ORIGINAL RESEARCH ARTICLE

Mapping EQ-5D Utility Scores from the PedsQL™ Generic Core Scales

Kamran A. Khan · Stavros Petrou ·
Oliver Rivero-Arias · Stephen J. Walters ·
Spencer E. Boyle

Abstract

Purpose The Pediatric Quality of Life Inventory™ (PedsQL™) General Core Scales (GCS) were designed to provide a modular approach to measuring health-related quality of life in healthy children, as well as those with acute and chronic health conditions, across the broadest, empirically feasible, age groups (2–18 years). Currently, it is not possible to estimate health utilities based on the PedsQL™ GCS, either directly or indirectly. This paper assesses different mapping methods for estimating EQ-5D health utilities from PedsQL™ GCS responses.

K. A. Khan (✉) · S. Petrou
Warwick Clinical Trials Unit, Division of Health Sciences,
Warwick Medical School, University of Warwick,
Coventry CV4 7AL, UK
e-mail: k.a.khan@warwick.ac.uk

O. Rivero-Arias
National Perinatal Epidemiology Unit, Nuffield Department of
Population Health, University of Oxford, Oxford, UK

O. Rivero-Arias
Red de Investigación de Servicios Sanitarios en Cronicidad
(REDISSEC), Madrid, Spain

S. J. Walters
School of Health and Related Research, University of Sheffield,
Yorkshire, UK

S. E. Boyle
Faculty of Health and Life Sciences, Northumbria University,
Newcastle upon Tyne, UK

Methods This study is based on data from a cross-sectional survey conducted in four secondary schools in England amongst children aged 11–15 years. We estimate models using both direct and response mapping approaches to predict EQ-5D health utilities and responses. The mean squared error (MSE) and mean absolute error (MAE) were used to assess the predictive accuracy of the models. The models were internally validated on an estimation dataset that included complete PedsQL™ GCS and EQ-5D scores for 559 respondents. Validation was also performed making use of separate data for 337 respondents.

Results Ordinary least squares (OLS) models that used the PedsQL™ GCS subscale scores, their squared terms and interactions (with and without age and gender) to predict EQ-5D health utilities had the best prediction accuracy. In the external validation sample, the OLS model with age and gender had a MSE (MAE) of 0.036 (0.115) compared with a MSE (MAE) of 0.036 (0.114) for the OLS model without age and gender. However, both models generated higher prediction errors for children in poorer health states (EQ-5D utility score <0.6). The response mapping models encountered some estimation problems because of insufficient data for some of the response levels.

Conclusion Our mapping algorithms provide an empirical basis for estimating health utilities in childhood when EQ-5D data are not available; they can be used to inform future economic evaluations of paediatric interventions. They are likely to be robust for populations comparable to our own (children aged 11–15 years in attendance at secondary school). The performance of these algorithms in childhood populations, which differ according to age or clinical characteristics to our own, remains to be evaluated.

Key points for decision makers

The Pediatric Quality of Life Inventory[TM] (PedsQL[TM]) General Core Scales (GCS) were designed to provide a modular approach to measuring health-related quality of life in healthy children, as well as those with acute and chronic health conditions, across the broadest, empirically feasible age groups (2–18 years).

It is not currently possible to estimate health utilities based on the PedsQL[TM] GCS. More broadly, a number of methodological constraints limit the estimation of health utilities across the childhood spectrum.

Our study uses a number of direct and response mapping approaches to predict EQ-5D health utilities from PedsQL[TM] GCS responses.

The results of this study can be used to inform utility estimation within future economic evaluations of paediatric interventions.

1 Introduction

The measurement of health-related quality of life (HRQoL) has become an integral component of studies aimed at measuring the health benefits of healthcare interventions, and there are a large number of specific and generic instruments available for the task [1].

Generic HRQoL measures may either be non-preference based or preference based. Preference-based HRQoL measures generate health state preference values or utilities that can be used in the calculation of quality-adjusted life years (QALYs) [1]. The QALY is a composite measure that combines the length and HRQoL following a healthcare intervention [2], where 'quality' of life is commonly measured in terms of health state utility values estimated using a preference-based technique or measure of health. Health state utility values, and by extension QALYs, can also be useful for assessing cost effectiveness as they translate outcomes for different interventions across disparate conditions into a common metric. Many decision makers, such as the National Institute for Health and Care Excellence (NICE) in England and Wales, recommend the use of the QALY as a standard measure of benefit for

economic evaluation purposes and the EQ-5D as the integral preference-based measure of HRQoL [3].

Despite the diffusion of preference-based HRQoL measures, a number of methodological concerns have constrained their development in children and, in particular, young children. Methodological concerns that are specific to this age group include issues surrounding the relevant attributes to incorporate into measurement instruments, appropriate respondents for measurement exercises (for example, children, parents or other proxies), potential sources of bias in the description and valuation processes, and the psychometric properties of existing measures [4–6]. Faced with these methodological constraints, several analysts conducting economic evaluations within the paediatric context have applied health state utility values derived for adult populations to childhood health states [7]. This ignores concerns by child development and other specialists that rapid developmental change throughout childhood makes it difficult to identify a common set of health attributes relevant to all age groups, and by extension a common set of utility values for health states applicable across all age groups [8]. In the light of these methodological concerns, there is clearly a pressing need for new approaches to measuring preference-based HRQoL outcomes in children and, in particular, young children. This has been recognised by the UK Medical Research Council-NICE scoping project for identifying methodological research priorities in the health sciences [9]. One solution to this problem is to derive a preference-based HRQoL measure de novo, such as the Child Health Utility 9D (CHU9D), a new measure developed for children aged 7–11 years [10]. Another solution is to apply a mapping (or 'crosswalk') function to convert non-preference-based HRQoL data into one of the generic preference-based measures where relevant data are available. 'Mapping' involves the development and use of an algorithm (or algorithms) to predict health-state utility values using data on other indicators or measures of health [11]. Mapping has proved a popular solution as it enables health state utility values to be predicted when no preference-based measure has been included in a study [11], but has not to our knowledge been used to predict health state utility values that are potentially applicable across the childhood spectrum.

This paper reports the results of a regression-based exercise that maps responses to the Pediatric Quality of Life Inventory[TM] 4.0 Generic Core Scales (hereafter PedsQL[TM] GCS for brevity), a generic non-preference-based HRQoL measure widely used in childhood, onto the EuroQol EQ-5D. To our knowledge, this is the first study that estimates conversion algorithms between the two measures.

2 Methods

2.1 Data

The data were obtained from a study that investigated the relationship between HRQoL, physical activity, diet and overweight status in children aged 11–15 years. A cross-sectional survey of four secondary schools in England was carried out using the PedsQL™ GCS, the youth version of the EQ-5D (EQ-5D-Y), the self-completed Western Australian Child and Adolescent Physical Activity and Nutrition Survey (CAPANS) and a food intake screener questionnaire. Self-reported demographic information including age, sex and ethnicity was also collected from the children that completed the survey. The schools were selected on the basis of a close match in examination results, percentage of children on free school meals and percentage of children with special educational needs. 2,858 children were asked to participate in an anonymous survey on two occasions, once in winter and again in summer. There were 869 respondents to the winter survey and 1,000 respondents to the summer survey and so the full dataset comprised 1,869 sets of responses. The study is described in detail elsewhere [12]. Because the data were collected anonymously, it is not possible to identify how many children completed the survey twice, once in the winter and then again in summer. This led us to focus on data collected at one time point for the purpose of the mapping exercises reported here. It was decided to use the 1,000 respondents to the summer survey for the modelling reported here as this constituted the larger sample, and to split this sample by geographical area to provide the estimation (children from two schools in north-west England) and validation (children from two schools in south-west England) samples.

