



Valuing EQ-5D-Y-3L Health States Using a Discrete Choice Experiment: Do Adult and Adolescent Preferences Differ?

David J. Mott¹, Koonal K. Shah², Juan Manuel Ramos-Goñi,
Nancy J. Devlin, and Oliver Rivero-Arias³

Background. An important question in the valuation of children's health is whether the preferences of younger individuals should be captured within value sets for measures that are aimed at them. This depends on whether younger individuals can complete valuation exercises and whether their preferences differ from those of adults. This study compared the preferences of adults and adolescents for EQ-5D-Y-3L health states using latent scale values elicited from a discrete choice experiment (DCE). **Methods.** An online DCE survey, comprising 15 pairwise choices, was provided to samples of UK adults and adolescents (aged 11–17 y). Adults considered the health of a 10-year-old child, whereas adolescents considered their own health. Mixed logit models were estimated, and comparisons were made using relative attribute importance (RAI) scores and a pooled model. **Results.** In total, 1000 adults and 1005 adolescents completed the survey. For both samples, level 3 in pain/discomfort was most important, and level 2 in self-care the least important, based on the relative magnitudes of coefficients. The RAI scores (normalized on self-care) indicated that adolescents gave less weight relative to adults to usual activities (1.18 v. 1.51; $P < 0.05$), pain/discomfort (1.77 v. 3.12; $P < 0.01$), and anxiety/depression (1.64 vs. 2.65; $P < 0.01$). The pooled model indicated evidence of differences between the two samples in both levels in pain/discomfort and anxiety/depression. **Limitations.** The perspective of the DCE task differed between the 2 samples, and no data were collected to anchor the DCE data to generate value sets. **Conclusions.** Adolescents could complete the DCE, and their preferences differed from those of adults taking a child perspective. It is important to consider whether their preferences should be incorporated into value sets.

Keywords

discrete choice experiment, EQ-5D-Y, UK, valuation exercise, youth health state valuation

Date received: January 6, 2021; accepted: January 12, 2021

The EQ-5D-Y-3L is a patient-reported outcome measure that was designed to measure the health-related quality of life (HRQOL) of children and adolescents.^{1,2} However, unlike the adult versions of the same instrument (EQ-5D-3L and EQ-5D-5L), value sets for translating EQ-5D-Y-3L responses to health state utilities are unavailable in most countries.³ This means that in economic evaluations of treatments aimed at younger populations, it is not possible to estimate quality-adjusted life-years (QALYs) based on EQ-5D-Y-3L data, unless a value set from another EQ-5D instrument is used. As the

EQ-5D-Y-3L instrument contains three severity levels, it is typical for EQ-5D-3L value sets to be used in lieu of a value set for the EQ-5D-Y-3L. However, evidence suggests that this is inappropriate because the 2 instruments are worded differently, and the perspective used in an EQ-5D-Y-3L valuation task may result in a value set

Corresponding Author

Oliver Rivero-Arias, National Perinatal Epidemiology Unit (NPEU), Nuffield Department of Population Health, University of Oxford, UK (oliver.rivero@npeu.ox.ac.uk).

with significantly different characteristics to an EQ-5D-3L value set.^{4,5} It is therefore important that preference elicitation studies are conducted to generate a value set for the EQ-5D-Y-3L instrument.

Although the development of value sets for the EQ-5D-Y-3L is desirable, valuing health states for children poses unique challenges. These challenges include the identification of the most appropriate methodology, perspective, and sample to be used in the health state valuation exercise.³ A methodological research program was pursued by the EuroQol Research Foundation to look into these issues, of which this study was part, which culminated in the development of a protocol for the valuation of EQ-5D-Y-3L.⁶ The protocol recommends that a discrete choice experiment (DCE) is used to obtain latent scale values, which should be anchored onto the QALY scale using composite time trade-off data. In both tasks, values are to be elicited from adult members of the general population, with the tasks framed as follows: “Considering your views about a 10-y-old child, what do you prefer?” Nonetheless, scientific protocols are subject to future improvements following further research, and one such area for further research relates to the choice of sample.

The question of whose preferences to elicit for health state valuation has been considered in many past studies, both theoretical and empirical.^{7–12} The normative debate centers around whether values should be sought from

individuals who are experiencing the health state that is being evaluated. In practice, most value sets are based on the preferences of adult (18+) members of the general population who have not, necessarily, experienced the health states under evaluation.⁸ This is often justified in countries with publicly funded health care systems on the basis that they represent taxpayers’ preferences.^{7,8} It has also been argued that general population preferences are important because all members of the public are potential users of health care services.

In the context of valuing children’s health, the debate differs. Given that the EQ-5D-Y-3L instrument was designed for self-completion by younger individuals, the question arises as to whether younger individuals, rather than adults, should value EQ-5D-Y-3L health states. On one hand, in many countries, younger individuals (<18) are typically not taxpayers and are often ineligible to vote, suggesting that adults’ preferences may be more appropriate. Furthermore, although the adult general population values are not experience based,¹³ nor are they necessarily fully informed,¹⁴ adults are arguably better informed about the impact of ill health on HRQOL than younger people, on average. On the other hand, adults are not potential users of health care services for younger people, nor can they feasibly be experiencing a child’s health state when completing a valuation task. Therefore, it can still be argued that it would be inappropriate to base resource allocation decisions that affect younger populations on adult preferences alone. In addition, there is limited guidance from international agencies around how to generate QALYs, and hence utilities, for use in the health technology assessment of interventions affecting young populations.³

The relevance of this debate ultimately depends on the feasibility of eliciting preferences for health states from younger individuals, and the existence of differences between their preferences and those of adults. A review by Crump et al.¹⁵ identified a total of 26 studies (up to May 2015) that elicited preferences for health states from children. Most studies used time trade-off or standard gamble, and only a third reported that the exercises were feasible in the target population. However, recent advances in health state valuation focus on less cognitively complex exercises such as DCEs and best-worst scaling (BWS).^{16,17} Recently published studies have elicited the preferences of children using BWS.^{18–20} All 3 studies found BWS to be feasible. Furthermore, 2 of the studies compared the preferences of adolescents with adults. Ratcliffe et al.¹⁸ found that Australian adults placed less weight on impairments in mental health and more weight on higher levels of pain relative to adolescents when valuing CHU9D. In alignment with

