to recruit more older adult respondents), Spain (Feb 2023), The Netherlands (February-April 2023; additional sample for adolescents in July 2023), Canada (February-April 2023), and Australia (February-April 2023; additional sample for adolescents in July 2023).

Quality Control

Given the risk of fraudulent responses (eg, bots) in online data collection,[15,16] participants first had to complete a Completely Automated Public Turing test to tell Computers and Humans Apart (CAPTCHA) to access the survey which is widely used to identify computer bots. We implemented several quality control (QC) measures to identify poor-quality data and exclude respondents from the sample, such as identifying speeders, an innovative "traffic light" task to ask respondents to rank 3 colors of a traffic light, examining answers to repeated questions, and identifying repeated identical comments and IP addresses. More

details on QC and results are provided in Supplemental Materials found at https://doi.org/10.1016/j.jval.2025.07.016.

Data Analysis

Descriptive statistics were used to summarize the QC results of collected data and sample characteristics of the final analytical sample. We descriptively reported responses to feedback questions about the valuation tasks (asking about ease of understanding, being able to tell the difference between states, and difficulty deciding between the states).

Latent-scale DCE data were analyzed using a mixed-logit model (MIXL) with a linear, additive utility function using incremental dummies. Incremental dummies represent the effect on preferences of a movement between consecutive levels within dimensions (the dummy variable for level 2 represents the increment from level 1 to level 2, the dummy for level 3 represents the increment from level 2 to level 3, etc). Standard MIXL model estimates were performed using Bayesian Markov-Chain Monte Carlo (MCMC) methods in Stata ("garbage_mixl" command)[17] and were based on 50 000 burn-in and 100 000 additional MCMC draws.

We expected statistically significant negative coefficients of the increment dummies, representing the utility decrement from a less-severe level to a more-severe level. Negative coefficients close to 0 or positive coefficients (not statistically significance at 5% level) were defined as coefficient indifference between levels.[18] Positive coefficients that were statistically significant at 5% level were defined as coefficient inconsistency (preference reversals).

We explored the relative importance of dimensions by country and by adults and adolescents. By taking the sum of the coefficients (including any positive ones) for each incremental dummy in each dimension, we calculate the latent-scale utility range of each dimension. Relative attribute importance scores were obtained by dividing the utility range for each dimension by the total utility range.[19]

To test whether any differences observed across countries and between adults and adolescents were of statistical significance, we applied heteroskedastic conditional logit models.[20,21] We introduced 2 sets of scale parameters in the models: gamma, to capture the relative choice consistency between 2 samples (eg, adults and adolescent within 1 country, or adult/adolescent samples from 2 countries), and thetas, representing the relative consistency of preference for each dimension, respectively. The models were estimated using Bayesian MCMC methods in OpenBUGS, using code developed by Jonker et al.[21] Details about the method, our model specification, and interpretation have been provided in Supplemental Materials found at https://doi.org/10.1016/j.jval.2025.07.016.

Results

Quality Assessment

The targeted sample size was achieved for the adult samples in all countries and for adolescent samples in China, Spain, and Canada. After additional sampling targeted specifically at adolescents, we were still unable to achieve the target adolescent sample in The Netherlands (fewer respondents accessed the survey) and in Australia (fewer respondents accessed the survey, and a higher proportion were excluded through QC).

Appendix Table 2 in Supplemental Materials found at https://doi.org/10.1016/j.jval.2025.07.016 shows the QC results and final sample sizes. Failing the traffic light task was the most common reason for adults failing QC in all countries except China, where more respondents speeded through DCE tasks or failed IP checks. Among adolescents, speeding on DCE tasks was the major cause of QC exclusion except Australia, where we found a high proportion (47%) of adolescents recruited from parents via Cint provided identical comments or comments written in other languages, suggesting fraudulent data. Therefore, we dropped this entire sample, only keeping self-recruited adolescents via Cint or Student Edge in Australia (QC results by sample source provided in Appendix Table 3 in Supplemental Materials found at https://doi.org/10.1016/j.jval.2025.07.016).

Sample Characteristics

Table 1 shows the characteristics of included respondents. Age groups were broadly representative across countries and samples, except for Chinese adults among whom only 2.3% were above 60 years. In Australia and China, there were slightly more female than male adult respondents, whereas there were more male than female respondents in Spanish adults and adolescents, and Canadian and Dutch adolescents. More than 40% of the adult respondents hold a bachelor's degree or higher across the 5 countries, including 78% and 64% of respondents in China and Spain, respectively. Between 40% and 60% of the respondents reported having or ever having primary caregiving responsibilities for a child in all countries except for China, where 83% were reported.

Feasibility

Table 2 reports results on 3 feedback questions on DCE tasks. Overall, more than 3 quarters and two-thirds of adult respondents agreed that it was easy to understand and differentiate between the Y-5L states, whereas between 16.6% (The Netherlands) and 33.3% (China) of adults agreed it was easy to decide. The corresponding proportions were similar in adolescent samples in all 5 counties.

Modeling DCE Data

Model coefficients for adult respondents by country are presented in Table 3. We did not find any evidence of statistically significant coefficient inconsistency. In Australia, China, and Spain, all coefficients were negative, as expected, albeit with a few instances of nonsignificant negative coefficients (indicated in orange), indicating indifference between levels. These include coefficients for the movement from level 2 to level 3 on MO in Australia and between levels 1 and 2 on UA and between all level pairs on LAM except for levels 3 and 4 in China. In Canada and The Netherlands, 90% of the coefficients were statistically significant and negative; however, there are 3 instances of positive coefficients (shown in red), including the movement from level 2 to level 3 on MO in both countries and between levels 4 and 5 on LAM in The Netherlands, none of which are statistically significant. Table 3 also shows that in Australia, Canada, China, and The Netherlands, the largest decrements in utility occurred between levels 3 to 4 on all dimensions except WSU, for which the largest decrements were from level 4 to 5 and level 2 to 3 in Australia and Canada, respectively. In Spain, the largest latent utility decrements were from levels 2 to 3 in MO, LAM, and UA and from levels 4 to 5 on PD and WSU.

Table 4 shows the results from adolescents. Similar to adults, we did not find any statistically significant coefficient inconsistency. However, we observed more coefficients suggesting level indifference, including coefficients for the movement from level 2 to level 3 on MO in Australia and Canada, on LAM in all countries except for Spain, on UA in Canada and The Netherlands, on PD in Canada and China, and WSU in China. In addition, the coefficients for movement from level 1 to level 2 on LAM (in Spain), on PD (in The Netherlands) and on WSU (in Canada and China) were not

Table 1. Sample characteristics.