2.2 Outcome Measures

The EQ-5D [13] is the most widely used generic preference-based measure of health outcomes [14]. It consists of two principal measurement components, a descriptive system and a visual analogue scale. This study concentrates on mapping using the information provided by the descriptive system, which is comprised of five dimensions of health: mobility, self-care, usual activities, pain/discomfort and anxiety/depression. Each dimension is assessed by a single question on a three-point ordinal scale (no problems, some or moderate problems, severe or extreme problems). This generates a total number of 243 (3^5) possible EQ-5D health states. In the UK, utility valuations for the 243 EQ-5D health states are commonly based on time trade-off valuations by 3,337 members of the general public [15]. Alternative tariffs have been developed for other countries [16], but in this study the UK (York A1) tariff set was applied. This study used the youth version of the EQ-5D, the EQ-5D-Y, which has been especially adapted in terms of language for children aged 8–11 years and for adolescents aged 12–18 years [17, 18]. Utility valuations in the York A1 tariff set range from no problems on any of the five dimensions in the EQ-5D descriptive system (value = 1.0) to severe or extreme impairment on all five dimensions (value = −0.594).

The PedsQL™ is an instrument that was designed to measure HRQoL in healthy children, as well as those with acute and chronic health conditions, covering an age spectrum in the range of 2–18 years [19]. The PedsQL™ GCS comprise parallel child self-report [age 5–7 years (young child), 8–12 years (child) and 13–18 years (adolescent)] and parent proxy-report [age 2–4 years (toddler), 5–7 years (young child), 8–12 years (child) and 13–18 years (adolescent)] formats. The PedsQL™ GCS were specifically designed to measure the core dimensions of health as delineated by the World Health Organisation, as well as role (school/day care) functioning. It contains 23 items that are grouped to create four sub-scales of (1) Physical Functioning (PF: 8 items), (2) Emotional Functioning (EF: 5 items), (3) Social Functioning (SF: 5 items) and (4) School Functioning (Sch F: 5 items). A 5-point response scale is applied across each item (0 = never a problem; 1 = almost never a problem; 2 = sometimes a problem; 3 = often a problem; 4 = almost always a problem). Items are reverse scored and linearly transformed to a 0–100 scale (0 = 100, 1 = 75, 2 = 50, 3 = 25, 4 = 0) with higher scores indicating better HRQoL. For scale and total scores, the mean is computed as the sum across all items divided by the number of items answered, thereby accounting for missing data if present [19].

2.3 Statistical Methods

To estimate EQ-5D-Y utility scores based on the PedsQL™ GCS, we employed two conversion techniques, namely direct and response mapping. We first estimated direct mapping models by regressing the PedsQL™ GCS scores directly onto the EQ-5D-Y utility scores. The direct mapping approach makes use of regression equations to predict the values of one outcome measure using scores/values from a second measure as regressors. The coefficients of the models can then be used to carry out the conversion from the source measure to the target measure in the required dataset. The following functional forms were used for the direct mapping:

- Ordinary least squares (OLS) regression with any utilities predicted to be >1 set to one.

- Generalized linear modelling (GLM) [20], which accommodates for skewness via variance weighting rather than through transformation and re-transformation. More specifically, such models explicitly specify a distribution that reflects the relationship between the mean and the variance, and a link function between the linear part $x'\beta$ and the mean $\mu = E(y/x')$ on the original scale of the dependent variable. The modified Parks test [21] was used to identify the preferred distributional family based on the lowest χ^2 value. For each prediction model, tests were carried out to identify the link function (including identity, square root, and log) using the Pearson correlation test [22], the Pregibon link test [23] and the modified Hosmer–Lemeshow test [24]. Where all three tests yield non-significant p values, the link function is said to fit well [25].

- Two-part logit-OLS regression was also used to deal with the high proportion of respondents having EQ-5D-Y utility scores of one (representing full health). Because OLS would not always predict a discrete score of one, a two-part model was formulated to be able to predict full-health states. The first part comprised a logistic regression model estimated on the entire estimation sample to predict which participants had perfect health, while the second part comprised an OLS model predicting EQ-5D-Y utility scores for those participants with utility scores <1. Utility predictions for the two-part model were estimated using an expected value method as follows: $\Pr(U = 1) + (1 - \Pr(U = 1)) \times U$, with $\Pr(U = 1)$ indicating the predicted probability of being in full health obtained from the first part of the model, and U the predicted utility conditional on imperfect health estimated from the second part of the model.

- Additional estimators, including Censored Least Absolute Deviations (CLAD) [26, 27] and Tobit [28], have been used in the broader literature for direct mapping purposes [29], and were also applied in this study for comparative purposes despite theoretical concerns and evidence to suggest that CLAD and Tobit estimators generate biased results when the true utility is conceptually bounded above at 1.0 [30, 31]. Another potential model, the Fractional Logistic regression (FLOGIT) estimator [32, 33] has not been extensively used in the modelling utility data literature and a recent application did not generate promising results [34].

We then used response mapping to predict the responses to the EQ-5D-Y dimensions as opposed to predicting the summary utility score directly [35]. A logistic regression model can be used to estimate the probabilities that each set of PedsQL^{TM} GCS responses would correspond to a response level for each EQ-5D-Y dimension. A multinomial logistic model can be used or an ordered logistic model if it is believed that the responses to the EQ-5D-Y questions are ordered. The models were estimated by fitting a separate multinomial logistic regression (MLOGIT) or ordinal logistic regression (OLOGIT) model for each EQ-5D-Y dimension, as described elsewhere [35]. Predictions for the response mapping model were generated using the expected value method, which is equivalent to the Monte Carlo approach given a large number of repeated Monte Carlo draws [36].

For each of the functional forms, eight models were estimated that differed in their predictors (or independent variables): The first model predicted the EQ-5D-Y utility score from the PedsQL^{TM} total scale score (model 1); the second model had the same specification as the first model but also included age and gender (model 2); the third model used the PedsQL^{TM} sub-scale scores (model 3); the fourth model had the same specification as the third model but also included age and gender (model 4); the fifth model used the PedsQL^{TM} sub-scale scores, their squared terms and interaction terms to formulate predictions (model 5); the sixth model had the same specification as the fifth model but also included age and gender (model 6); in the seventh model, the explanatory variables were 92 dummy variables (23 questions each with five possible responses, 'never' used as reference) indicating whether or not participants had a particular response level on each PedsQL^{TM} GCS question (model 7); and the eighth model had the same specification as the seventh model but also included age and gender (model 8). All were estimated in Stata version 11 (Stata-Corp, College Station, TX, USA).

2.4 Assessing Model Performance

In line with external guidance [11, 29], the mean squared error (MSE) and the mean absolute error (MAE) were used to measure the goodness of fit of the models. The MSE equals the mean of squared differences between the observed EQ-5D-Y utility score and the EQ-5D-Y utility score predicted from the model, whilst the MAE is the mean of absolute differences between observed and predicted EQ-5D-Y utility scores.