Office of Health Economics, London, UK (DJM, KKS, NJD); PHMR, London, UK (KKS); Maths in Health, Rotterdam, The Netherlands (JMR-G); Melbourne School of Population and Global Health, University of Melbourne, Australia (NJD); National Perinatal Epidemiology Unit (NPEU), Nuffield Department of Population Health, University of Oxford, Oxford, UK (OR-A). The author(s) declared the following potential conflicts of interest with respect to the research, authorship, and/or publication of this article: All authors received grants from the EuroQol Research Foundation during the conduct of the study and are members of the EuroQol Group. David J. Mott is an employee of the Office of Health Economics, a registered charity and independent research organization, that receives funding from a variety of private and public sector sources. Koonal K. Shah is an employee of PHMR, which receives research and consulting income from a variety of sources, including the EuroQol Research Foundation and pharmaceutical industry clients. Nancy Devlin, Juan Manuel Ramos-Goñi, and Oliver Rivero-Arias have no conflicts of interest to disclose. The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: Financial support for this study was provided entirely by grants from the EuroQol Research Foundation (EQ projects 2016630 and 2016631). The funding agreement ensured the authors’ independence in designing the study, interpreting the data, writing, and publishing the report. The following author was employed by the sponsor during the conduct of the study: Juan Manuel Ramos-Goñi.

this finding, Dalziel et al.¹⁹ found that Australian adults place less weight on being very worried, sad, or unhappy and more weight on having pain or discomfort relative to adolescents when valuing EQ-5D-Y-3L. However, they also found that Spanish adults and adolescents aligned in their greatest weight being placed on pain or discomfort.

Thus, there is some evidence that health state valuation is feasible with younger individuals and that preferences may indeed differ between younger individuals and adults. However, the evidence base is limited, and no comparison studies have been published to date using DCE methodology, despite its increasing prominence in health state valuation.¹⁷ Therefore, this study sought to compare latent scale values from a DCE elicited from a sample of adults and a sample of adolescents in the UK. The objectives were to determine whether a DCE is feasible as a health state valuation method in an adolescent population and to determine if preferences differ between an adult and an adolescent sample.

Method

Overview

Two online surveys were administered, one to a sample of adults and another to a sample of adolescents. The surveys were developed in collaboration with epiGenesys, a software development company. Both surveys comprised the following elements (in order): screening questions, information sheet and informed consent, self-reported health using EQ-5D-Y-3L and visual analogue scale (EQ-VAS), instructions, 16 paired comparison tasks, 3 debrief questions, and background questions. The background questions differed slightly between samples, and adults were asked some additional debrief questions relating to the framing of the DCE task. The full surveys can be found in the supplementary materials. Ethics approval to conduct this study was obtained from the Medical Sciences Inter-Divisional Research Ethics Committee (IDREC) at the University of Oxford (reference: R47732/RE002).

The remainder of this section will describe the sample recruitment, the EQ-5D-Y-3L instrument, and the various components of the DCE design.

Sample

Data were collected from a sample of adult members of the UK general public (target sample size: $n = 1000$) as well as a sample of UK adolescent members of the general public aged between 11 and 17 y (target sample size:

$n = 1000$). All adult respondents were members of an online panel managed by a market research agency, Survey Sampling International. The adolescent respondents were the children of adult panel members. Selected panel members who had not been contacted for the adult survey but who had been identified as having children according to the agency's database were contacted with an invitation for their children to take part.

Quotas, combined with a targeted recruitment strategy, were used to ensure that the sample was representative of the general population in terms of gender, age, social grade (adult sample only), and nation (within the UK; adult sample only). Respondents were awarded "panel points" (which can be redeemed for cash vouchers and other rewards) following completion of the survey. Based on piloting that suggested that the survey should take 7.5 min to complete on average, it was agreed to exclude any respondents completing the entire survey in less than 2.5 min (i.e., one-third of that time) on data-quality grounds.

EQ-5D-Y-3L

The EQ-5D-Y-3L instrument consists of 5 dimensions, each with 3 severity levels, as detailed in Table 1. In contrast to the commonly used EQ-5D-3L instrument, the "self-care" dimension is labeled as "looking after myself," and the "anxiety/depression" dimension is labeled as "feeling worried, sad, or unhappy," as these were deemed to be more easily understood by younger individuals. However, in the interest of brevity, the "traditional" labels/codes are used throughout this article. A total of 243 (3^5) health states are possible when using the EQ-5D-Y-3L.

Discrete Choice Experiment

Overview

The DCE required respondents to make a choice between 2 EQ-5D-Y-3L health states labeled as options A and B. All 5 dimensions of the EQ-5D-Y-3L were included as attributes, along with the 3 severity levels for each dimension.

In terms of the perspective of the choice tasks, the adult sample were asked, "Considering your views about a 10-y-old child: which do you prefer, A or B?" This choice of perspective is not a straightforward one and has been shown to potentially influence results.³ Our choice of a 10-y-old child perspective was based on past studies and the fact that the EQ-5D-Y-3L instrument is

Table 1 EQ-5D-Y-3L Instrument

Dimension	Levels	Coding
Mobility (walking about)	I have no problems walking about	MO1
	I have some problems walking about	MO2
	I have a lot of problems walking about	MO3
Looking after myself ^a	I have no problems washing or dressing myself	SC1
	I have some problems washing or dressing myself	SC2
	I have a lot of problems washing or dressing myself	SC3
Doing usual activities (for example, going to school, hobbies, sports, playing, doing things with friends or family)	I have no problems doing my usual activities	UA1
	I have some problems doing my usual activities	UA2
	I have a lot of problems doing my usual activities	UA3
Having pain or discomfort	I have no pain or discomfort	PD1
	I have some pain or discomfort	PD2
	I have a lot of pain or discomfort	PD3
Feeling worried, sad, or unhappy ^b	I am not worried, sad or unhappy	AD1
	I am a bit worried, sad or unhappy	AD2
	I am very worried, sad or unhappy	AD3

^aReferred to as “self-care” by convention.

^bReferred to as “anxiety/depression” by convention.

intended for use in a population aged 8 to 15 y.⁴⁻⁶ In contrast, the adolescent sample were asked, “Which do you prefer, A or B?” No opt-out or indifference options were provided. The visual presentation of the choice tasks (see Figure 1) was designed to mimic the format used for DCE tasks in the EuroQol Group’s international EQ-5D-5L valuation protocol.²¹

Experimental Design

The experimental design for the DCE was a Bayesian efficient design allowing for the estimation of main effects and all 2-way interactions, with a minimal number of unrealistic health states, overlapping of health states in 2-dimensional levels, and good level and utility balance. The design phase was split into 2 phases. In the first phase, an initial design was produced and tested in a soft launch with 127 participants. Subsequently, simulations were conducted to select the final design, which incorporated the priors estimated in the soft launch. The final design, which can be found in the supplementary materials, contained 150 pairs and was split into 10 blocks, resulting in 15 tasks per respondent. The design used in this study has since been adopted by the recent international EQ-5D-Y-3L protocol.⁶