Demographic characteristics	Australia		Canada		China		The Netherlands		Spain	
	Adults	Adolescents	Adults	Adolescents	Adults	Adolescents	Adults	Adolescents	Adults	Adolescents
	N = 1009	*N* = 383	*N* = 908	*N* = 1054	*N* = 1319	*N* = 845	*N* = 857	*N* = 280	*N* = 1103	*N* = 956
Age										
Adolescents*										
12-14				50.09		53.14		39.60		57.32
15-17				1.71		46.86		60.40		42.68
14-15		28.72								
16-17		71.28		48.20						
Adults										
18-29	15.96		16.74		25.63		20.42		20.31	
30-39	17.64		17.95		38.29		16.80		12.42	
40-49	14.37		12.78		23.81		12.84		15.5	
50-59	16.06		19.38		10.01		18.67		23.48	
60-69	17.15		18.94		2.2		20.89		20.49	
70+	18.83		14.21		0.08		10.39		7.80	
Gender										
female	55.20	50.91	50.66	38.48	57.70	51.36	48.42	41.44	41.7	35.98
male	44.40	46.74	49.23	61.01	42.30	48.64	51.23	57.84	58.2	64.02
other	0.40	2.35	0.11	0.52			0.35	0.72	0.09	
Being a parent										
Yes	47.57		39.76		83.17		38.62		57.39	
Educational level										
University and above	42.46		42.95		78.58		41.19		63.92	

*Mixed recruitment strategy was used for adolescent sample, including self-recruiting and or recruiting from parents according to country-specific regulations. In Australia where mixed strategies were used to recruit adolescent sample (minimum self-recruitment age 14 years old), 2 survey links were used because of different consent process and different choices (14-15 and 16-17) were given to the age question. In Canada, 2 survey links were used for adolescents aged 12 to 15 years old and 16 to 17 years old because of additional discrete choice experiment tasks administered to the older adolescents (reported separately).[14]

statistically significant. The utility decrement pattern among adolescents varied across countries.

Relative Dimension Importance

Tables 3 and 4 show the latent-scale utility range of each dimension and the total scale of the model. Figure 1 shows relative dimension importance. Among adults, all 5 countries except for China shared the same ranking: PD, WSU, MO, UA, and LAM. In China, MO was considered as the most important dimension. Relative dimension importance varied across adolescents' samples in the 5 countries. MO was considered the most important dimension in Canada, China, and Spain and second most important in Australia and The Netherlands. LAM and UA were the least 2 important dimensions, except in China, where WSU was least important.

Differences Between Adults and Adolescents and Across Countries

Table 5 reports the scale heterogeneity between adult and adolescent samples. Compared with the adult sample, the adolescent sample had a statistically significantly smaller scale parameter of choice consistency ($\gamma < 1$) in all countries except Australia, suggesting increased error variance in adolescents. In all countries, we also found significant differences in choice consistency for preference for dimensions between adults and adolescents after correcting for differences in overall choice consistency. Mobility was between 4.3% and 66% more important in adolescents compared with adults. PD was between 8.1% and 47.1% less important to adolescents than adults, and WSU was between 16.6% and 38.8% less important to adolescents. This suggests that adults and adolescents had different preferences regarding Y-5L dimensions. Full model results and further description of the results are provided in Appendix Table 4 in Supplemental Materials found at https://doi.org/10.1016/j.jval.2025.07.016.

As shown in Appendix Table 5 in Supplemental Materials found at https://doi.org/10.1016/j.jval.2025.07.016, Australian, Dutch, and Spanish adults had comparable choice consistency, whereas consistency was much lower in the Chinese adult sample. Among adolescents (see Appendix Table 6 in Supplemental Materials found at https://doi.org/10.1016/j.jval.2025.07.016), choice consistency was comparable among Australia, Spain, and The Netherlands but much lower in Canada and China, suggesting more randomness in the choices. More similarities were found in adults' relative preference for dimensions across countries, compared with adolescents across countries.

Discussion

Overall, the results imply that adults can interpret the Y-5L level labels and their intended order of severity when completing latent-scale DCE tasks. We observed generally similar patterns of stated preferences for the Y-5L among adults across countries, in terms of both the relative importance of dimensions and patterns of utility decrements within dimensions. Adolescent samples were largely able to differentiate between level labels, but there was more randomness in their choices. The relative importance of dimensions and patterns of latent utility decrements by level varied in the adolescent samples across the 5 countries and were different compared with their adult counterparts.

Among adults, more than 80% of incremental dummies were statistically significant and negative in all countries. Overall results suggest that the level labels in all language versions of the Y-5L included in this study can be readily differentiated by adult

Table 2. Feasibility (time taken) and self-reported understanding and difficulty of the tasks, by adults and adolescents, in each of the 5 countries.

Completion time and feedback to the DCE tasks	Australia		Canada		China		The Netherlands		Spain	
	Adults	Adolescents	Adults	Adolescents	Adults	Adolescents	Adults	Adolescents	Adults	Adolescents
	N = 1009	*N* = 383	*N* = 908	*N* = 1054	*N* = 1319	*N* = 845	*N* = 857	*N* = 280	*N* =1103	*N* = 956
DCE time taken (in seconds), median	415	235	375	280	222	206	360	232.5	383	241
DCE time take, IQR	298-582	175-316	277-555	195-437	165-313	150-298	264-501	169.5-332	270-545	177-347
Easy to understand (%)										
Strongly agree	36.87	41.0	38.0	40.2	38.06	44.5	30.0	27.1	49.9	43.1
Agree	44.3	43.3	42.5	41.8	44.73	42.37	45.0	46.8	32.6	37.6
Neutral	10.51	13.1	12.2	15.2	13.65	10.18	14.9	15.4	10.2	11.2
Disagree	6.64	2.4	5.7	2.5	3.11	2.96	7.5	8.2	6.4	7.0
Strongly disagree	1.68	0.3	1.5	0.4	0.45	0.0	2.6	2.5	1.0	1.2
Easy to tell difference (%)										
Strongly agree	30.53	31.6	29.4	31.8	24.94	32.78	20.8	21.1	34.9	31.9
Agree	45.99	40.2	43.4	46.9	42.15	43.67	46.1	47.1	38.3	39.2
Neutral	15.56	18.8	18.8	16.6	23.73	17.99	19.8	20.4	16.1	15.9
Disagree	6.54	8.6	6.9	4.2	8.19	5.21	11.0	7.9	9.0	11.1
Strongly disagree	1.39	0.8	1.4	0.6	0.99	0.36	2.3	3.6	1.8	1.9
Difficult to decide (%)										
Strongly agree	16.95	14.9	18.5	10.4	10.46	11.24	22.6	18.6	28.7	21.3
Agree	33.7	36.8	34.6	22.8	24.87	21.3	38.9	35.7	30.8	32.1
Neutral	22.3	25.9	22.5	29.8	31.39	19.88	21.9	26.4	21.7	22.6
Disagree	20.02	19.1	18.9	25.3	23.65	27.34	13.2	13.9	13.2	15.4
Strongly disagree	7.04	3.4	5.5	11.7	9.63	20.24	3.4	5.4	5.6	8.8

