This is a repository copy of *Mapping CHU9D Utility Scores from the PedsQL™ 4.0 SF-15*.

White Rose Research Online URL for this paper:
http://eprints.whiterose.ac.uk/167574/

Version: Accepted Version

Article:
Mpundu-Kaambwa, C., Chen, G., Russo, R. et al. (3 more authors) (2017) Mapping CHU9D Utility Scores from the PedsQL™ 4.0 SF-15. PharmacoEconomics, 35 (4). pp. 453-467. ISSN 1170-7690

https://doi.org/10.1007/s40273-016-0476-y

This is a post-peer-review, pre-copyedit version of an article published in PharmacoEconomics. The final authenticated version is available online at: http://dx.doi.org/10.1007/s40273-016-0476-y.

Reuse
Items deposited in White Rose Research Online are protected by copyright, with all rights reserved unless indicated otherwise. They may be downloaded and/or printed for private study, or other acts as permitted by national copyright laws. The publisher or other rights holders may allow further reproduction and re-use of the full text version. This is indicated by the licence information on the White Rose Research Online record for the item.

Takedown
If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

White Rose
university consortium
Universities of Leeds, Sheffield & York

eprints@whiterose.ac.uk
https://eprints.whiterose.ac.uk/

Mapping CHU9D utility scores from the PedsQL™ 4.0 SF-15

Authors: Christine Mpundu-Kaambwa[1], Gang Chen[2], Remo Russo[3, 4], Katherine Stevens[5], Karin Dam Petersen[6], Julie Ratcliffe[7]

Affiliations:

[1] University of South Australia Business School, University of South Australia, Adelaide Australia

[2] Centre for Health Economics, Monash Business School, Monash University, Melbourne, Australia

[3] Faculty of Health Sciences, School of Medicine, Flinders University, South Australia, Australia

[4] Department of Paediatric Rehabilitation, Women's and Children's Hospital, North Adelaide, South Australia

[5] Health Economics and Decision Science, University of Sheffield, Sheffield, United Kingdom

[6] Department of Business and Management, Faculty of Social Sciences, Aalborg University, Aalborg East, Denmark

[7] Flinders Health Economics Group, Flinders University, South Australia, Australia

*Corresponding Author: Christine Mpundu-Kaambwa

Address: University of South Australia Business School, University of South Australia, North Terrace, Adelaide SA 5000 Australia

P: +61 8302 0228 Email: mpucc001@mymail.unisa.edu.au

Abstract

Background: The PedsQL™ 4.0 Short Form 15 Generic Core Scales (hereafter the PedsQL) and the Child Health Utility 9 Dimension (CHU9D) are two generic instruments designed to measure health-related quality of life in children and adolescents in the general population and paediatric patient groups living with specific health conditions. Although the PedsQL is widely used among paediatric patient populations, presently it is not possible to directly use the scores from the instrument to calculate quality adjusted life years (QALYs) for application in economic evaluation because it produces summary scores which are not preference-based.

Objective: This paper examines different econometric mapping techniques for estimating CHU9D utility scores from the PedsQL for the purpose of calculating QALYs for cost-utility analysis.

Methods: The PedsQL and the CHU9D were completed by a community sample of 755 Australian adolescents aged 15-17 years. Seven regression models were estimated: ordinary least squares estimator, generalised linear model, robust MM-estimator, multivariate factorial polynomial estimator, beta-binomial estimator, finite mixture model and multinomial logistic model. The mean absolute error (MAE) and the mean squared error (MSE) were used to assess predictive ability of the models.

Results: The MM-estimator with stepwise-selected PedsQL dimension scores as explanatory variables had the best predictive accuracy using MAE and the equivalent beta-binomial model had the best predictive accuracy using MSE.

Conclusions: Our mapping algorithm facilitates the estimation of health state utilities for use within economic evaluations where only PedsQL data is available and is suitable for use in

community based adolescents aged 15-17 years. Applicability of the algorithm in younger populations should be assessed in further research.