An additional choice scenario was added as a dominance test (see section Logic Checks), resulting in 16 tasks per respondent. However, this scenario was not part of the experimental design and was therefore not included in the choice analysis. In addition, the order in

which the alternatives were displayed in the survey was randomized to minimize left-right bias.²²

Discrete Choice Modeling

Choice data are typically modeled using a random utility model framework, in which the utility obtained by decision maker n choosing alternative j is given by equation 1.

$$U_{nj} = V_{nj} + \varepsilon_{nj} \quad (1)$$

where V_{nj} is an observable (or deterministic) component made up of the attributes of the alternatives X_{nj} and observable characteristics of the decision maker Z_n . ε_{nj} is an unknown (or stochastic) component and treated as random. It follows that the probability that the decision maker n chooses alternative i is:

$$P_{ni} = \Pr(\varepsilon_{nj} - \varepsilon_{ni} < V_{ni} - V_{nj}) \forall j \neq i \quad (2)$$

Different choice models are obtained from different assumptions about the distribution of the random terms. The most commonly used choice model, the multinomial logit (MNL), assumes that the random terms are independent and identically distributed (IID) type one extreme value and suffers from the restrictive independence of irrelevant alternatives (IIA) property. In addition, the MNL model assumes that preferences are homogenous across individuals, unless systematic differences across

Adult

Considering your views about a 10 year old child: which do you prefer, A or B?

- no problems walking about
- a lot of problems washing or dressing
- a lot of problems doing usual activities
- some pain or discomfort
- very worried, sad or unhappy

- some problems walking about
- some problems washing or dressing
- a lot of problems doing usual activities
- some pain or discomfort
- a bit worried, sad or unhappy

A

B

Adolescent

Which do you prefer, A or B?

- no problems walking about
- a lot of problems washing or dressing
- a lot of problems doing usual activities
- some pain or discomfort
- very worried, sad or unhappy

- some problems walking about
- some problems washing or dressing
- a lot of problems doing usual activities
- some pain or discomfort
- a bit worried, sad or unhappy

A

B

Figure 1 Example choice scenarios.

participants are included in the observable component of utility (e.g., gender, age). Because of this, alternative models are typically preferred on the basis that actual choice behavior can be better represented by flexible models that attempt to control for various sources of random heterogeneity.

There are 2 types of random heterogeneity that alternative choice models typically try to account for. The first, preference heterogeneity, occurs if individuals' preferences differ from one another for reasons beyond differences in observable characteristics. The second, scale heterogeneity, is a specific type of correlation across utility coefficients. It occurs when the impact of factors not included (in the model) affect individuals differently, giving the impression that some individuals' responses are "more random" than others.²³ Several suggestions exist in the econometric literature for incorporating preference and scale heterogeneity in a discrete choice model.²⁴⁻²⁶ In this study, we selected the MIXL model with correlated parameters as the basis of our comparison between adult and adolescent samples. Correlated MIXL models allow parameters to be estimated for each respondent in the sample and hence take preference and scale heterogeneity into account.²⁴

In the choice models, a linear, additive utility function was estimated with all variables dummy coded and "level 1s" used as base levels, as in equation 3. In the MIXL models, each parameter was modeled as random and normally distributed.

$$V_j = \beta_1 MO2 + \beta_2 MO3 + \beta_3 SC2 + \beta_4 SC3 + \beta_5 UA2 + \beta_6 UA3 + \beta_7 PD2 + \beta_8 PD3 + \beta_9 AD2 + \beta_{10} AD3 \quad (3)$$

To compare the models between samples, predicted probabilities (i.e., estimated by the models) and observed probabilities for the 150 DCE pairs (i.e., the choices made) were compared.

Preference Comparisons

An increasingly well-documented issue when comparing the preferences of different samples using DCE data is the confounding between preference and scale.^{27,28} It is possible to determine whether differences in scale exist between samples using the Swait-Louviere test; this is typically conducted using MNL models and is the approach used in this study.²⁷ However, this does not allow for both scale and preference heterogeneity to be

controlled for. In fact, it has been argued that it is impossible to disentangle the two.²³ We therefore use 2 approaches to compare the preferences of the 2 samples while controlling for scale and preference heterogeneity.

The first approach is to examine relative attribute importance (RAI) scores by dimension. This approach involves estimating the utility range for each attribute and subsequently applying a normalization to enable sample comparisons; in this case, an attribute-based normalization was used,²⁹ as in equation 4.

$$RAI_X = \frac{\beta_X}{\beta_Y} \cdot S \quad (4)$$

RAI_X is the RAI score for attribute X . β_X is the coefficient for the level 3 variable of attribute X (this provides the utility range for attribute X , as long as level 3 is worse than level 2). Using the same logic, β_Y is the coefficient for the level 3 variable of attribute Y , which is the attribute chosen for the normalization for all RAI scores (in this case, the least important attribute overall). S is a scaling factor; in this case, $S = 1$ was chosen. Thus, the RAI score for attribute X is ≥ 1 and indicates the extent to which attribute X is preferred to attribute Y , which can be compared between samples. The delta method was used to estimate standard errors associated to the RAI scores.

The second approach is to estimate a pooled model that includes additional interaction parameters that interact each variable with a sample dummy $Adol$ ($= 1$ if the respondent is in the adolescent sample), as in equation 5. In the MIXL model, the main parameters were modeled as random and normally distributed; however, the interaction parameters were modeled as fixed.

$$\begin{aligned} V_j = & \beta_1 MO2 + \beta_2 MO3 + \beta_3 SC2 + \beta_4 SC3 + \beta_5 UA2 \\ & + \beta_6 UA3 + \beta_7 PD2 + \beta_8 PD3 \\ & + \beta_9 AD2 + \beta_{10} AD3 + \beta_{11} (MO2 \times Adol) \\ & + \beta_{12} (MO3 \times Adol) + \beta_{13} (SC2 \times Adol) \\ & + \beta_{14} (SC3 \times Adol) + \beta_{15} (UA2 \times Adol) \\ & + \beta_{16} (UA3 \times Adol) + \beta_{17} (PD2 \times Adol) \\ & + \beta_{18} (PD3 \times Adol) + \beta_{19} (AD2 \times Adol) \\ & + \beta_{20} (AD3 \times Adol) \end{aligned} \quad (5)$$

In this model, the coefficients on the main parameters reflect the preferences of the adult sample. The coefficients on the interaction terms indicate the shift in the parameter distribution for the adolescent sample. Thus, mean coefficients for the adolescent sample can be derived by adding the interaction coefficient to the main

parameter coefficient, and statistically significant interaction terms indicate differences in preferences between the 2 samples.