DCE indicates discrete choice experiment; IQR, interquartile range.

samples. Although the level labels differed from the adult EQ-5D-5L instrument,[1,22] the percentage of statistically significant negative coefficients in our study is comparable with existing evidence on modeling the EQ-5D-5L valuation data.[23-27] Although the literature lacks any formal or definitive "threshold" percentage for judging this characteristic of the coefficients (ie, what prevalence of coefficient indifference is considered acceptable), the high levels of statistically significant negative coefficients we observed—combined with the absence of "preference reversals"—suggest no cause for concern about the Y-5L descriptive system. In the very few cases in which we observed insignificant coefficients of incremental dummies between the levels, the pattern appeared to be country specific. There could be several explanations for these insignificant coefficients. Some respondents may have found the labels for some adjacent levels difficult to differentiate between. As shown in Appendix Figure 1 in Supplemental Materials found at https://doi.org/10.1016/j.jval.2025.07.016, in the English version (and the simplified Chinese and Dutch versions, which were translated from English), when expending the severity levels for the Y-5L from the Y-3L, a new severity level label ("a little bit," level 2) was added between the original level 1 ("no problem") and level 2 ("some problem") of the Y-3L on the first 4 dimension. Some respondents may perceive the semantic differences between "a little bit" and "some" to be smaller than the difference between "no" and "a little bit."

Another explanation for the insignificant incremental dummies may be that that respondents could tell which is descriptively "worse" but did not have very strong preferences about an improvement between the 2 adjacent levels. For example, in the English version, the first 4 dimensions use identical label wording for levels 2 and 3, respectively, but we only found insignificant coefficients on MO in Australian and Canadian adults and on PD in Canadian adults—not on other dimensions with the same wording. This suggests that respondents understood the ordinal severity intended by these labels but did not have strong preference for the improvement from level 3 to 2 on MO. The insignificant coefficients for movement from level 3 to 2 are not uncommon in modeling the EQ-5D-5L valuation data in English-speaking countries[27,28] and other countries.[23-26,29] We also found insignificant coefficients between levels 4 and 5 on LAM in China and The Netherlands, and a similar pattern (or even preference reversals) between levels 4 and 5 were found in several EQ-5D-5L valuation studies.[18,23,30]

The statistical insignificance of some incremental dummies may, however, also be affected by the DCE design and sample size. This could be further investigated by using alternative DCE designs (eg, Bayesian efficient design, nonequal decrements as priors between levels or more tasks among certain adjacent level pairs) that enable us to pick up differences between adjacent levels,[21,31] or increasing sample size, particularly among adolescents. Nevertheless, it is relevant to note that, if applied in economic evaluation, all reductions in severity would represent some level of improvement in HRQoL, even if some of the coefficients of incremental dummies are not statistically significant.

Overall, preferences regarding the relative importance of Y-5L dimensions were similar among adult samples across countries. Among adults, PD was the most important dimension (except for China, where MO was the most important), and LAM was the least important dimensions in all 5 countries. The importance ranking is the same as Y-3L in Spain.[32] There are striking similarities between adults' preferences for the Y-5L dimensions observed in this study and those in Y-3L value sets[33-35] and other child HRQoL instruments, eg, CHU9D[36-38] and HUI2,[39] in which pain is typically considered the most important dimension.

We found differences in choice consistency and preference patterns between the adolescent and adult samples in each country. The error variance was larger in the adolescent sample (12-17 years) compared with the corresponding adult sample in all countries except Australia, where the adolescent sample included only those aged 14 years and above. The insignificance

Table 3. Model estimates from DCE analysis of preferences for EQ-5D-Y-5L by country, among adult respondents.