Models were initially selected on the basis of the MSE in the validation sample. To compare the models further, a number of analyses were carried out using the results from the validation sample. First, the distributions of the predicted and observed EQ-5D-Y utility scores were examined to see how closely the predicted values matched the observed scores [37], and the proportions of predictions deviating from observed values by <0.10 or <0.25 were also calculated because these indicate the distribution of errors and how often the models fail to produce useful

predictions [1]. This was followed by plotting the histograms of the residuals (observed minus predicted EQ-5D-Y utility scores) to ascertain the bias in the predicted EQ-5D-Y utility scores. Finally, the errors were reported across subsets of the EQ-5D-Y utility score range, as this is useful for indicating whether or not there is systematic bias in the predictions [11]. The best fitting model(s) were then re-estimated using data for the entire summer sample.

3 Results

3.1 Study Sample Characteristics

There were a total of 1,000 respondents to the summer survey and of these 896 had complete PedsQL™ GCS and EQ-5D-Y data, from here referred to as the whole sample. Table 1 presents descriptive statistics for the whole, estimation and validation samples. Pairs of complete PedsQL™ GCS and EQ-5D responses were needed to estimate the mapping models, therefore the estimation sample included 559 children whereas the validation sample comprised 337 children. Just over one half (54.0 %) of the respondents in the whole sample were boys, the average age of the respondents was 13.3 years (standard deviation (SD) 1.3) and approximately 40 % of the respondents were non-white. There were no differences between the respondents who completed the HRQoL assessments and those who declined to complete the HRQoL assessments by body mass index (BMI), fruit intake and receipt of free school meals. However, those who did not complete the HRQoL assessments were more likely to be girls, of a younger age, and of white ethnicity. There were significant differences in the characteristics of the estimation and validation samples across all variables except for whether or not the child had achieved the target level of physical activity per day.

Table 2 shows the summary characteristics of the outcome measures across the whole, estimation and validation samples. The proportion of responses to each level of the five EQ-5D-Y dimensions was similar across the samples for all but the anxiety/depression dimension. The mean EQ-5D-Y utility score and total PedsQL™ GCS score for the whole sample were 0.89 (SD 0.21) and 85.4 (SD 14.3), respectively. The mean total PedsQL™ GCS scores differed between the estimation and validation samples as did the physical functioning, emotional functioning and school functioning sub-scale scores. Figure 1 shows the distribution of the outcome measures across the data sets. Both the EQ-5D-Y utility scores and PedsQL™ GCS total scores are negatively skewed with a large mass of values at full health, and the EQ-5D-Y utility scores are bi-modally distributed.

When using the GLM estimator, the Poisson family with log link was identified as appropriate for modelling the data. For the response mapping approach, when using the mlogit we found that that four of the eight model specifications did not converge. A requirement for using an ordered logistic model is that the parallel regression assumption holds. A likelihood ratio test was used to assess whether this assumption held, but it did not in all but one model. However, the results obtained using these models have been presented for information purposes only to show how the results from using the response mapping approach compare with those using the direct utility mapping approach. There were convergence issues with several of the CLAD models and therefore no results have been presented in Table 3 for the CLAD estimator.

3.2 Validation

The performance of all of the models using different groups of independent variables was assessed using the estimation sample and, separately, the validation sample. Table 3 shows summary performance indicators for each model using both the estimation and validation samples. The results show that most of the models were able to fairly accurately predict the mean EQ-5D-Y utility score in the validation sample (0.88) with predicted mean EQ-5D-Y utility scores in the range of 0.87–0.90. A number of models were able to predict utility scores into the negative range; the models predicting the largest negative value closest to the actual one for the validation sample were the GLM 1 (−0.72), two-part 7 (−0.43) and two-part 8 (−0.35) models in the external validation sample. Table 3 also shows that some of the OLS and two-part models over predicted the highest EQ-5D-Y utility score. The models were ranked according to their MSE, the primary measure of prediction accuracy, in the validation sample to shortlist the two best performing models. Based on the MSE, the two models that gave the best predictions were the OLS 6 (MSE 0.0363) and OLS 5 (MSE 0.0364) models. However, the ranking of the two best models was not the same by the MAE with the models ordered as OLS 5 (MAE 0.1140) and OLS 6 (MAE 0.1151).

3.3 Performance of Best Models

When considering the use of a mapping algorithm, its ability to accurately predict the mean EQ-5D-Y utility score is important but should not be the only consideration. For the two best models, in addition to assessing how accurately the models estimated the mean EQ-5D-Y utility score, we also examined the distributions of the predicted scores. Table 4 contains results from analyses carried out using either the estimation sample or the validation sample,

Table 1 Descriptive statistics for the whole, estimation and validation samples

Characteristic	Whole sample n = 896 n (%)	Estimation sample n = 559 n (%)	Validation sample n = 337 n (%)	p Value[†]
Gender				
Male	484 (54.02)	347 (62.08)	137 (40.65)	<0.001
Female	412 (45.98)	212 (37.92)	200 (59.35)	
Age (years)				
11	76 (8.48)	47 (8.41)	29 (8.61)	<0.001
12	246 (27.46)	176 (31.48)	70 (20.77)	
13	81 (9.04)	53 (9.48)	28 (8.31)	
14	328 (36.61)	202 (36.14)	126 (37.39)	
15	165 (18.42)	81 (14.49)	84 (24.93)	
Ethnicity				
White	541 (60.38)	217 (38.82)	324 (96.14)	<0.001
Non-white	355 (39.62)	342 (61.18)	13 (3.86)	
Free meals				
Yes	738 (82.37)	438 (78.35)	300 (89.02)	<0.001
No	157 (17.52)	121 (21.65)	36 (10.68)	
Missing	1 (0.11)		1 (0.30)	
Weight				
Under weight	59 (6.58)	45 (8.05)	14 (4.15)	0.038
Normal	296 (33.04)	185 (33.09)	111 (32.94)	
Over weight	34 (3.79)	16 (2.86)	18 (5.34)	
Very over-weight	22 (2.46)	15 (2.68)	7 (2.08)	
Missing	485 (54.13)	298 (53.31)	187 (55.49)	
Achieving 60 min of moderate/vigorous physical activity per day				
No	620 (69.20)	380 (67.98)	240 (71.22)	0.309
Yes	276 (30.80)	179 (32.02)	97 (28.78)	
Optimal fruit/vegetable intake				
No	386 (43.08)	274 (49.02)	112 (33.23)	<0.001
Yes	510 (56.92)	285 (50.98)	225 (66.77)	

Underweight, below 2nd percentile; normal, 2nd percentile up to the 91st percentile; overweight, 91st percentile up to 98th percentile; very overweight, 98th percentile or higher

[†] Comparisons of estimation and validation samples carried out using the Student's t tests for continuous variables and χ^2 test for categorical variables

but we only focus on the results for the validation sample in the narrative. Table 4 shows that, on average, the OLS 6 model gave the widest range of predicted scores and the OLS 5 model gave the narrowest range of predictions. Both models (OLS 6 and OLS 5) were able to predict negative values for the EQ-5D-Y utility score, but only came within 0.465 and 0.482 of the lowest observed EQ-5D-Y utility score, respectively. Both models reported lower variability across predicted utility values with SDs at 51 and 50 % of the magnitude of the observed EQ-5D-Y scores, on average, for the OLS 6 and OLS 5 models, respectively.

If we compare the 25th percentile, median and 75th percentile predicted scores to the respective observed EQ-5D-Y utility scores, the OLS 6 model has the closest

matching score for the 25th percentile, and the OLS 5 model has the closest matching score for the median and 75th percentile. Both models achieved around 65 and 94 % of individual estimations within 0.10 and 0.25 of the observed values, respectively. The plots of residuals for the two best performing models are shown in Fig. 2 and they are very similar for both preferred models.