Key points for decision makers

- The PedsQL is a validated, well known and widely used measure of health-related quality of life among children and adolescents.
- Currently, the PedsQL is not suitable for measuring and valuing health-related quality of life in children and adolescents for the purposes of directly calculating quality adjusted life years for application in economic evaluation.
- This study generated mapping algorithms for transforming PedsQL scores to CHU9D utility scores

1 Introduction

Health-related quality of life (HRQoL) has increasingly become a key outcome measure in clinical trials, public health and health services research aimed at evaluating the effectiveness and efficacy of interventions in children and adolescents [1]. HRQoL is a multidimensional construct that measures an individual's subjective assessment of their physical health and psychosocial functioning [1, 2]. HRQoL measures are classified into non-preference-based and preference-based measures. Non-preference-based measures are widely used in paediatric populations and are premised on simple summary scoring of individual items or dimensions to generate HRQoL scores[1]. However, preference-based measures (whereby the individual items or dimensions are weighted according to their relative importance in determining overall HRQoL) are required for economic evaluation. Valuation methods such as the standard gamble (SG), time trade-off (TTO) or the rating scale (visual analogue scale) techniques have typically been employed in general population samples to generate the health state utilities (or weights) for preference-based measures on a cardinal scale which is anchored at 0.0 (representing being death) and 1.0 (representing full health) [3].

Preference-based measures can be applied to a broad set of conditions and population groups, and can be used to calculate QALYs within cost utility analysis (CUA) [3]. Many decision-making bodies recommend the use of the QALY as a standard measure of benefit in economic evaluation of new health care technologies and pharmaceuticals [4, 5].

One of the most predominant non-preference-based HRQoL measures frequently employed in paediatric health research and evaluations, is the Pediatric Quality of Life Inventory™ Version 4.0 Generic Core Scales (PedsQL™) [6]. The PedsQL™ uses a simple summative scoring system to generate dimension and total scores. [6]. Currently, the PedsQL™ is not

preference-based and therefore cannot be directly or readily used in CUA. Directly collecting preference-based data using a purposefully designed measure remains the optimal way of collecting such data for purposes of conducting a CUA [1]. However, in the absence of such data and provided there is a certain degree of overlap in their construction and conceptual development, it is possible to derive a mapping algorithm to predict health state utilities from the PedsQL^{TM} by observing its statistical relationship with a generic preference-based measure. This technique is referred to as 'mapping' or 'cross walking' [1]. A recent review of mapping models reports that mapping is a valid method of predicting health state utility values for the purposes of conducting an economic evaluation when no preference-based measure is used [7]. To date few mapping studies have been conducted exclusively within child and adolescent populations [8-11]. Of these, two studies have generated mapping algorithms for transforming PedsQL scores onto multi-attributable instruments [10, 11]. Payakachat and colleagues derived Health Utilities Index Mark 3 scores from the 23-item PedsQL using proxy-reports for a sample of 4-17 year old children with autism [11]. On the other hand, Khan and colleagues generated health utilities from the 23-item PedsQL^{TM} questionnaire responses onto the youth version of the EuroQoL 5 dimensions (EQ-5D-Y) for a sample of 11-15 year old children based on the tariff set derived from the general population of UK adults (York A1)[10].

This paper reports the results of a regression-based exercise that maps response items from the Short Form 15 item (SF-15) version of the PedsQL^{TM} or PedsQL for brevity [6, 12], to the CHU9D [13-15]. The 15 items within the PedsQL detailed in section 2.2 are a sub-set of the 23 items in the PedsQL^{TM}. This empirical examination focuses on data from a community-based cross-sectional sample of Australian adolescents. To our knowledge, this is the first study internationally to estimate a mapping algorithm between these two measures

and the first to apply the newly updated CHU9D Australian adolescent specific scoring algorithm [16].

2 Methods

2.1 Study Design

The study used data obtained from a web-based, randomly selected representative sample of Australian adolescents aged 15-17 years. Following participant and parent dyad consent, adolescents completed a 3-section questionnaire. The first section comprised the CHU9D, the second section contained the PedsQL and the final section included a questions on socio-demographic and general health characteristics including age, gender, socioeconomic status as measured by the Family Affluence Scale (FAS) (a four item measure of family wealth) [17], self-reported general health and the presence or absence of long term disability, illness or medical condition/s. Ethical approval for this study was obtained from the Social and Behavioural Research Ethics Committee, Flinders University (project number 5508).

2.2 Outcome Measures

The source instrument for mapping was the PedsQL and the target instrument was the CHU9D. The CHU9D is a relatively new generic preference-based measure of HRQoL originally developed for application with children aged 7-11 years [13, 14, 18]. A programme of recent research has also demonstrated its practicality and validity in adolescents aged 11 - 17 years [19, 20]. The CHU9D consists of nine dimensions (worry, sadness, pain, tiredness, annoyance, school, sleep, daily routine and activities), each with five impairment levels ranging from 1 (no problems) to 5 (severe problems) that assesses the child/adolescent's assessment of their own HRQoL "today". In this study, responses to the CHU9D were scored using an Australian adolescent population-specific scoring algorithm which was generated using profile case best-worst scaling methods [16]. The Australian scoring algorithm was

generated on a utility scale and ranges from -0.1059 for the ‘pits’ (most severe) health state to 1.0 for the best state