Incremental dummies	Australia	Canada	China	The Netherlands	Spain
	mean [95% CI]	mean [95% CI]	mean [95% CI]	mean [95% CI]	mean [95% CI]
MO1-MO2	−0.680 [−0.873; −0.494]	−0.582 [−0.77; −0.347]	−0.547 [−0.687; −0.432]	−0.587 [−0.783; −0.385]	−0.558 [−0.737; −0.35]
MO2-MO3	−0.012 [−0.225; 0.160]†	0.040 [−0.170; 0.215]*	−0.347 [−0.471; −0.208]	0.000 [−0.108; 0.161]*	−1.041 [−1.216; −0.869]
MO3-MO4	−1.426 [−1.667; −1.222]	−1.636 [−1.905; −1.408]	−0.8 [−0.937; −0.672]	−1.406 [−1.600; −1.195]	−0.637 [−0.83; −0.441]
MO4-MO5	−1.27 [−1.561; −1.057]	−1.081 [−1.316; −0.859]	−0.666 [−0.793; −0.530]	−0.794 [−1.005; −0.616]	−0.863 [−1.057; −0.663]
LAM1-LAM2	−0.304 [−0.522; −0.089]	−0.654 [−0.882; −0.43]	−0.122 [−0.273; 0.069]†	−0.238 [−0.461; −0.046]	−0.364 [−0.596; −0.144]
LAM2-LAM3	−0.184 [−0.371; −0.021]	−0.135 [−0.270; −0.024]	−0.073 [−0.225; 0.071]†	−0.212 [−0.361; −0.013]	−0.833 [−1.018; −0.667]
LAM3-LAM4	−1.048 [−1.271; −0.843]	−0.999 [−1.222; −0.778]	−0.529 [−0.630; −0.385]	−0.939 [−1.127; −0.780]	−0.310 [−0.478; −0.175]
LAM4-LAM5	−0.283 [−0.496; −0.096]	−0.376 [−0.619; −0.163]	−0.030 [−0.150; 0.063]†	0.061 [−0.060; 0.216]*	−0.176 [−0.358; −0.026]
UA1-UA2	−0.421 [−0.555; −0.316]	−0.499 [−0.693; −0.324]	−0.087 [−0.198; 0.013]†	−0.494 [−0.737; −0.292]	−0.446 [−0.618; −0.288]
UA2-UA3	−0.311 [−0.485; −0.092]	−0.198 [−0.410; −0.073]	−0.142 [−0.251; −0.035]	−0.275 [−0.483; −0.029]	−1.113 [−1.257; −0.929]
UA3-UA4	−0.904 [−1.187; −0.701]	−1.127 [−1.380; −0.918]	−0.418 [−0.534; −0.295]	−1.029 [−1.291; −0.833]	−0.157 [−0.297; −0.007]
UA4-UA5	−0.644 [−0.851; −0.489]	−0.529 [−0.731; −0.340]	−0.177 [−0.244; −0.116]	−0.807 [−1.001; −0.621]	−0.558 [−0.719; −0.402]
PD1-PD2	−0.996 [−1.278; −0.777]	−0.964 [−1.227; −0.749]	−0.419 [−0.536; −0.308]	−1.096 [−1.295; −0.923]	−0.978 [−1.199; −0.754]
PD2-PD3	−0.318 [−0.518; −0.140]	−0.195 [−0.401; 0.017]†	−0.168 [−0.228; −0.076]	−0.160 [−0.254; −0.063]	−1.914 [−2.185; −1.678]
PD3-PD4	−1.976 [−2.361; −1.651]	−2.201 [−2.526; −1.897]	−1.083 [−1.225; −0.943]	−2.043 [−2.343; −1.756]	−0.677 [−0.855; −0.507]
PD4-PD5	−1.892 [−2.288; −1.584]	−2.001 [−2.396; −1.682]	−0.326 [−0.470; −0.170]	−1.292 [−1.585; −0.985]	−2.000 [−2.297; −1.724]
WSU1-WSU2	−0.815 [−1.046; −0.629]	−0.641 [−0.855; −0.429]	−0.321 [−0.421; −0.217]	−0.997 [−1.233; −0.794]	−0.935 [−1.134; −0.748]
WSU2-WSU3	−1.093 [−1.369; −0.916]	−1.459 [−1.734; −1.194]	−0.167 [−0.256; −0.100]	−0.410 [−0.598; −0.243]	−1.232 [−1.453; −1.036]
WSU3-WSU4	−0.609 [−0.828; −0.403]	−0.496 [−0.751; −0.271]	−0.95 [−1.069; −0.828]	−1.655 [−1.863; −1.445]	−0.517 [−0.718; −0.337]
WSU4-WSU5	−1.397 [−1.700; −1.146]	−1.198 [−1.504; −0.930]	−0.248 [−0.385; −0.078]	−1.280 [−1.574; −1.000]	−1.337 [−1.597; −1.081]
Latent scale					
MO	−3.387	−3.259	−2.360	−2.787	−3.098
LAM	−1.819	−2.164	−0.754	−1.328	−1.683
UA	−2.281	−2.353	−0.825	−2.606	−2.273
PD	−5.182	−5.361	−1.996	−4.591	−5.570
WSU	−3.914	−3.794	−1.685	−4.342	−4.021
Model scale	−16.582	−16.931	−7.619	−15.655	−16.645

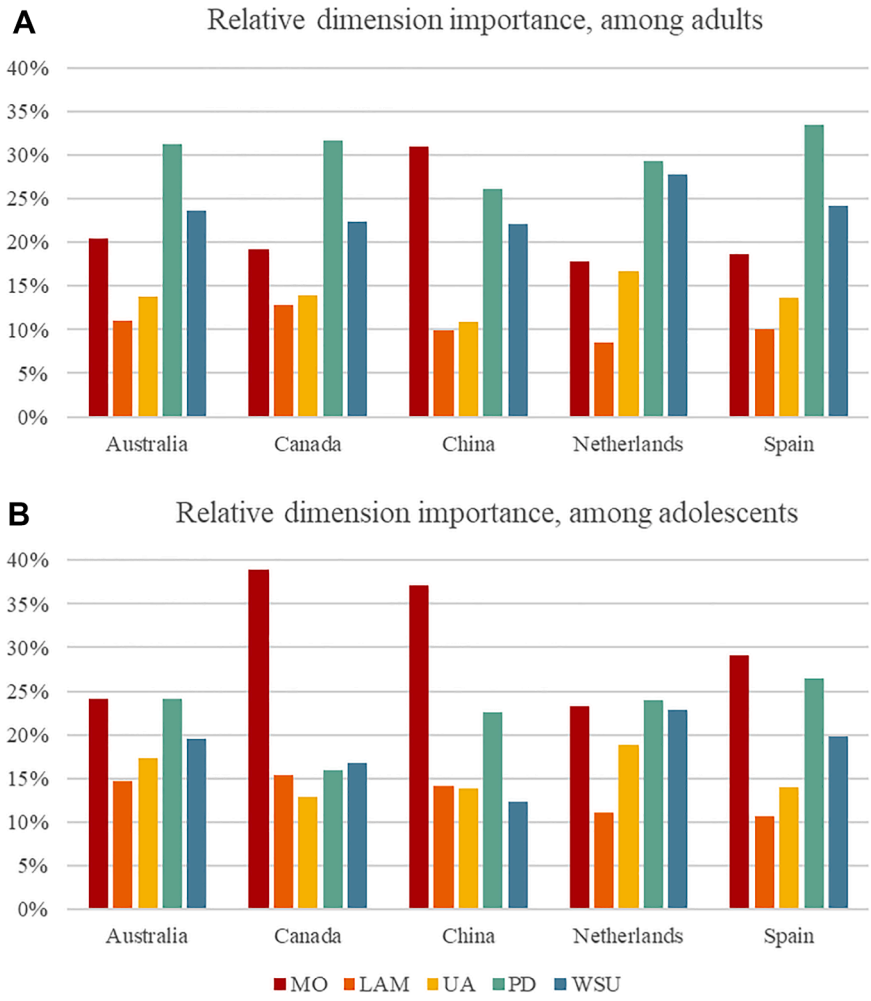
Note. Incremental dummies are used to represent movement between consecutive levels within dimensions.
Latent-scale DCE data were analyzed using a standard mixed-logit model. Model estimates were performed using Bayesian Markov-Chain Monte Carlo (MCMC) methods in Stata ("garbage_mixl" command).
DCE indicates discrete choice experiment; MO, Mobility; LAM, Looking After Myself; UA, Usual Activity; PD, Pain/Discomfort; WSU, Worried.
*Indicates 0 or positive coefficients that were not statistically significant at 5% level.
†Indicates the negative coefficients that were not statistically significant at 5% level.

Table 4. Model estimates from DCE analysis by country, among adolescent respondents.