Table 5 shows the MSE and MAE across the range of EQ-5D-Y utility scores for the two best performing models. Although the prediction accuracy for the mean EQ-5D-Y utility scores was similar across models, the level of accuracy was not uniform across the full range of scores. If we first look at the MSE for the models, both models were better predictors at the upper end of the EQ-5D-Y range.

**Table 2** Summary characteristics of outcome measures across the samples

Characteristic	Whole sample $n = 896$	Estimation sample $n = 559$	Validation sample $n = 337$	p value[†]
EQ-5D domain				
Mobility, n (%)				0.589
Level 1	849 (94.75)	533 (95.35)	316 (93.77)	
Level 2	36 (4.02)	20 (3.58)	16 (4.75)	
Level 3	11 (1.23)	6 (1.07)	5 (1.48)	
Self-care, n (%)				0.476
Level 1	872 (97.32)	546 (97.67)	326 (96.74)	
Level 2	13 (1.45)	6 (1.07)	7 (2.08)	
Level 3	11 (1.23)	7 (1.25)	4 (1.19)	
Usual activities, n (%)				0.124
Level 1	838 (93.53)	530 (94.81)	308 (91.39)	
Level 2	47 (5.25)	23 (4.11)	24 (7.12)	
Level 3	11 (1.23)	6 (1.07)	5 (1.48)	
Pain/discomfort, n (%)				0.359
Level 1	686 (76.56)	423 (75.67)	263 (78.04)	
Level 2	196 (21.88)	129 (23.08)	67 (19.88)	
Level 3	14 (1.56)	7 (1.25)	7 (2.08)	
Anxiety/depression, n (%)				<0.001
Level 1	695 (77.5)	464 (83.01)	231 (68.55)	
Level 2	177 (19.75)	81 (14.49)	96 (28.49)	
Level 3	24 (2.68)	14 (2.50)	10 (2.97)	
EQ-5D VAS score, mean (SD) (range), (IQR)	78.09 (17.12) (10, 100), 20	77.53 (17.49) (10, 100), 22	79.07 (16.43) (25, 100), 20	0.0961
EQ-5D utility score, mean (SD) range, (IQR)	0.89 (0.21) (−0.59, 1.00), 0.19	0.90 (0.20) (−0.59, 1.00), 0.20	0.88 (0.23) (−0.59, 1.00), 0.15	0.0693
PedsQL score, mean (SD) range, (IQR)				
Total scale score	85.39 (14.25) (0.00, 100.00), 18.75	86.26 (13.94) (0.00, 100.00), 17.03	83.95 (14.64) (10.00, 100.00), 22.50	0.0182
Physical functioning	90.39 (14.20) (0.00, 100.00), 15.63	90.57 (13.91) (0.00, 100.00), 15.63	90.09 (14.68) (0.00, 100.00), 15.63	0.6188
Emotional functioning	82.71 (19.65) (0.00, 100.00), 30.00	84.01 (18.99) (0.00, 100.00), 25.00	80.56 (20.55) (0.00, 100.00), 30.00	0.0110
Social functioning	88.53 (17.66) (0.00, 100.00), 20.00	89.46 (17.86) (0.00, 100.00), 15.00	86.97 (17.22) (0.00, 100.00), 25.00	0.0408
School functioning	79.94 (19.80) (0.00, 100.00), 35.00	81.01 (19.22) (0.00, 100.00), 30.00	78.16 (20.64) (0.00, 100.00), 40.00	0.0368

[†] Comparisons of estimation and validation samples carried out using the Student's t tests for continuous variables and χ^2 test for categorical variables

IQR interquartile range, *SD* standard deviation, *VAS* visual analogue scale

For EQ-5D-Y utility scores above 0.6, the models had MSEs between 0.010 and 0.022 whereas for predicted values below 0.6 the MSE was in the range of 0.117–1.03. The OLS 6 model reported the lowest MSE in more EQ-5D-Y utility range categories than the OLS 5 model, but the OLS 5 model reported the smallest errors for the 0.8–1 category of the EQ-5D-Y utility score range. The results for the MAEs followed a similar pattern to those of the MSE with models predicting more accurately at the upper end of the EQ-5D-Y utility score range. This result can also be seen in the quantile-quantile plot in supplementary Fig. 1, which shows the over predictions for the lower end of the EQ-5D-Y range. Model coefficients for the two best-performing models are presented in Table 6.
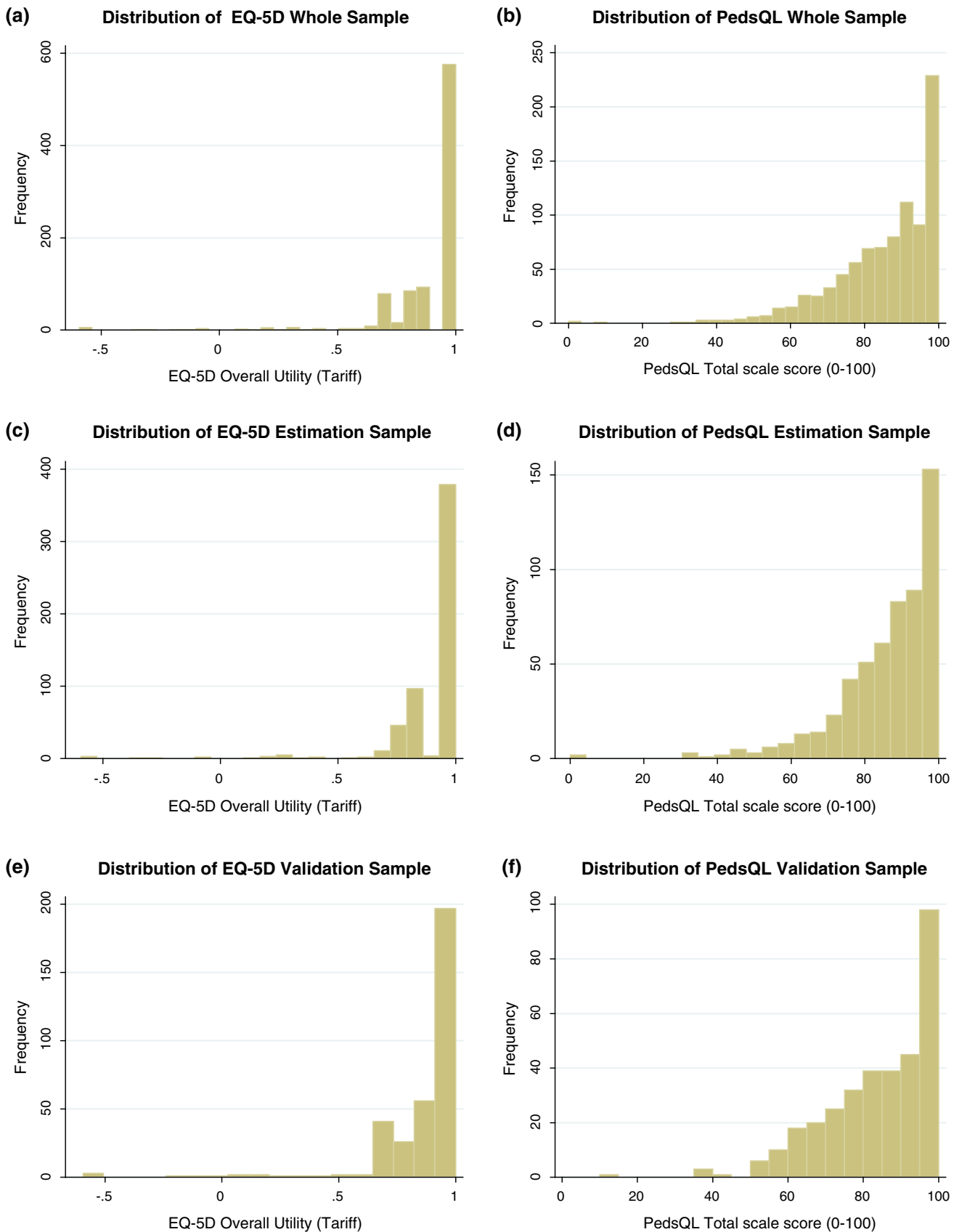
Fig. 1 Distribution of outcome measures across the data sets