Logic Checks

In addition to the choice sets drawn from the experimental design, all respondents completed one further fixed pair, in which one health state (11122) could be considered to logically dominate the other (22233). This was included as a dominance test to examine data quality³⁰; these data were excluded from the modeling exercise.

Another data quality check involves an examination of the proportion of respondents who chose health states with a lower level sum score (LSS). The LSS for any given health state is calculated by taking the sum of its levels.^{31,32} For example, the LSS for 11111 is 5 and the LSS for 33333 is 15. It follows that a higher LSS corresponds to a more severe state. The larger the difference in LSS between any 2 health states, the greater the expectation that a respondent would choose the option with the lower LSS. However, it should be noted that this is a limited approach, as not all dimensions will be valued equally.

Results

Response Rates and Sample Composition

Main data collection for the adult sample was carried out in February/March 2017. Of the 1187 individuals who accessed the survey, 87 (7.3%) declined consent, 72 (6.1%) started but did not provide a complete set of data, and 28 (2.4%) completed the survey in less than the agreed minimum time of 2.5 min. This left a total of 1000 respondents for analysis. The main data collection for the adolescent sample was carried out in November/December 2017. Of the 1449 individuals who accessed the survey, 192 (13.2%) were outside the eligible age range, 136 (9.4%) declined consent, 56 (3.9%) started but did not provide a complete set of data, and 60 (4.1%) completed the survey in less than the agreed minimum time. This left a total of 1005 respondents for analysis.

Background characteristics of the 2 samples are summarized in Table 2. Quotas were used to generate representative samples of the UK general population. By construction, the adult sample was representative of the UK general population in terms of age group, gender, social grade and nation, and the adolescent sample was representative of the UK adolescent population in terms of age group (i.e., split between 11- to 14-y-olds and

Table 2 Sample Background Characteristics

		Adult Sample, <i>n</i> = 1000 (%)	Adolescent Sample, <i>n</i> = 1005 (%)	General Population ^a
Gender	Female	512 (51.2%)	494 (49.1%)	51%
	Male	488 (48.8%)	511 (50.9%)	49%
Age, y	11	N/A	78 (7.8%)	15%
	12		132 (13.1%)	14%
	13		181 (18.0%)	14%
	14		174 (17.3%)	14%
	15		162 (16.1%)	14%
	16		139 (13.8%)	14%
	17		139 (13.8%)	15%
	18–29	199 (19.9%)	N/A	20%
	30–44	272 (27.2%)		25%
	45–59	255 (25.5%)		26%
	60+	274 (27.4%)		30%
Nation	England	845 (84.5%)	857 (85.3%)	84%
	Scotland	85 (8.5%)	72 (7.2%)	16%
	Wales	49 (4.9%)	58 (5.8%)	
	Northern Ireland	21 (2.1%)	18 (1.8%)	
Social grade ^b	Higher (ABC1)	542 (54.2%)		55%
	Lower (C2DE)	458 (45.8%)		44%
Family affluence scale	Low score (0–2)		30 (3%)	N/A
	Medium score (3–5)		456 (45%)	N/A
	High score (6–9)		519 (52%)	N/A
Self-reported health (EQ-5D-Y-3L)	Health state 11111	148 (14.8%)	587 (58.4%)	N/A
	All other health states	852 (85.2%)	418 (41.6%)	N/A

^aGeneral population gender stats refer to percentage of the entire UK population, whereas age stats refer to percentage of the 11- to 17-y-old and 18+ y populations, respectively. General population gender and age stats were taken from the Office for National Statistics, 2017. Population estimates for UK, England and Wales, Scotland and Northern Ireland (data set). Available from: <https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/datasets/populationestimatesforukenglandandwalescotlandandnorthernireland> (accessed January 15, 2021). General population social grade stats taken from National Readership Survey, 2016. Social grade. Available from: <http://www.nrs.co.uk/nrs-print/lifestyle-and-classification-data/social-grade/> (accessed January 15, 2021).

^bHigher (ABC1) indicates that the chief income earner in the respondent's household works in a managerial, administrative, or professional occupational group; lower (C2DE) indicates that they are a skilled, semi-skilled, or unskilled manual worker, stated pensioner, casual/lowest grade worker, or unemployed with state benefits only.

15- to 17-olds) and gender. Self-reported EQ-5D-Y-3L indicated that the adult sample was in worse health than the adolescent sample (15% and 58% self-reporting 11111, respectively); a figure summarizing the self-reported EQ-5D-Y-3L responses can be found in the supplementary materials.

After excluding speeders, the mean (median) amount of time taken to complete the survey was 11 min (7 min) for the adult sample and 9 min (6 min) for the adolescent sample.

DCE Results

The main regression results can be found in Table 3. The coefficients for every dimension level included in the MNL and MIXL models were negative and statistically

significant at the 1% level. In addition, in the MIXL models for both samples, every standard deviation (except MO2 in the adolescent sample) was statistically significant at the 1% level, indicating evidence of preference heterogeneity, spanning most dimensions and levels of the EQ-5D-Y-3L. Such a result is an indication of the suitability of the MIXL model over the MNL model in this study, which is also captured by better fit of the data with lower log-likelihood and BIC. When implementing the Swait-Louviere test using the MNL model, it was found that differences in coefficients were not explained solely by differences in scale, indicating that differences in preferences do exist between the 2 samples.

Based on the MIXL model results, the rank order of the various dimensions levels (based on the relative magnitudes of the coefficients within each model) is similar

Table 3 Discrete Choice Modeling Estimation Results, by Sample^a

	Adults				Adolescents			
	MNL		MIXL		MNL		MIXL	
	Est	SE	Est	SE	Est	SE	Est	SE
MO2	-0.158	0.048	-0.408	0.067	-0.255	0.046	-0.407	0.062
MO3	-0.611	0.079	-1.200	0.114	-0.896	0.074	-1.419	0.106
SC2	-0.247	0.039	-0.365	0.057	-0.196	0.037	-0.332	0.053
SC3	-0.592	0.065	-0.979	0.090	-0.723	0.063	-1.123	0.090
UA2	-0.372	0.042	-0.607	0.061	-0.310	0.040	-0.496	0.054
UA3	-0.894	0.051	-1.478	0.090	-0.819	0.051	-1.328	0.085
PD2	-0.581	0.043	-1.128	0.077	-0.492	0.039	-0.818	0.060
PD3	-1.553	0.075	-3.057	0.159	-1.414	0.064	-2.319	0.114
AD2	-0.602	0.043	-0.951	0.070	-0.363	0.039	-0.566	0.056
AD3	-1.504	0.069	-2.592	0.131	-1.310	0.065	-2.162	0.114
Number of parameters	10		65		10		65	
σ (MO2)			0.547	0.112			0.086	0.151
σ (MO3)			1.246	0.158			1.166	0.186
σ (SC2)			0.240	0.083			0.481	0.075
σ (SC3)			0.806	0.123			1.148	0.119
σ (UA2)			0.702	0.082			0.615	0.077
σ (UA3)			1.171	0.097			1.326	0.121
σ (PD2)			1.100	0.080			0.865	0.081
σ (PD3)			2.560	0.138			1.996	0.140
σ (AD2)			0.900	0.095			0.722	0.087
σ (AD3)			2.048	0.121			1.952	0.138
Number of choices	15,000		15,000		15,075		15,075	
Number of participants	1000		1000		1005		1005	
LL	-8300		-7225		-8907		-8013	
BIC	16,696		15,074		17,910		16,651	