Incremental dummies	Australia	Canada	China	The Netherlands	Spain
	mean [95% CI]	mean [95% CI]	mean [95% CI]	mean [95% CI]	mean [95% CI]
MO1-MO2	−0.612 [−0.706; −0.519]	−0.775 [−0.956; −0.602]	−0.375 [−0.526; −0.231]	−0.203 [−0.339; −0.019]	−0.618 [−0.817; −0.477]
MO2-MO3	−0.095 [−0.415; 0.107]†	0.073 [−0.084; 0.265]*	−0.209 [−0.364; −0.064]	−0.148 [−0.279; −0.048]	−0.888 [−1.032; −0.744]
MO3-MO4	−1.368 [−1.704; −0.872]	−0.828 [−1.036; −0.637]	−0.954 [−1.152; −0.781]	−1.222 [−1.503; −0.912]	−0.599 [−0.748; −0.444]
MO4-MO5	−1.407 [−1.769; −1.151]	−1.169 [−1.382; −0.969]	−0.943 [−1.122; −0.776]	−1.043 [−1.484; −0.679]	−1.013 [−1.208; −0.83]
LAM1-LAM2	−0.531 [−0.599; −0.346]	−0.301 [−0.476; −0.115]	−0.287 [−0.424; −0.126]	−0.265 [−0.388; −0.100]	−0.039 [−0.136; 0.072]†
LAM2-LAM3	−0.078 [−0.401; 0.103]†	−0.079 [−0.196; 0.071]†	0.020 [−0.066; 0.172]*	−0.234 [−0.527; 0.074]†	−0.698 [−0.843; −0.533]
LAM3-LAM4	−0.844 [−1.125; −0.652]	−0.307 [−0.495; −0.158]	−0.476 [−0.649; −0.350]	−0.731 [−0.970; −0.500]	−0.293 [−0.490; −0.121]
LAM4-LAM5	−0.675 [−0.875; −0.436]	−0.373 [−0.525; −0.231]	−0.205 [−0.353; −0.064]	−0.017 [−0.190; 0.197]†	−0.110 [−0.272; 0.031]†
UA1-UA2	−0.457 [−0.708; −0.308]	−0.142 [−0.265; −0.009]	−0.158 [−0.281; −0.045]	−0.351 [−0.547; −0.215]	−0.318 [−0.462; −0.156]
UA2-UA3	−0.116 [−0.121; −0.094]	−0.015 [−0.037; 0.001]†	−0.135 [−0.154; −0.109]	−0.062 [−0.292; 0.142]†	−0.692 [−0.82; −0.576]
UA3-UA4	−0.960 [−1.149; −0.818]	−0.268 [−0.407; −0.114]	−0.362 [−0.441; −0.273]	−0.837 [−1.175; −0.523]	−0.058 [−0.173; 0.044]†
UA4-UA5	−0.973 [−1.177; −0.833]	−0.470 [−0.642; −0.300]	−0.273 [−0.415; −0.142]	−0.866 [−1.132; −0.615]	−0.434 [−0.545; −0.305]
PD1-PD2	−0.648 [−0.81; −0.557]	−0.383 [−0.544; −0.235]	−0.361 [−0.505; −0.218]	−0.274 [−0.638; 0.015]†	−0.467 [−0.658; −0.283]
PD2-PD3	−0.162 [−0.274; −0.042]	0.061 [−0.025; 0.166]*	−0.065 [−0.219; 0.063]†	−0.657 [−0.959; −0.437]	−0.896 [−1.068; −0.715]
PD3-PD4	−1.444 [−1.803; −1.132]	−0.429 [−0.598; −0.282]	−0.948 [−1.110; −0.774]	−0.755 [−1.021; −0.477]	−0.412 [−0.571; −0.213]
PD4-PD5	−1.235 [−1.414; −0.969]	−0.351 [−0.498; −0.208]	−0.134 [−0.284; −0.003]	−1.009 [−1.454; −0.656]	−1.069 [−1.268; −0.886]
WSU1-WSU2	−0.261 [−0.442; −0.042]	−0.057 [−0.120; 0.007]†	−0.106 [−0.244; 0.066]†	−0.361 [−0.642; −0.143]	−0.315 [−0.479; −0.165]
WSU2-WSU3	−0.952 [−1.069; −0.735]	−0.588 [−0.733; −0.439]	−0.006 [−0.101; 0.160]†	−0.475 [−0.691; −0.169]	−0.802 [−0.950; −0.651]
WSU3-WSU4	−0.637 [−0.776; −0.476]	0.05 [−0.086; 0.200]*	−0.515 [−0.638; −0.390]	−1.062 [−1.282; −0.804]	−0.331 [−0.465; −0.216]
WSU4-WSU5	−0.98 [−1.261; −0.731]	−0.568 [−0.743; −0.412]	−0.199 [−0.396; −0.086]	−0.667 [−1.151; −0.251]	−0.682 [−0.869; −0.509]
Latent scale					
MO	−3.482	−2.698	−2.482	−2.616	−3.118
LAM	−2.127	−1.059	−0.948	−1.248	−1.139
UA	−2.506	−0.896	−0.928	−2.116	−1.502
PD	−3.489	−1.103	−1.509	−2.694	−2.844
WSU	−2.829	−1.163	−0.826	−2.566	−2.130
Model scale	−14.433	−6.919	−6.692	−11.240	−10.733

Note. Incremental dummies are used to represent movement between consecutive levels within dimensions.
Latent-scale DCE data were analyzed using a standard mixed-logit model. Model estimates were performed using Bayesian Markov-Chain Monte Carlo (MCMC) methods in Stata ("garbage_mixl" command).
DCE indicates discrete choice experiment; MO, Mobility; LAM, Looking After Myself; UA, Usual Activity; PD, Pain/Discomfort; WSU, Worried.
*Indicates zero or positive coefficients that were not statistically significant at 5% level.
†Indicates the negative coefficients that were not statistically significant at 5% level.

Figure 1. Relative dimension importance, results based on coefficients estimated from mixed-logit model of latent-scale DCE tasks per sample and country. (A) Relative dimension importance among adult samples in 5 countries. (B) Relative dimension importance among adolescent samples in 5 countries.



DCE indicates discrete choice experiment; LAM, Looking After Myself; MO, Mobility; PD, Pain/Discomfort; UA, Usual Activity; WSU, Worried, Sad, Unhappy.

Table 5. Relative choice consistency and dimension-specific preference between adult and adolescent samples.