Table 3 Model performance

Estimation method	Estimation sample (n = 559)					Validation sample (n = 337)				
	Mean (SD)	Min	Max	MSE	MAE	Mean (SD)	Min	Max	MSE	MAE
Observed EQ-5D	0.90 (0.20)	−0.59	1.00	–	–	0.88 (0.23)	−0.59	1.00	–	–
OLS (1)	0.90 (0.09)	0.34	0.99	0.0329	0.1000	0.89 (0.10)	0.40	0.99	0.0409	0.1149
OLS (2)	0.90 (0.09)	0.33	1.01	0.0327	0.0994	0.88 (0.10)	0.40	1.01	0.0409	0.1166
OLS (3)	0.90 (0.10)	0.28	0.99	0.0320	0.0968	0.89 (0.10)	0.30	0.99	0.0389	0.1116
OLS (4)	0.90 (0.10)	0.27	1.01	0.0318	0.0967	0.89 (0.10)	0.28	1.01	0.0388	0.1134
OLS (5)	0.90 (0.11)	−0.30	1.05	0.0291	0.1006	0.89 (0.10)	−0.11	1.07	0.0364	0.1140
OLS (6)	0.90 (0.11)	−0.31	1.05	0.0290	0.1000	0.89 (0.10)	−0.13	1.06	0.0363	0.1151
OLS (7)	0.90 (0.13)	−0.27	1.15	0.0232	0.0870	0.89 (0.15)	−0.19	1.18	0.0372	0.1120
OLS (8)	0.90 (0.13)	−0.26	1.15	0.0230	0.0868	0.89 (0.15)	−0.20	1.18	0.0371	0.1126
Tobit (1)	0.90 (0.11)	−0.14	0.97	0.0309	0.1020	0.89 (0.11)	0.02	0.97	0.0380	0.1153
Tobit (2)	0.90 (0.11)	−0.13	0.98	0.0308	0.1003	0.88 (0.11)	0.02	0.98	0.0384	0.1179
Tobit (3)	0.90 (0.12)	−0.17	0.97	0.0302	0.0982	0.89 (0.12)	−0.10	0.97	0.0368	0.1128
Tobit (4)	0.90 (0.12)	−0.20	0.98	0.0300	0.0977	0.89 (0.12)	−0.13	0.98	0.0372	0.1154
Tobit (5)	0.90 (0.12)	−0.21	1.00	0.0292	0.0960	0.89 (0.11)	0.02	1.00	0.0371	0.1134
Tobit (6)	0.90 (0.12)	−0.24	1.00	0.0292	0.0955	0.88 (0.12)	−0.03	0.99	0.0373	0.1156
Tobit (7)	0.90 (0.15)	−0.34	1.00	0.0240	0.0820	0.88 (0.21)	−0.50	1.00	0.0462	0.1235
Tobit (8)	0.90 (0.15)	−0.34	1.00	0.0240	0.0818	0.87 (0.21)	−0.46	1.00	0.0472	0.1263
2 Part (1)	0.90 (0.10)	0.35	0.98	0.0327	0.1022	0.89 (0.10)	0.40	0.98	0.0401	0.1169
2 Part (2)	0.90 (0.10)	0.32	0.98	0.0328	0.1014	0.88 (0.10)	0.36	0.98	0.0403	0.1183
2 Part (3)	0.90 (0.10)	0.29	0.98	0.0315	0.0983	0.89 (0.11)	0.27	0.98	0.0378	0.1125
2 Part (4)	0.90 (0.10)	0.24	0.98	0.0316	0.0982	0.89 (0.11)	0.21	0.98	0.0382	0.1150
2 Part (5)	0.90 (0.11)	−0.33	1.06	0.0293	0.0996	0.90 (0.10)	−0.04	1.00	0.0377	0.1150
2 Part (6)	0.90 (0.11)	−0.36	1.07	0.0294	0.0996	0.89 (0.10)	−0.07	1.00	0.0378	0.1162
2 Part (7)	0.92 (0.10)	0.33	1.22	0.0234	0.0812	0.90 (0.16)	−0.43	1.21	0.0443	0.1231
2 Part (8)	0.92 (0.10)	0.32	1.20	0.0232	0.0804	0.90 (0.15)	−0.35	1.26	0.0433	0.1204
GLM (1)	0.90 (0.17)	−1.58	0.96	0.0374	0.1091	0.89 (0.12)	−0.72	0.96	0.0373	0.1151
GLM (2)	0.89 (0.21)	−2.80	0.97	0.0490	0.1121	0.88 (0.16)	−1.49	0.97	0.0405	0.1215
GLM (3)	0.90 (0.18)	−1.68	0.96	0.0383	0.1071	0.89 (0.16)	−1.43	0.96	0.0401	0.1170
GLM (4)	0.89 (0.24)	−3.42	0.97	0.0572	0.1108	0.87 (0.25)	−2.91	0.97	0.0582	0.1267
GLM (5)	0.90 (0.18)	−1.55	1.00	0.0379	0.1068	0.89 (0.09)	0.35	1.00	0.0414	0.1205
GLM (6)	0.89 (0.18)	−2.00	1.00	0.0409	0.1085	0.88 (0.11)	0.12	1.00	0.0411	0.1248
GLM (7)	Convergence not achieved									
GLM (8)	Convergence not achieved									
MLOGIT (1)	0.89 (0.10)	0.01	0.95	0.0311	0.1092	0.88 (0.10)	0.10	0.95	0.0382	0.1207
MLOGIT (2)	0.89 (0.11)	−0.07	0.97	0.0316	0.1095	0.87 (0.10)	0.01	0.97	0.0383	0.1222
MLOGIT (3)	0.89 (0.11)	−0.02	0.96	0.0306	0.1056	0.88 (0.11)	−0.12	0.96	0.0368	0.1169
MLOGIT (4)	0.89 (0.11)	−0.12	0.97	0.0309	0.1062	0.87 (0.12)	−0.23	0.97	0.0372	0.1193
MLOGIT (5)	Convergence not achieved									
MLOGIT (6)	Convergence not achieved									
MLOGIT (7)	Convergence not achieved									
MLOGIT (8)	Convergence not achieved									
OLOGIT (1)	0.89 (0.11)	−0.16	0.95	0.0310	0.1089	0.88 (0.10)	−0.01	0.95	0.0379	0.1202
OLOGIT (2)	0.89 (0.11)	−0.17	0.97	0.0310	0.1075	0.87 (0.11)	−0.03	0.97	0.0385	0.1235
OLOGIT (3)[a]	0.89 (0.11)	−0.20	0.96	0.0302	0.1051	0.88 (0.12)	−0.17	0.96	0.0370	0.1170
OLOGIT (4)	0.89 (0.11)	−0.23	0.97	0.0300	0.1046	0.87 (0.12)	−0.20	0.97	0.0374	0.1199
OLOGIT (5)	0.89 (0.11)	−0.10	0.99	0.0290	0.1028	0.88 (0.11)	0.10	0.99	0.0376	0.1195