^a Bold estimates are statistically significant at 1%. MO2: I have some problems walking about; MO3: I have a lot of problems walking about; SC2: I have some problems washing or dressing myself; SC3: I have a lot of problems washing or dressing myself; UA2: I have some problems doing my usual activities; UA3: I have a lot of problems doing my usual activities; PD2: I have some pain or discomfort; PD3: I have a lot of pain or discomfort; AD2: I am a bit worried, sad, or unhappy; AD3: I am very worried, sad, or unhappy; σ : standard deviation; LL: log-likelihood; BIC: Bayesian information criteria; MNL: multinomial logit; MIXL: mixed logit with correlated parameters; MIXL models were estimated using 5000 Halton draws and the Broyden-Fletcher-Goldfarb-Shanno (BFGS) algorithm. Correlated parameters MIXL used MIXL with uncorrelated parameters as starting values.

between the 2 samples but not entirely consistent. In both samples, the 2 most important attribute levels were PD3 and AD3, respectively. Similarly, the 4 least important attribute levels were AD2, UA2, MO2, and SC2, respectively. However, there was some disagreement between the 2 samples in between. In the adolescent sample, MO3 was considered more important relative to UA3 (the third and fourth most important levels), whereas the reverse was true in the adult sample. Similarly, in the adolescent sample, SC3 was considered more important relative to PD2 (the fifth and sixth most important levels), whereas the reverse was true in the adult sample.

Table 4 presents the RAI scores for each sample and associated standard errors. The attribute-based normalization was employed using the least important attribute overall, which was SC. This is because the coefficient for

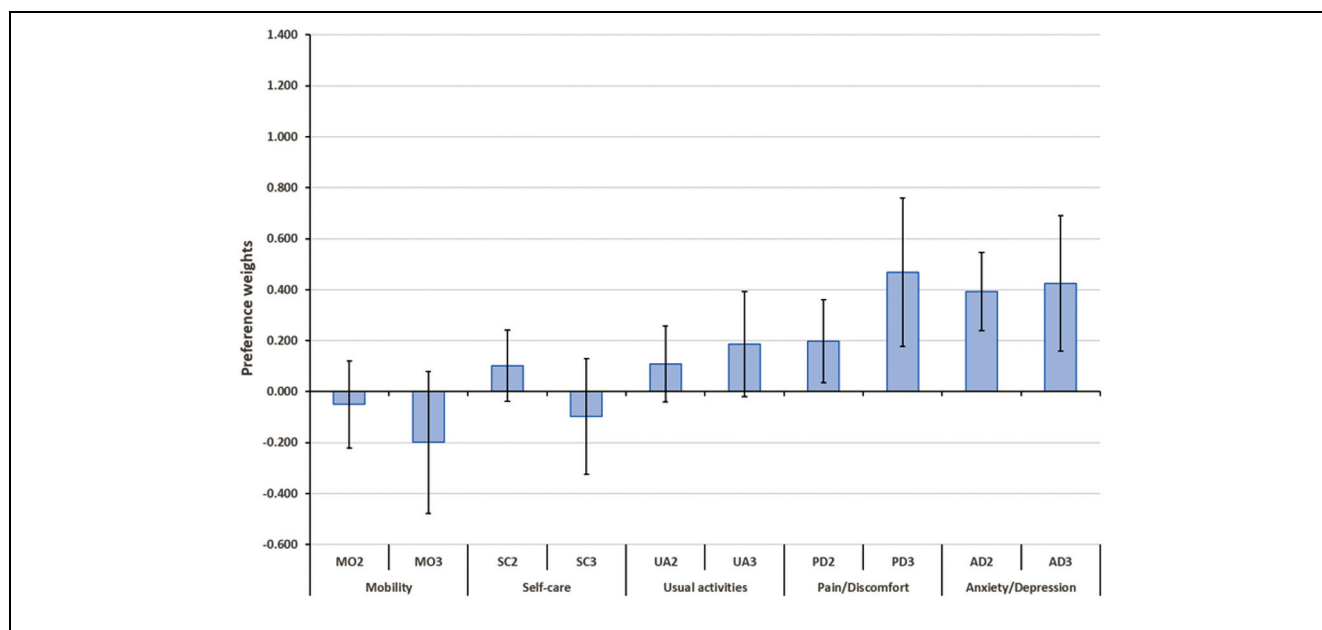
SC3 was the smallest in magnitude relative to all other level 3 coefficients in both MIXL models in Table 3. The interpretation of the RAI scores is as follows: the score of 3.12 for PD for adults indicates that respondents in this sample considered PD to be more than 3 times as important as SC on average. Comparatively, adolescent respondents considered PD to be 2.07 times more important than SC on average. The difference between the 2 samples is statistically significant ($P < 0.01$). A similar, albeit less substantial, statistically significant difference can be seen between the 2 samples in relation to AD ($P < 0.01$) and, to a lesser extent, UA ($P < 0.05$). However, there is no significant difference in the extent to which MO is preferred to SC between the 2 samples.

Unlike RAI scores, the pooled model can indicate differences in preferences between the samples by dimension

Table 4 Relative Attribute Importance Scores by Sample and RAI Differences with 95% Confidence Intervals^a

	Adults		Adolescents		RAI Difference (95% Confidence Interval)	P Value
	RAI	SE	RAI	SE		
Mobility	1.23	0.08	1.26	0.07	-0.04 (-0.25 to 0.17)	0.719
Self-care	1.00		1.00			
Usual activities	1.51	0.12	1.18	0.08	0.33 (0.04 to 0.61)	0.025
Pain/discomfort	3.12	0.27	2.07	0.15	1.06 (0.46 to 1.66)	0.001
Anxiety/depression	2.65	0.23	1.93	0.14	0.72 (0.19 to 1.26)	0.008

^aRelative attribute importance scores were calculated based on the MIXL models from Table 3. An attribute-based normalization was applied using the least important attribute (SC) and a scaling factor of 1. Standard errors (SE) were calculated using the Delta method.