Scale parameters	Australia	Canada	China	The Netherlands	Spain
Scale parameter -Relative choice consistency					
$\gamma adolescent$	1.027 (0.952 -1.109)	0.459 (0.424-0.493)*	0.888 (0.805-0.976)*	0.875 (0.798-0.959)*	0.797 (0.749-0.847)*
Scale parameter -Relative preference for dimensions					
MO (θ_1)	1.043 (0.953-1.136)	1.66 (1.542-1.787)*	1.177 (1.043-1.314)*	1.127 (1.021-1.241)*	1.397 (1.305-1.495)*
LAM(θ_2)	1.208 (1.061-1.364)*	1.086 (0.941-1.237)	1.243 (1.025-1.458)*	1.153 (0.958-1.371)	1.068 (0.941-1.208)
UA (θ_3)	1.158 (1.039-1.28)*	0.893 (0.773-1.017)	0.972 (0.841-1.104)	1.189 (0.995-1.392)	0.93 (0.836-1.024)
PD (θ_4)	0.786 (0.715-0.86)*	0.529 (0.465-0.598)*	0.793 (0.702-0.89)*	0.919 (0.8181.025)*	0.779 (0.724-0.838)*
WSU (θ_5)	0.806 (0.722-0.895)*	0.834 (0.744-0.927)*	0.815 (0.715-0.92)*	0.612 (0.518-0.709)*	0.826 (0.758-0.893)*

Note. For each country, we pooled adult data and adolescent data. Data were modeled using heteroscedastic conditional logit model to allow the variance of the error term to vary across samples. A smaller gamma (<1) means that the error variance is larger in the adolescent sample (compared with the adult sample). A larger theta (>1) for a specific dimension suggests that after accounting for differences in scale, that dimension is more important to adolescents compared with adults. Full model results and details on the method has been provided in Appendix Table 4 in Supplemental Materials found at https://doi.org/10.1016/j.jval.2025.07.016.
LAM indicates Looking After Myself; MO, Mobility; PD, Pain/Discomfort; UA, Usual Activity; WSU, Worried, Sad, Unhappy.
*Indicates 95% CI of the parameters that do not comprise 1.0.

of coefficients representing the movement from levels 2 and 3 was slightly more prevalent in adolescent samples than in the adult samples. These may indicate data quality or feasibility issues or may be suggestive of less-strong preferences underpinning adolescents' responses. However, we found that adolescents' self-reported understanding of DCE tasks, being able to tell the differences between states, and difficulty in making choices between states was similar to or better than that of adults. More than 50% of adults found it difficult to choose between the Y-5L states in the tasks. One possible explanation for these observed differences is that adult respondents were being asked to value child HRQoL from a third person perspective, which may involve more difficulty in imagining the states and choosing between them,[40,41] whereas adolescents completed the tasks from a "self" perspective. Further research, including qualitative research such as Sullivan et al,[42] is needed to explore reasons contributing to the self-reported difficulty choosing between the states.

Another key difference between adult and adolescent samples is that adolescents generally valued MO as being more important compared with PD and other dimensions, which contrasts with adult preferences. Previous studies also found different preferences between adults and adolescents for the Y-3L and CHU9D dimensions, with adolescents being more concerned with mobility and mental health when preferences are elicited using DCE or best-worst scaling.[43-45]

Our findings have several implications for the valuation for Y-5L and, arguably, other child HRQoL instruments. The different preferences apparent between adult and adolescent respondents presents further normative and methodological questions about the role of involving adolescents in valuation studies, for example, whether responses from adults and adolescents can or should be combined to create a single value set for child HRQoL or whether separate value sets based on adults and adolescent preferences are needed. In addition, the latent-scale DCE tasks used in this study do not provide values anchored at 0 and 1 as required for the estimation of quality-adjusted life-years in economic evaluation, and the results here do not provide evidence about any differential preferences for the "dead" health state between adults and adolescents. Alternative valuation methods or additional tasks (eg, time trade-off[11]or DCEs that include a duration attribute) are needed to anchor the preferences produced by latent-scale DCE. Such tasks may be more challenging for adolescents to complete. Although we did not observe inconsistent coefficients from the adolescent sample, choice consistency was lower in the adolescent sample. Further research is needed to investigate the feasibility (and minimum age) of adolescents completing latent-scale DCE and other valuation tasks.

This study used online data collection and has several limitations in terms of sample representativeness, fraudulent responses, and different strategies used to recruit adolescents. First, the Chinese adult sample is overrepresented by respondents aged 30 to 39 years and underrepresented by older adults (>60 years). This may be because the panel company was unable to reach populations with limited access to a computer or who do not have sufficient technical skills, which may be more prevalent among older people in China. In all countries except for Canada, the adult samples were overrepresented by individuals with a bachelor's degree and above, which is a recurring problem in valuation studies conducted via online surveys.[46,47] In addition, the proportion of the sample with higher education, as well as parental status, was not comparable across the 5 countries. Individuals with lower education may find greater challenges differentiating between levels; and studies have found different preferences for child health by parental status.[36,41,48] Therefore, the comparability of adult preferences across countries should be interpreted with caution. Second, we also encountered a high proportion of fraudulent data. In one subsample in Australia (adolescents 12-17 years recruited via parents), more than 60% of respondents failed at least 1 aspect of QC, which led to our decision to exclude that entire sample from the analysis. Our findings highlighted the importance of implementing QC both during and after the survey to identify fraudulent data. Third, we adopted different recruitment strategies for adolescents because of local ethics considerations and challenges in recruiting adolescents in certain countries (The Netherlands and Australia), which may limit our comparison of adolescent results across the 5 countries. In this study, we required completion using computers rather than phones. Allowing completion via phone may improve recruitment but may also result in other quality issues, eg, surveys being completed in environments not conducive to concentration, or the inability to have all relevant information on the screen at the same time. There are other challenges recruiting adolescents for online valuation studies. For example, we mainly recruited adolescents via parents (ie, parents provided consent and then "handed over" the survey to their child[ren]). However, we cannot directly establish whether adolescents or their parents were completing the adolescent survey. Asking age questions twice was helpful in identifying responses that appeared likely to be from parents. Latent-scale DCE has the advantage of allowing tasks to be self-completed online. However, with the increasing issues around fraudulent data, it is important to implement strict QC procedures to ensure data quality. Given that the error variance was larger among Chinese respondents and Canadian adolescents, further research to explore feasibility issues or other reasons for this situation may be needed. Face-to-face interviewer-led data collection may also be needed among hard-to-reach populations or adolescents to ensure sample representativeness and data quality.

Conclusions

Adults could differentiate between the Y-5L level labels in health state valuation tasks, whereas more randomness was observed in adolescents' choices. Among adults, PD was considered the most important dimension in all countries except for China; LAM and UA were the least important dimensions. Adolescents' preferences regarding the relative importance of Y-5L dimensions differed from those of adults. Further research is needed to explore adolescents' preferences, and close attention to online data quality is required.

Author Disclosures

Author disclosure forms can be accessed below in the Supplemental Material section.

Opinions expressed in this article are those of the authors and do not necessarily represent the views of the EuroQol Research Foundation.

Drs Mulhern, Devlin, and Norman are editors for *Value in Health* and had no role in the peer-review process of this article.