Table 3 continued

Estimation method	Estimation sample ($n = 559$)					Validation sample ($n = 337$)				
	Mean (SD)	Min	Max	MSE	MAE	Mean (SD)	Min	Max	MSE	MAE
OLOGIT (6)	0.89 (0.11)	−0.12	0.98	0.0290	0.1024	0.87 (0.11)	0.09	0.99	0.0381	0.1219
OLOGIT (7)	Convergence not achieved									
OLOGIT (8)	Convergence not achieved									

Dependent variable for OLS, Tobit, Fractional Logistic, CLAD, Two-part and GLM models was EQ-5D score

Dependent variables for OLOGIT models were EQ-5D Dimensions

Independent variable(s): Model (1) PedsQL Total scale score; Model (2) PedsQL Total scale score, age and gender; Model (3) PedsQL Subscale scores; Model (4) PedsQL Subscale scores, age and gender; Model (5) PedsQL Subscale scores, squared terms and interactions; Model (6) PedsQL Subscale scores, squared terms, interactions, age and gender; Model (7) PedsQL items entered as categorical variables; Model (8) PedsQL items entered as categorical variables, age and gender

[a] Parallel regression assumption holds

GLM generalized linear modelling, *MSE* mean squared error, *MAE* mean absolute error, *Min* minimum, *Max* maximum, *OLS* ordinary least squares, *PedsQL* Pediatric Quality of Life Inventory, *SD* standard deviation

4 Discussion

The PedsQL[TM] GCS is an important measure that has been widely used to assess the HRQoL of children, but is not currently preference based. This study shows that it is possible to accurately map PedsQL[TM] GCS scores onto summary EQ-5D-Y utility scores. A range of mapping models was developed in accordance with external guidelines for this purpose [11]. Of all of the models examined, the best two models for predicting utility scores in terms of prediction error were two OLS models based on sub-scale scores, their squared terms and interaction terms (with and without age and gender).

The performance of the two preferred models in terms of overall MAE was similar to previous mapping models, which have obtained MAEs between 0.0011 and 0.19 [29].Their performance in terms of the overall MSE (0.0315–0.0316) was also within the range of other reported mapping studies [29]. The prediction accuracy for both models varied considerably across the range of fitted EQ-5D-Y utility scores (see Table 5) with the MAE for predicted EQ-5D-Y utility scores at the lower end of the utility range more than double that of higher fitted values, a tendency also found in previous mapping studies [38, 39]. Because of the higher prediction errors at the lower end of the utility scale it would be useful to replicate our analyses using data for clinical populations with a broader range of health utility index scores. However, if the purpose is to use predicted scores in cost-effectiveness analyses, then the mean scores across populations will be of greater interest than the individual scores [37, 38]. Therefore, a predictive model that gives group means that are close to the observed values would be suitable for the purpose of mapping, and other reported metrics such as the range of individual errors should complement the assessment of the model's performance.

The mapping models shown in Table 6 can be used to estimate EQ-5D-Y health utilities in situations where only the PedsQL[TM] GCS has been administered. These models require the analyst to have access to individual-level PedsQL[TM] GCS data. They have been estimated using PedsQL[TM] GCS sub-scale scores, their squared terms and interactions, and additionally in the case of one model (OLS 6) basic sociodemographic information. The models are likely to be most accurate for children aged 11–15 years who have similar distributions of PedsQL[TM] GCS scores to those of our study population. The generalisability of our algorithms to other childhood populations requires careful consideration. The items for each of the age-specific modules and self-report or proxy-report formats of the PedsQL[TM] GCS are essentially identical across 2–18 years of age, differing only subtly in terms of developmentally appropriate language, or first- or third-person tense. Similarly, the EQ-5D-Y currently only differs from the adult EQ-5D in terms of its age-appropriate use of language rather than in terms of its structure or underpinning preference weights. In principle, therefore, our mapping algorithms may have broad applications across a wide spectrum of childhood years, potentially inclusive of 2- to 18-year-olds, the age range currently considered empirically amenable to HRQoL measurement. Nevertheless, the performance of our mapping algorithms in childhood populations, which differ according to age or clinical characteristics to our own, remains to be evaluated.

Our study is, to our knowledge, the first to estimate associations between the PedsQL[TM] GCS and the EQ-5D-Y and, therefore, we are not able to gauge our results with those of comparable studies. The PedsQL (with the
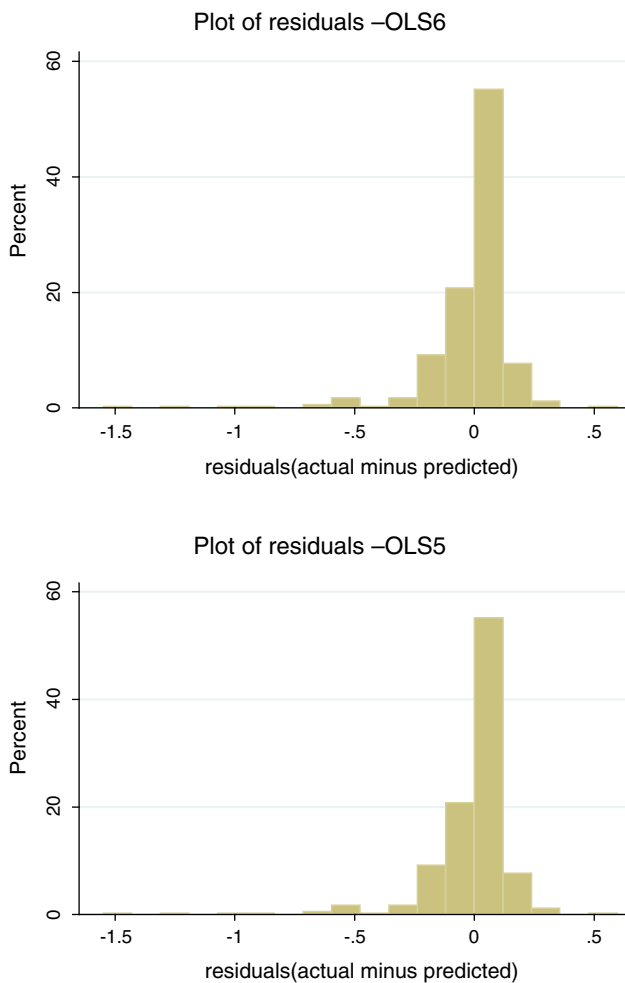
Table 4 Model performance of two best-fitting models

Model no.	Mean (SD)	Min	P.25	Median	P.75	Max	MSE	MAE	Abs diff. <0.10 (%)	Abs diff. <0.25 (%)
Estimation sample										
Observed EQ-5D	0.903 (0.203)	−0.594	0.796	1	1	1	–	–	–	–
OLS (6)	0.903 (0.111)	−0.309	0.890	0.935	0.955	1.047	0.0290	0.1000	70.84	94.99
OLS (5)	0.903 (0.110)	−0.301	0.890	0.938	0.955	1.045	0.0291	0.1006	69.77	95.35
Validation sample										
Observed EQ-5D	0.879 (0.226)	−0.594	0.848	1	1	1	–	–	–	–
OLS (6)	0.889 (0.104)	−0.129	0.861	0.925	0.946	1.060	0.0363	0.1151	65.28	92.88
OLS (5)	0.893 (0.102)	−0.112	0.866	0.928	0.955	1.067	0.0364	0.1140	64.69	93.77