**Figure 2** Mean preference weights for adolescent interaction coefficients from the pooled mixed logit model and associated 95% confidence intervals.

levels, rather than dimensions alone. The adolescent sample interactions from the pooled MIXL model are illustrated in Figure 2. A table with all coefficients from the pooled model is presented in the supplementary material. The statistically significant differences between the 2 samples relate to PD2 (less important in the adolescent sample; $P < 0.05$), PD3, AD2, and AD3 (all less important in the adolescent sample; $P < 0.01$). As is the case with the RAI scores, this suggests that respondents in the adolescent sample put less weight on attributes such as PD and AD relative to the adult sample. The interaction of UA3 was borderline statistically significant, indicating less importance in the adolescent sample ($P = 0.08$).

However, there were no significant differences between the 2 samples in relation to MO and SC.

Figure 3 shows a scatter plot of observed versus predicted choice probabilities for each of the 150 DCE pairs. When comparing the 2 samples, there are fewer observations in the middle range (where predicted and observed probabilities are about 0.5) for the adult sample when compared with the adolescent sample. This suggests that, in the adult sample, there were fewer cases in which the probability of respondents choosing each option was relatively equal as compared with the adolescent sample.

Approximately 90% of the adult sample chose the dominant option in the “fixed pair” relative to 88% in

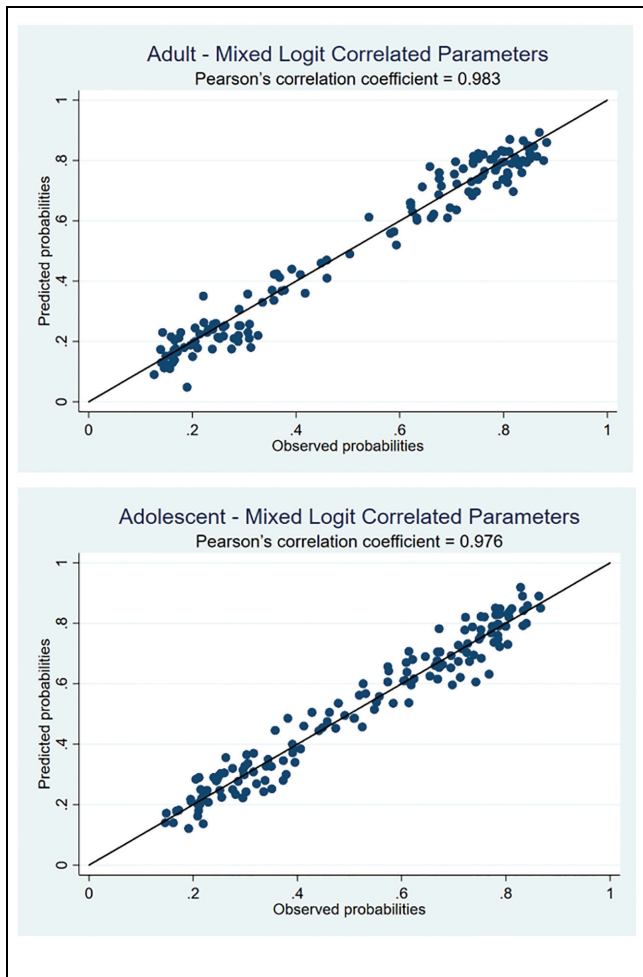


Figure 3 Observed versus predicted probabilities for mixed logit with correlated parameters, by sample.

the adolescent sample, indicating little difference between samples. The results did not differ when individuals who failed the test were excluded. Figure 4 illustrates that, when comparing the responses based on differences in LSS, the expected pattern was observed for both samples overall. Adolescents were generally slightly less likely to choose the “less severe” option in the tasks, and this was exacerbated when the difference in LSS was very small. For example, when option A had an LSS that was 1 point lower than that of option B (i.e., option A was “less severe” than option B), 76% of adults chose option A relative to 68% of adolescents. In addition, the proportion of the adolescent sample choosing each option when there was no difference in LSS was further from 50:50, relative to the adult sample.

A similar proportion of respondents agreed or strongly agreed that the tasks were difficult between the

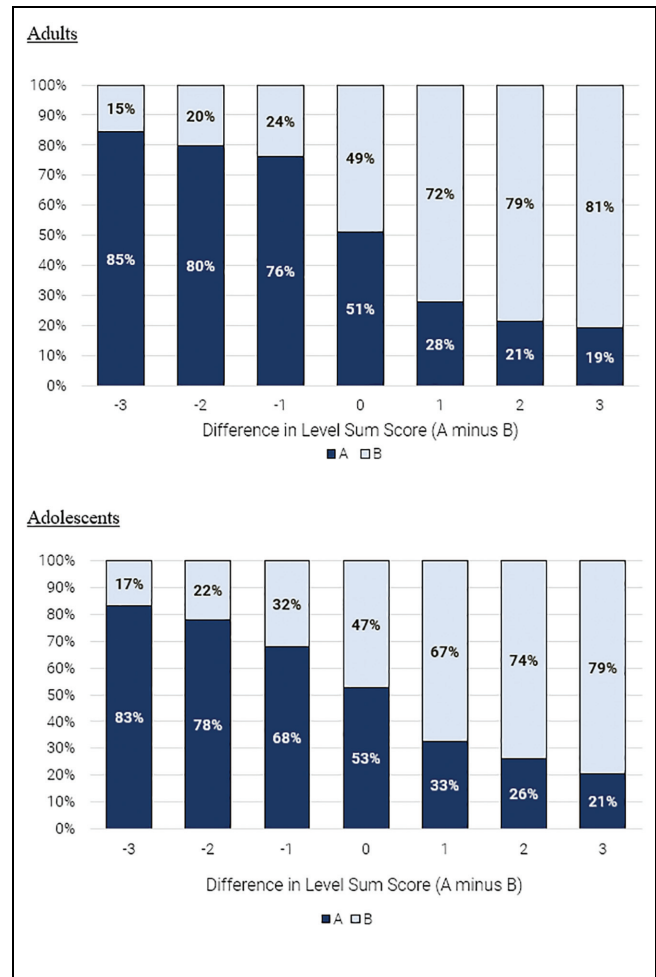


Figure 4 Proportion of respondents choosing A/B.

2 samples (27% adults; 26% adolescents). However, a higher proportion of adolescent respondents reported that they found it difficult to tell the difference between profiles (15% adults; 24% adolescents) and that they found it difficult to imagine the health problems described (28% adults; 44% adolescents).