Supplemental Material

Supplementary data associated with this article can be found in the online version at https://doi.org/10.1016/j.jval.2025.07.016.

Article and Author Information

Accepted for Publication: July 14, 2025

Published Online: September 10, 2025

doi: https://doi.org/10.1016/j.jval.2025.07.016

Author Affiliations: Melbourne School of Population and Global Health, University of Melbourne, Melbourne, Australia (Pan, Devlin); Maths in Health, Klimmen, The Netherlands (Ramos-Goni); Warsaw School of Economics, Warsaw, Poland (Ramos-Goni); EuroQol Research Foundation, Rotterdam, The Netherlands (Roudijk); School of Pharmaceutical Science and Technology, Tianjin University, Tianjin, China (S. Xie); Department of Health Research Methods, Evidence, and Impact, McMaster University, Hamilton, Canada (S. Xie, F. Xie); Centre for Health Economics and Policy Analysis, McMaster University, Hamilton, Canada (F. Xie); Health Services Management Department, Guizhou Medical University, Guizhou, China (Yang); Centre for Health Economics Research and Evaluation, University of Technology Sydney, Sydney, Australia (Mulhern); School of Population Health, Curtin University, Perth, Australia (Norman).

Correspondence: Tianxin Pan, PhD, Melbourne School of Population and Global Health, University of Melbourne, Level 4, 207 Bourverie St, Melbourne Victoria 3010, Australia. Email: tianxin.pan1@unimelb.edu.au

Authorship Confirmation: All authors certify that they meet the ICMJE criteria for authorship.

Funding/Support: The project was funded by EuroQol Research Foundation grant 1404-RA. Open Access funding enabled and organized by CAUL and its Member Institutions.

Role of the Funder/Sponsor: The funder had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; or decision to submit the manuscript for publication.

Acknowledgment: The authors are grateful to Mike Herdman for comments on a previous draft of this article.

Ethical Approval: This study was approved by the University of Melbourne Human Research Ethics Committee (2022-24583-31815-4) in Australia, Hamilton Integrated Research Ethics Board (202-#15051) in Canada, the Ethics Committee of Guizhou Medical University (2022-249), and the Research Ethics Committee at the Canary Islands University Hospital (2022-11-24) in Spain and received a waiver by the Medical Ethical Committee of Erasmus University Medical Center (MEC-2022-0555) in The Netherlands.

Consent to Participate: Informed consent was obtained from all individual participants included in the study.

Data Availability: The data sets generated and/or analyzed during this study are available from the corresponding author on reasonable request.

REFERENCES

1. Kreimeier S, Åström M, Burström K, et al. EQ-5D-Y-5L: developing a revised EQ-5D-Y with increased response categories. *Qual Life Res*. 2019;28(7):1951–1961.
2. Wong CKH, Cheung PWH, Luo N, Cheung JPY. A head-to-head comparison of five-level (EQ-5D-5L-Y) and three-level EQ-5D-Y questionnaires in paediatric patients. *Eur J Health Econ*. 2019;20(5):647–656.
3. Fitriana TS, Purba FD, Rahmatika R, et al. Comparing measurement properties of EQ-5D-Y-3L and EQ-5D-Y-5L in paediatric patients. *Health Qual Life Outcomes*. 2021;19(1):256.
4. Cheng LJ, Schieskow S, Chen LA, Cheng JY, Herdman M, Luo N. Head-to-head comparisons of the distributional characteristics and measurement properties of the 3-level and 5-level versions of the EQ-5D-Y: a systematic review. *Value Health*. published online April 15, 2025. https://doi.org/10.1016/j.jval.2025.03.020.
5. Devlin N, Pan T, Kreimeier S, et al. Valuing EQ-5D-Y: the current state of play. *Health Qual Life Outcomes*. 2022;20(1):105.
6. Ungar WJ, Herdman M. Meeting the challenges of preference-weighted health-related quality-of-life measurement in children. *Pharmacoeconomics*. 2024;42(suppl 1):3–8.
7. Rabin R, Gudex C, Selai C, Herdman M. From translation to version management: a history and review of methods for the cultural adaptation of the EuroQol five-dimensional questionnaire. *Value Health*. 2014;17(1):70–76.
8. Powell PA, Rowen D, Keetharuth A, Mukuria C, Shah K. Who should value children's health and how? An international Delphi study. *Soc Sci Med*. 2024;355:117127.
9. Nazari JL, Pickard AS, Gu NY. Findings from a roundtable discussion with US stakeholders on valuation of the EQ-5D-Y-3L. *Pharmacoeconomics*. 2022;40(suppl 2):139–146.
10. Xie F, Xie S, Pullenayegum E, Ohinmaa A. Understanding Canadian stakeholders' views on measuring and valuing health for children and adolescents: a qualitative study. *Qual Life Res*. 2024;33(5):1415–1422.
11. Ramos-Goñi JM, Oppe M, Stolk E, et al. International valuation protocol for the EQ-5D-Y-3L. *Pharmacoeconomics*. 2020;38(7):653–663.
12. Mulhern B, Norman R, Street DJ, Viney R. One method, many methodological choices: a structured review of discrete-choice experiments for health state valuation. *Pharmacoeconomics*. 2019;37(1):29–43.
13. Wang H, Rowen DL, Brazier JE, Jiang L. Discrete choice experiments in health state valuation: a systematic review of progress and new trends. *Appl Health Econ Health Policy*. 2023;21(3):405–418.
14. Xie S, Pan T, Ramos-Goni JM, et al. Eliciting and anchoring health state preferences using discrete choice experiments among adults, adolescents, and children. *PharmacoEconomics*. 2025. https://doi.org/10.1007/s40273-025-01530-y.
15. Goodrich B, Fenton M, Penn J, Bovay J, Mountain T. Battling bots: experiences and strategies to mitigate fraudulent responses in online surveys. *Appl Econ Perspect Policy*. 2023;45(2):762–784.
16. Wang J, Calderon G, Hager ER, et al. Identifying and preventing fraudulent responses in online public health surveys: lessons learned during the COVID-19 pandemic. *PLOS Glob Public Health*. 2023;3(8):e0001452.
17. Jonker MF. Fitting garbage class mixed logit models in Stata. *Stata J*. 2024;24(3):427–445.
18. Craig BM, Pickard AS, Rand-Hendriksen K. Do health preferences contradict ordering of EQ-5D labels? *Qual Life Res*. 2015;24(7):1759–1765.
19. Gonzalez JM. A guide to measuring and interpreting attribute importance. *Patient*. 2019;12(3):287–295.
20. Jensen CE, Sørensen SS, Gudex C, Roudijk B. Stability of Danish population health preferences over time. *Value Health*. published online June 5, 2025. https://doi.org/10.1016/j.jval.2025.05.014.
21. Jonker MF, Donkers B, de Bekker-Grob E, Stolk EA. Attribute level overlap (and color coding) can reduce task complexity, improve choice consistency, and decrease the dropout rate in discrete choice experiments. *Health Econ*. 2019;28(3):350–363.
22. Reyes N, Pan T, Jones R, Dalziel K, Devlin N. Quality of Life in Kids: Key evidence to strengthen decisions in Australia (QUOKKA) project team. Understanding the transition between age-specific measures of health-related quality of life: evidence on the relationship between and comparative performance of the EQ-5D-Y-5L and EQ-5D-5L. *Value Health*. published online August 2025. https://doi.org/10.1016/j.jval.2025.04.2161.
23. Ramos-Goñi JM, Craig BM, Oppe M, et al. Handling data quality issues to estimate the Spanish EQ-5D-5L value set using a hybrid interval regression approach. *Value Health*. 2018;21(5):596–604.
24. Andrade LF, Ludwig K, Goni JMR, Oppe M, De Pouvourville G. A French value set for the EQ-5D-5L. *Pharmacoeconomics*. 2020;38(4):413–425.
25. Mai VQ, Sun S, Minh HV, et al. An EQ-5D-5L value set for Vietnam. *Qual Life Res*. 2020;29(7):1923–1933.
26. Pattanaphesaj J, Montarat T, Manuel R-GJ, Tongsiri S, Ingsrisawang L, Teerawattananon Y. The EQ-5D-5L valuation study in Thailand. *Expert Rev Pharmacoecon Outcomes Res*. 2018;18(5):551–558.
27. Norman R, Mulhern B, Lancsar E, et al. The use of a discrete choice experiment including both duration and dead for the development of an EQ-5D-5L value set for Australia. *Pharmacoeconomics*. 2023;41(4):427–438.
28. Pickard AS, Law EH, Jiang R, et al. United States valuation of EQ-5D-5L health states using an international protocol. *Value Health*. 2019;22(8):931–941.
29. Jensen CE, Sørensen SS, Gudex C, Jensen MB, Pedersen KM, Ehlers LH. The Danish EQ-5D-5L value set: a hybrid model using cTTO and DCE data. *Appl Health Econ Health Policy*. 2021;19(4):579–591.
30. Versteegh MM, Vermeulen KM, Evers SMAA, de Wit GA, Prenger R, Stolk EA. Dutch tariff for the five-level version of EQ-5D. *Value Health*. 2016;19(4):343–352.
31. Jonker MF, Roudijk B. A new and improved experimental design for the discrete choice experiment module of the EuroQol valuation technology protocol. *Value Health*. 2024;27(10):1311–1317.
32. Ramos-Goñi JM, Oppe M, Estévez-Carrillo A, Rivero-Arias O, IMPACT HTA HRQoL Group. Accounting for unobservable preference heterogeneity and evaluating alternative anchoring approaches to estimate country-specific