Dependent variable = EQ-5D score

Independent variable(s): Model (5) PedsQL Subscale scores, squared terms and interactions; Model (6) PedsQL Subscale scores, squared terms, interactions, age and gender

Abs diff. absolute differences, *MSE* mean squared error, *MAE* mean absolute error, *OLS* ordinary least squares, *PedsQL* Pediatric Quality of Life Inventory, *SD* standard deviation

Table 5 Distribution of errors by observed EQ-5D utility score

	OLS (6)	OLS (5)
MSE		
EQ-5D range		
<0	1.0154	1.0339
0–0.2	0.3249	0.3282
0.2–0.4	0.1170	0.1214
0.4–0.6	0.1474	0.1527
0.6–0.8	0.0219	0.0223
0.8–1	0.0101	0.0095
MAE		
EQ-5D range		
<0	0.9422	0.9522
0–0.2	0.5677	0.5708
0.2–0.4	0.2860	0.2889
0.4–0.6	0.3797	0.3871
0.6–0.8	0.1329	0.1342
0.8–1	0.0784	0.0762

Dependent variable = EQ-5D score. Independent variable(s): Model (5) PedsQL Subscale scores, squared terms and interactions; Model (6) PedsQL Subscale scores, squared terms, interactions, age and gender

MSE mean squared error, *MAE* mean absolute error, *OLS* ordinary least squares

Fig. 2 Plot of residuals (actual minus predicted) for the two best fitting models

exception of the toddler module covering 2- to 4-year-olds) can be completed by the child and/or a parent proxy, including for adolescents, and there are have been shown to be differences in agreement between self and proxy reports [40]. Future research should explore the performance of our mapping algorithms in data obtained from proxy

Table 6 Models results for the two best-fitting models, $n = 896$

	OLS (6)		OLS (5)	
	Coefficient	Standard error	Coefficient	Standard error
Age	−0.006136	(0.004741)		
Gender	−0.009385	(0.012292)		
PedsQL PF	0.009067***	(0.002571)	0.009127***	(0.002568)
PedsQL EF	0.006807**	(0.002533)	0.006611**	(0.002530)
PedsQL SF	0.005630*	(0.002831)	0.005705*	(0.002829)
PedsQL SchF	0.005802*	(0.002371)	0.006011*	(0.002367)
PedsQL PF Squared	0.000020	(0.000025)	0.000020	(0.000025)
PedsQL EF Squared	−0.000049**	(0.000018)	−0.000048**	(0.000018)
PedsQL SF Squared	0.000011	(0.000016)	0.000011	(0.000016)
PedsQL SchF Squared	−0.000017	(0.000015)	−0.000017	(0.000015)
PedsQL PF × EF	−0.000005	(0.000027)	−0.000004	(0.000027)
PedsQL PF × SF	−0.000053	(0.000029)	−0.000055	(0.000029)
PedsQL PF × SchF	−0.000066*	(0.000030)	−0.000066*	(0.000030)
PedsQL EF × SF	−0.000011	(0.000023)	−0.000009	(0.000023)
PedsQL EF × SchF	0.000061**	(0.000021)	0.000059**	(0.000021)
PedsQL SF × SchF	−0.000026	(0.000022)	−0.000027	(0.000022)
Constant	−0.335861**	(0.118035)	−0.428496***	(0.094210)
Observations	896		896	
Adjusted R^2	0.2870		0.2868	
MSE	0.0315		0.0316	
MAE	0.1063		0.1067	

Standard errors in parentheses

PF physical functioning, *PedsQL* Pediatric Quality of Life Inventory, *EF* emotional functioning, *MSE* mean squared error, *MAE* mean absolute error, *OLS* ordinary least squares, *SF* social functioning, *SchF* school functioning

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

reports as opposed to self-reports. The results presented in this paper are preliminary and we have shown that it is possible to map PedsQL™ GCS scores onto the summary EQ-5D-Y utility score. There is need for further validation and testing before these results can be broadly applied in economic evaluations.

A caveat to the study findings is that although we used a fairly large sample for the mapping analyses, we encountered problems when attempting to use the response mapping approaches to generate predictions. This was because of the fact that our study population largely comprised healthy children with a mean EQ-5D-Y utility score of 0.89. This was reflected in the distribution of scores across the three response levels for each of the five dimensions of the EQ-5D-Y, where the proportion of children scoring a question as level three or "severe or extreme problems" was very low (see Table 2). We have not tested the performance of the estimated models in populations of less healthy children. Future research could overcome this issue by estimating mapping algorithms using data for children with acute or chronic health conditions; this is a further line of enquiry that we are pursuing. A further caveat to the study findings is that we have not estimated mapping algorithms between the PedsQL™ GCS and the EQ-5D-5L, a recently developed modification of the standard EQ-5D that provides five response levels in each dimension [41]. Early evidence suggests that the five-level classification system is less prone to ceiling and floor effects and has greater discriminative ability in some clinical populations [42–45]. Future research in this area may focus on the development of new mapping algorithms between non-preference-based measures of HRQoL developed for childhood and the EQ-5D-5L, if the latter measure becomes the recommended preference-based measure of HRQoL for QALY measurement.

It will, of course, always be preferable to have preference- or utility-based outcomes data collected as part of an evaluation of services for children [11]. When this is not possible, however, the results of our analyses show that it is possible to reasonably predict EQ-5D-Y utility scores from PedsQL™ GCS responses using a regression model framework. It is important to stress that, if predicted scores from mapping are to be used as part of economic evaluations, the uncertainty surrounding these predictions should

be accounted for in the analyses. Nevertheless, our results can be used to inform utility estimation within future economic evaluations of paediatric interventions.

5 Conclusions

This study has shown that it is possible to predict EQ-5D health utilities from PedsQL™ GCS responses. Our mapping algorithms provide an empirical basis for estimating health utilities in childhood when EQ-5D data are not available; they can be used to inform future economic evaluations of paediatric interventions. They are likely to be robust for populations comparable to our own (children aged 11–15 years in attendance at secondary school). The performance of these algorithms in childhood populations, which differ according to age or clinical characteristics to our own, remains to be evaluated.

References

1. Dakin H, Petrou S, Haggard M, Benge S, Williamson I. Mapping analyses to estimate health utilities based on responses to the OM8-30 otitis media questionnaire. Qual Life Res. 2010;19(1):65–80. doi:10.1007/s11136-009-9558-z.
2. Culyer AJ. The dictionary of health economics. Cheltenham: Edward Elgar Publishing; 2005.
3. National Institute for Health and Care Excellence, (NICE). Guide to the methods of technology appraisal. National Institute for Health and Clinical Excellence (NICE); 2013.
4. Petrou S. Methodological issues raised by preference-based approaches to measuring the health status of children. Health Econ. 2003;12(8):697–702.
5. Tilford JM, Payakachat N, Kovacs E, Pyne JM, Brouwer W, Nick TG, et al. Preference-based health-related quality-of-life