Discussion

In this study, latent scale values for the EQ-5D-Y-3L instrument were obtained from a DCE that was completed by both a sample of adults and a sample of adolescents in the UK. The results indicate that preferences differ between the 2 samples. In particular, it appears that adolescents give less weight to PD, AD, and UA relative to those in the adult sample. This may be due to differences in experience; fewer adolescents are likely to

have suffered with these issues relative to adults. It could be the case that adults are more concerned with these issues on average, which could be due to having a greater level of experience with health issues. Our results align with those from other recent studies with respect to adolescents giving less weight to pain relative to adults.^{18,19} In contrast, our results differ in relation to mental health issues (i.e., in this study, adults gave more weight to AD than adolescents did).^{18,19} However, there are substantial differences between the studies in relation to the methodology and country.

In relation to feasibility and data quality, it is interesting to observe that there were fewer choice probabilities centered on 0.5 for adults relative to adolescents. This might imply that the adolescent sample on average were less consistent than the adult sample. For example, for any given choice, a larger majority of the adult sample may have chosen a particular alternative relative to the adolescent sample (i.e., a smaller majority). It was also the case that adolescents were typically less likely to choose the option with a lower LSS when the difference in LSS between the 2 options was small. These findings suggest that some adolescents may have struggled with the task to a greater extent than adults, despite an equal proportion of the 2 samples reporting that they did not find the tasks difficult. It could be the case that this is better explained by the relatively greater difficulty that adolescents had compared with adults when it came to differentiating between the health descriptions and imagining the health problems that were described. This result should certainly be the subject of future further research. Nonetheless, it is important to note that these are minor differences and that the proportion of adolescents passing the dominance test was almost identical to the proportion of adults passing. Furthermore, the dominance test pass rates were in line with rates observed in other DCE studies.³⁰ Overall, it would seem fair to conclude that the adolescent sample did not struggle significantly with the DCE and that this methodology would appear to be suitable for use in this age range.

Given that it appears to be feasible to elicit preferences from adolescents and that their preferences differ slightly from that of adults, a logical question follows relating to how this information could, or should, be used. It is important to note that an additional set of data would be required to “anchor” these DCE data onto the QALY scale.^{33,34} It is not necessarily the case that the anchoring task would be feasible in a sample of adolescents, especially if an approach such as time trade-off is used, as in the EQ-5D-Y-3L protocol.⁶ If it were feasible, it would theoretically be possible to generate an

EQ-5D-Y-3L value set based entirely on adolescent preferences, which could be used in evaluations of interventions aimed at younger populations. Another alternative might be to combine the DCE data to create a general population EQ-5D-Y-3L value set, which includes preference data from a smaller (representative) number of individuals aged between 11 and 17 y. In such a case, it might not be necessary for adolescents to be involved in the anchoring task, as their preferences would at least be captured within the DCE data.

This study has a number of limitations. First, the perspective of the task differed between the 2 samples. Adults were asked to express preferences with respect to another individual, whereas adolescents were asked for their own individual preferences. Various theoretical frameworks have highlighted the importance of differences in perspective when eliciting preferences in health.^{35–37} It is therefore important to highlight that both the respondent sample and the perspective of the task differed in our study, which reduces our ability to accurately determine why preferences between the 2 samples differ. Another limitation is that the adult sample was asked to think about a 10-y-old child experiencing the health states to be valued, without specifying who that child is. Our intention was to avoid specific ways of framing the questions that may have limited the generalizability of the preferences elicited. However, the risk with this approach is that we do not know the cognitive process employed by respondents in completing the tasks; for example, some may have considered themselves as a 10-y-old, considered a 10-y-old they know, or imagined a hypothetical 10-y-old. The approaches might differ across respondents and could have been different had the reference child been framed in a different manner. For example, the age of the reference child may have made a difference. Nobody in the adolescent sample was younger than 11 y, and therefore, there was no direct comparison between the 2 samples. This mismatch could potentially have been avoided had we asked adolescents to take the perspective of a 10-y-old child too. An alternative could have been to ask both samples to take the perspective of a 14-y-old child instead, as this was the midpoint of the ages covered by the adolescent sample. However, using the 10-y-old child perspective in adult samples is consistent with earlier research, which is why we opted for this perspective.^{4,5} Furthermore, as we were exploring the feasibility of providing a DCE to a sample of adolescents, it did not seem sensible to include a further complication by asking adolescent participants to consider the health of a hypothetical child, rather than themselves. Ultimately it is not possible to disentangle whether the framing of the questions for each sample may have

played a role in the differences that were observed, and only further research will be able to uncover the influence that framing may have. In addition, another limitation is that the DCE tasks did not include any consideration of dead or the duration of the health states, and the latent scale results reported here are therefore not anchored in a manner that would enable them to be used in the estimation of QALYs, without further information. This means that comparisons are limited to the relative importance of the different levels rather than comparisons of (anchored) utilities, which would be more meaningful. However, the inclusion of dead or duration may have made the task too difficult for the adolescent sample and may have raised ethical issues. Post hoc anchoring of latent scale values may enable a value set to be created at a later date. Finally, online data collection can be susceptible to data quality issues relative to other modes of administration, which may be exacerbated when recruiting adolescents. However, it has been noted that other modes of administration may also have such issues,³⁸ and in this study, the results showed good face validity and the logic checks generally indicated a good level of respondent understanding.


Conclusion


Our evidence suggests that adolescents' preferences differ from those of adults taking the perspective of a child. It may be that these differences exist because of the relative experience of adults, who might have a better understanding of ill health and its effect on HRQOL. However, a normative argument can be made that adolescents' preferences should be considered in decision making that is directly relevant to them. Although the cognitive demands of other valuation methods may have ruled this possibility out, this study provides evidence to suggest that adolescents are capable of completing a DCE. Future research should further explore the possible differences that may occur in value sets as a result of these latent scale differences.


Acknowledgments

We are grateful to Simone Kreimeier and 4 anonymous reviewers for feedback received on earlier drafts.

ORCID iDs

David J. Mott  <https://orcid.org/0000-0001-5959-8447>

Koonal K. Shah  <https://orcid.org/0000-0002-4927-7858>

Oliver Rivero-Arias  <https://orcid.org/0000-0003-2233-6544>

Supplemental Material

Supplementary material for this article is available on the *Medical Decision Making* website at <http://journals.sagepub.com/home/mdm>.