EQ-5D-Y value sets: a case study using Spanish preference data. *Value Health*. 2021;25(5):835–843.
33. Roudijk B, Sajjad A, Essers B, Lipman S, Stalmeier P, Finch AP. A value set for the EQ-5D-Y-3L in the Netherlands. *Pharmacoeconomics*. 2022;40(suppl 2):193–203.
34. Yang Z, Jiang J, Wang P, et al. Estimating an EQ-5D-Y-3L value set for China. *Pharmacoeconomics*. 2022;40(suppl 2):147–155.
35. Pan T, Roudijk B, Devlin N, Mulhern B, Norman R. An Australian Value Set for the EQ-5D-Y-3L. *Health Qual Life Outcomes*. 2025 Jul 15;23(1):72.
36. Xiong X, Huang L, Carvalho N, Dalziel K, Devlin N. Do the age of children and parental status matter in valuing the child health utility 9D (CHU9D)? *Pharmacoeconomics*. 2025;43(7):819–833.
37. Stevens K. Valuation of the child health Utility 9D index. *Pharmacoeconomics*. 2012;30(8):729–747.
38. Rowen D, Mulhern B, Stevens K, Vermaire JH. Estimating a Dutch value set for the pediatric preference-based CHU9D using a discrete choice experiment with duration. *Value Health*. 2018;21(10):1234–1242.
39. Torrance GW, Feeny DH, Furlong WJ, Barr RD, Zhang Y, Wang Q. Multiattribute utility function for a comprehensive health status classification system. Health Utilities Index Mark 2. *Med Care*. 1996;34(7):702–722.
40. Lipman SA, Reckers-Droog VT, Karimi M, Jakubczyk M, Attema AE. Self vs. other, child vs. adult. An experimental comparison of valuation perspectives for valuation of EQ-5D-Y-3L health states. *Eur J Health Econ*. 2021;22(9):1507–1518.
41. De Silva A, van Heusden A, Zhongyu L, et al. How do health state values differ when respondents consider adults vs children living in those states? A systematic review. *Pharmacoeconomics*. 2025;43(7):723–740.
42. Sullivan TA, Wyeth EH, Turner RM, et al. Can adolescents value the EQ-5D-Y-5L and EQ-5D-5L, and how do the values compare? A feasibility study. *Value Health*. 2025;28(6):936–944.
43. Mott DJ, Shah KK, Ramos-Goñi JM, Devlin NJ, Rivero-Arias O. Valuing EQ-5D-Y-3L health states using a discrete choice experiment: do adult and adolescent preferences differ? *Med Decis Mak*. 2021;41(5):584–596.
44. Ratcliffe J, Huynh E, Stevens K, Brazier J, Sawyer M, Flynn T. Nothing about us without us? A comparison of adolescent and adult health-state values for the child health Utility-9D using profile case best-worst scaling. *Health Econ*. 2016;25(4):486–496.
45. Dalziel K, Catchpool M, García-Lorenzo B, Gorostiza I, Norman R, Rivero-Arias O. Feasibility, validity and differences in adolescent and adult EQ-5D-Y health state valuation in Australia and Spain: an application of best-worst scaling. *Pharmacoeconomics*. 2020;38(5):499–513.
46. Ozdemir S, Quaife M, Mohamed AF, Norman R. An overview of data collection in health preference research. *Patient*. 2025;18(4):303–315.
47. Cao Y, Xu J, Norman R, et al. Chinese utility weights for the EORTC cancer-specific utility instrument QLU-C10D. *Qual Life Res*. 2024;33(12):3335–3349.
48. Kreimeier S, Mott D, Ludwig K, Greiner W, IMPACT HTA HRQoL Group. EQ-5D-Y value set for Germany. *Pharmacoeconomics*. 2022;40(Suppl 2):217–229.