6. outcomes in children with autism spectrum disorders. Pharmacoeconomics. 2012;30(8):661–79.
6. Ungar WJ, Boydell K, Dell S, Feldman BM, Marshall D, Willan A, et al. A parent-child dyad approach to the assessment of health status and health-related quality of life in children with asthma. Pharmacoeconomics. 2012;30(8):697–712.
7. Griebsch I, Coast J, Brown J. Quality-adjusted life-years lack quality in pediatric care: a critical review of published cost-utility studies in child health. Pediatrics. 2005;115(5):e600–14.
8. Eiser C, Morse R. Quality-of-life measures in chronic diseases of childhood. Health Technol Assess (Winchester, England). 2001;5(4):1.
9. Longworth L, Bojke L, Tosh J, Sculpher M. MRC-NICE scoping project: identifying the National Institute for Health and Clinical Excellence's methodological research priorities and an initial set of priorities. Centre for Health Economics, University of York Working Papers; 2009.
10. Stevens K. Valuation of the child health utility index 9D (CHU9D). 2010.
11. Longworth L, Rowen D. NICE DSU Technical Support Document 10: the use of mapping methods to estimate health state utility values; 2011.
12. Boyle SE, Jones GL, Walters SJ. Physical activity, quality of life, weight status and diet in adolescents. Qual Life Res. 2010;19(7):943–54.
13. The EQG. EuroQol-a new facility for the measurement of health-related quality of life. Health Policy. 1990;16(3):199–208.
14. Räsänen P, Roine E, Sintonen H, Semberg-Konttinen V, Ryynänen OP, Roine R. Use of quality-adjusted life years for the estimation of effectiveness of health care: a systematic literature review. Int J Technol Assess Health Care. 2006;22(02):235–41.
15. Dolan P. Modeling valuations for EuroQol health states. Med Care. 1997;35(11):1095–108.
16. Szende A, Oppe M, Devlin N. EQ-5D value sets: inventory, comparative review and user guide, vol. 2. Berlin: Springer; 2006.
17. Eidt-Koch D, Mittendorf T, Greiner W. Cross-sectional validity of the EQ-5D-Y as a generic health outcome instrument in children and adolescents with cystic fibrosis in Germany. BMC Pediatr. 2009;9(1):55.
18. Wille N, Ravens-Sieberer U. Age-appropriateness of the EQ-5D adult and child-friendly version: testing the feasibility, reliability and validity in children and adolescents. In: 23rd Scientific Plenary Meeting of the EuroQol Group in Barcelona, Spain, September 14, vol 16. 2006. p. 217–9.
19. Varni JW, Burwinkle TM, Seid M. The PedsQL TM 4.0 as a school population health measure: feasibility, reliability, and validity. Qual Life Res. 2006;15(2):203–15.
20. McCullagh P, Nelder JA. Generalized linear models, vol. 37. London: Chapman & Hall/CRC; 1989.
21. Manning WG, Mullahy J. Estimating log models: to transform or not to transform? J Health Econ. 2001;20(4):461–94.
22. Pearson E, Please N. Relation between the shape of population distribution and the robustness of four simple test statistics. Biometrika. 1975;62(2):223–41.
23. Pregibon D. Goodness of link tests for generalized linear models. Appl Stat. 1980;29:15–23.
24. Hosmer D. In: Lemeshow S (ed). Applied logistic regression. New York: Wiley; 1989. p. 8–20.
25. Glick HA, Doshi JA, Sonnad SS, Polsky D. Economic evaluation in clinical trials. USA: Oxford University Press; 2007.
26. Powell JL. Least absolute deviations estimation for the censored regression model. J Econometr. 1984;25(3):303–25.
27. Chay KY, Powell JL. Semiparametric censored regression models. J Econ Perspect. 2001;15(4):29–42.

28. Tobin J. Estimation of relationships for limited dependent variables. Econometrica: J Econ Soc. 1958;26(1):24–36.

29. Brazier J, Yang Y, Tsuchiya A, Rowen D. A review of studies mapping (or cross walking) non-preference based measures of health to generic preference-based measures. Eur J Health Econ. 2010;11(2):215–25. doi:10.1007/s10198-009-0168-z.

30. Pullenayegum EM, Tarride J-E, Xie F, Goeree R, Gerstein HC, O'Reilly D. Analysis of health utility data when some subjects attain the upper bound of 1: are Tobit and CLAD models appropriate? Val Health. 2010;13(4):487–94.

31. Pullenayegum EM, Tarride J-E, Xie F, O'Reilly D. Calculating utility decrements associated with an adverse event marginal Tobit and CLAD coefficients should be used with caution. Med Decis Making. 2011;31(6):790–9.

32. Papke LE, Wooldridge JM. Econometric methods for fractional response variables with an application to 401 (K) plan participation rates. J Appl Econ. 1996;11(6):619–32. doi:10.2307/2285155.

33. Levy A, Christensen T, Johnson J. Utility values for symptomatic non-severe hypoglycaemia elicited from persons with and without diabetes in Canada and the United Kingdom. Health Qual Life Outcomes. 2008;6(1):73.

34. Dakin H, Gray A, Murray D. Mapping analyses to estimate EQ-5D utilities and responses based on Oxford Knee Score. Qual Life Res. 2012;1–12.

35. Gray AM, Rivero-Arias O, Clarke PM. Estimating the association between SF-12 responses and EQ-5D utility values by response mapping. Med Decis Making. 2006;26(1):18–29. doi:10.1177/0272989x05284108.

36. Le QAPP, Doctor JNP. Probabilistic mapping of descriptive health status responses onto health state utilities using Bayesian networks: an empirical analysis converting SF-12 into EQ-5D utility index in a National US sample. Med Care. 2011;49(5):451–60.

37. Kaambwa B, Billingham L, Bryan S. Mapping utility scores from the Barthel index. Eur J Health Econ. 1–11. doi:10.1007/s10198-011-0364-5.

38. Pinedo-Villanueva R, Turner D, Judge A, Raftery J, Arden N. Mapping the Oxford hip score onto the EQ-5D utility index. Qual Life Res. 1–11. doi:10.1007/s11136-012-0174-y.

39. Tsuchiya A, Brazier J, McColl E, Parkin D. Deriving preference-based single indices from non-preference based condition-specific instruments: converting AQLQ into EQ5D indices. 2002. http://eprints.whiterose.ac.uk/10952/

40. Cremeens J, Eiser C, Blades M. Factors influencing agreement between child self-report and parent proxy-reports on the Pediatric Quality of Life Inventory[TM] 4.0 (PedsQL[TM]) Generic Core Scales. Health Qual Life Outcomes. 2006;4(1):58.

41. Herdman M, Gudex C, Lloyd A, Janssen M, Kind P, Parkin D, et al. Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L). Qual Life Res. 2011;20(10):1727–36.

42. Janssen MF, Pickard AS, Golicki D, Gudex C, Niewada M, Scalone L, et al. Measurement properties of the EQ-5D-5L compared to the EQ-5D-3L across eight patient groups: a multi-country study. Qual Life Res. 2012. 1–11. doi:10.1007/s11136-012-0322-4.

43. Pickard AS, De Leon MC, Kohlmann T, Cella D, Rosenbloom S. Psychometric comparison of the standard EQ-5D to a 5 level version in cancer patients. Med Care. 2007;45(3):259–3. doi:10.1097/01.mlr.0000254515.63841.81.

44. Kim S, Kim H, Lee S-I, Jo M-W. Comparing the psychometric properties of the EQ-5D-3L and EQ-5D-5L in cancer patients in Korea. Qual Life Res. 2012;21(6):1065–73. doi:10.1007/s11136-011-0018-1.

45. van Hout B, Janssen M, Feng Y-S, Kohlmann T, Busschbach J, Golicki D, et al. Interim scoring for the EQ-5D-5L: mapping the EQ-5D-5L to EQ-5D-3L value sets. Val Health. 2012;15(5):708–15.