References

1. Wille N, Badia X, Bonsel G, et al. Development of the EQ-5D-Y: a child-friendly version of the EQ-5D. *Qual Life Res.* 2010;19(6):875–86.
2. Ravens-Sieberer U, Wille N, Badia X, et al. Feasibility, reliability, and validity of the EQ-5D-Y: results from a multinational study. *Qual Life Res.* 2010;19(6):887–97.
3. Rowen D, Rivero-Arias O, Devlin N, Ratcliffe J. Review of valuation methods of preference-based measures of health for economic evaluation in child and adolescent populations: where are we now and where are we going? *Pharmacoeconomics.* 2020;38(4):325–40.
4. Kind P, Klose K, Gusi N, Olivares PR, Greiner W. Can adult weights be used to value child health states? Testing the influence of perspective in valuing EQ-5D-Y. *Qual Life Res.* 2015;24(10):2519–39.
5. Kreimeier S, Oppe M, Ramos-Goñi JM, et al. Valuation of EuroQol Five-Dimensional Questionnaire, Youth Version (EQ-5D-Y) and EuroQol Five-Dimensional Questionnaire, Three-Level Version (EQ-5D-3L) health states: the impact of wording and perspective. *Value Health.* 2018; 21(11):1291–8.
6. Ramos-Goñi JM, Oppe M, Stolk E, et al. International valuation protocol for the EQ-5D-Y-3L. *Pharmacoeconomics.* 2020;38(7):653–63.
7. Brazier J, Akehurst R, Brennan A, et al. Should patients have a greater role in valuing health states? *Appl Health Econ Health Policy.* 2005;4(4):201–8.
8. Versteegh MM, Brouwer WBF. Patient and general public preferences for health states: a call to reconsider current guidelines. *Soc Sci Med.* 2016;165:66–74.
9. Gandjour A. Theoretical foundation of patient v. Population preferences in calculating QALYs. *Med Decis Making.* 2010;30(4):E57–63.
10. Stiggelbout AM, de Vogel-Voogt E. Health state utilities: a framework for studying the gap between the imagined and the real. *Value Health.* 2008;11(1):76–87.
11. Mott DJ, Najafzadeh M. Whose preferences should be elicited for use in health-care decision-making? A case study using anticoagulant therapy. *Expert Review of Pharmacoeconomics & Outcomes Research.* 2016;16(1):33–9.
12. Peeters Y, Stiggelbout AM. Health state valuations of patients and the general public analytically compared: a meta-analytical comparison of patient and population health state utilities. *Value Health.* 2010;13(2):306–9.
13. Brazier J, Rowen D, Karimi M, Peasgood T, Tsuchiya A, Ratcliffe J. Experience-based utility and own health state valuation for a health state classification system: why and how to do it. *Eur J Health Econ.* 2018;19(6):881–91.

14. Karimi M, Brazier J, Paisley S. Are preferences over health states informed? *Health Qual Life Outcomes*. 2017;15(1):105.
15. Crump RT, Beverung LM, Lau R, Sieracki R, Nicholson M. Reliability, validity, and feasibility of direct elicitation of children's preferences for health states: a systematic review. *Med Decis Making*. 2017;37(3):314–26.
16. Krucien N, Sicsic J, Ryan M. For better or worse? Investigating the validity of best–worst discrete choice experiments in health. *Health Econ*. 2019;28(4):572–86.
17. 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.
18. 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: adolescent versus adult values for the CHU-9D. *Health Econ*. 2016;25(4):486–96.
19. 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.
20. Chen G, Xu F, Huynh E, Wang Z, Stevens K, Ratcliffe J. Scoring the Child Health Utility 9D instrument: estimation of a Chinese child and adolescent-specific tariff. *Qual Life Res*. 2019;28(1):163–76.
21. Oppe M, Devlin NJ, van Hout B, Krabbe PFM, de Charro F. A program of methodological research to arrive at the new international EQ-5D-5L valuation protocol. *Value Health*. 2014;17(4):445–53.
22. Hensher D, Rose J, Greene W. *Applied Choice Analysis*. 2nd ed. Cambridge, UK: Cambridge University Press;
23. Hess S, Train K. Correlation and scale in mixed logit models. *J Choice Model*. 2017;23:1–8.
24. McFadden D, Train K. Mixed MNL models for discrete response. *J Appl Econ*. 2000;15(5):447–70.
25. Fiebig DG, Keane MP, Louviere J, Wasi N. The generalized multinomial logit model: accounting for scale and coefficient heterogeneity. *Market Sci*. 2010;29(3):393–421.
26. Zhou M, Thayer WM, Bridges JFP. Using latent class analysis to model preference heterogeneity in health: a systematic review. *Pharmacoeconomics*. 2018;36(2):175–87.
27. Vass CM, Wright S, Burton M, Payne K. Scale heterogeneity in healthcare discrete choice experiments: a primer. *Patient*. 2018;11(2):167–73.
28. Wright SJ, Vass CM, Sim G, Burton M, Fiebig DG, Payne K. Accounting for scale heterogeneity in healthcare-related discrete choice experiments when comparing stated preferences: a systematic review. *Patient*. 2018;11(5):475–88.
29. Gonzalez JM. A guide to measuring and interpreting attribute importance. *Patient*. 2019;12(3):287–95.
30. Tervonen T, Schmidt-Ott T, Marsh K, Bridges JFP, Quaipe M, Janssen E. Assessing rationality in discrete choice experiments in health: an investigation into the use of dominance tests. *Value Health*. 2018;21(10):1192–7.
31. Shah K, Mulhern B, Longworth L, Janssen MF. An empirical study of two alternative comparators for use in time trade-off studies. *Value Health*. 2016;19(1):53–9.
32. Devlin N, Parkin D, Janssen B. *Methods for Analysing and Reporting EQ-5D Data*. Cham (UK): Springer International; 2020. Available from: <http://link.springer.com/10.1007/978-3-030-47622-9>. Accessed August 22, 2020.
33. Webb EJD, O'Dwyer J, Meads D, Kind P, Wright P. Transforming discrete choice experiment latent scale values for EQ-5D-3L using the visual analogue scale. *Eur J Health Econ*. 2020;21(5):787–800.
34. Shah KK, Ramos-Goñi JM, Kreimeier S, Devlin NJ. An exploration of methods for obtaining 0 = dead anchors for latent scale EQ-5D-Y values. *Eur J Health Econ*. 2020;21(7):1091–103.
35. Dolan P, Olsen JA, Menzel P, Richardson J. An inquiry into the different perspectives that can be used when eliciting preferences in health. *Health Econ*. 2003;12(7):545–51.
36. Cubi-Molla P, Shah K, Burström K. Experience-based values: a framework for classifying different types of experience in health valuation research. *Patient*. 2018;11(3):253–70.
37. Tsuchiya A, Watson V. Re-thinking 'the different perspectives that can be used when eliciting preferences in health.' *Health Econ*. 2017;26(12):e103–7.
38. Ratcliffe J, Flynn T, Terlich F, Stevens K, Brazier J, Sawyer M. Developing adolescent-specific health state values for economic evaluation: an application of profile case best-worst scaling to the Child Health Utility 9D. *Pharmacoeconomics*. 2012;30(8):713–27.