The PedsQL is a generic non-preference-based, four-dimension instrument designed to measure HRQoL in healthy children, as well as in those with different health conditions [12]. It has four dimensions (consisting of 15 items altogether) which measure the following: physical functioning (5 items), emotional functioning (4 items), social functioning (3 items), and school functioning (3 items). The PedsQL has a child self-report for ages 5-7 years (young child), 8-12 years (child) and 13-18 years (adolescent), as well as a parent proxy-report for ages 2-4 years (toddler), 5-7 years (young child), 8-12 years (child) and 13-18years (adolescent). For the purposes of this study, the adolescent self-report version of the instrument was utilised. Respondents rate their answers on a 5-point likert scale with one of the following preferences: 0 “never a problem”, 1 “almost never a problem”, 2 “sometimes a problem”, 3 “often a problem”, 4 “almost always a problem”. Items are then reverse-scored and linearly transformed into a total score ranging from 0 to 100 (where 0=100, 1=75, 2=50, 3=25, 4=0). Higher total scores represent better HRQoL [6, 12]. The mean total score is a summation of all the items over the number of items answered.

2.3 Statistical Analysis

Participant characteristics were summarized as means (± standard deviations (SD)) for continuous variables and frequency (%) for categorical variables. We tested for normality of variables using the Shapiro-Wilks test. The correlation between the CHU9D and the PedsQL was estimated using a scatterplot and Spearman correlation coefficients.

This study was conducted in accordance with the newly developed 'Mapping onto Preference-based measures reporting Standards' (MAPS) checklist [21]. To develop the optimal mapping algorithm of CHU9D utility scores from the PedsQL, a direct mapping technique was firstly used which included regressing the PedsQL total, dimension and item scores directly onto the CHU9D utilities. An indirect mapping approach was further considered, in which the response levels of each of the CHU9D dimensions were predicted. To improve predictive performance, socio-demographic characteristics, specifically age and gender were also included [7]. Three regression model specifications were estimated, henceforth simply referred to as "models". The models were depicted algebraically as:

$$CHU9D\ /CHU9D\_Dimen = \alpha + \beta_1 * PedsQL + \beta_2 * PedsQL^2 + \delta_1 * Age + \delta_2 * Gender$$

(Model 1)

$$CHU9D\ /CHU9D\_Dimen = \alpha + \sum_{i=1}^{k} \gamma_i * PedsQL\_Dimen\_sw_{1i} + \sum_{i=1}^{k} PedsQL\_Dimen\_sw_{1i}^2 + \delta_1 * Age + \delta_2 * Gender$$ (Model 2)

$$CHU9D\ /CHU9D\_Dimen = \alpha + \sum_{j=1}^{m} \gamma_j * PedsQL\_Item\_sw_{1j} + \sum_{j=1}^{m} PedsQL\_Item\_sw_{1j}^2 + \delta_1 * Age + \delta_2 * Gender$$ (Model 3)

where CHU9D is the CHU9D utility score, CHU9D_Dimen is one of the 9 CHU9D dimensions listed in section 2.2., PedsQL is the PedsQL score, PedsQL_Dimen_sw is the PedsQL dimension scores, PedsQL_Item_sw is the PedsQL items, i is number of PedsQL dimensions selected based upon statistical significance using stepwise regression methodsi<= 4and j is number of PedsQL items selected using stepwise regression techniques j<=15. Despite its limitations, stepwise regression is still widely used within regression literature

reported in high impact journals [22-26]. The significance level of statistical inference is 5% in this study.

Age and gender were omitted as they were statistically insignificant. In Model (1) the dependent variable was predicted from the PedsQL raw total scores and squared terms. To estimate the non-linear relationship between the PedsQL and CHU9D PedsQL squared terms were added as explanatory continuous variables. In Model (2), dimension scores and their squared terms were modelled as the only independent continuous variables. Similarly, Model (3), contained the item scores and their squared terms entered as continuous variables. Forward selection stepwise regression modelling was used to determine which of the 15 items and 4 dimensions were to be included in the final predictive models. Item responses and dimension scores that were not statistically significant (p-value >0.05) in the stepwise regression models were not included in the final models. All analyses were conducted in Stata version 14.1 [27].

A number of regression models have been used in direct and indirect (response) mapping exercises and were therefore considered for the work undertaken in this paper [28, 29]. For direct mapping, candidate models included: ordinary least squares (OLS), Tobit, censored least absolute deviations (CLAD), generalised linear modelling (GLM), robust MM-estimator (MM), fractional logistic regression (FLOGIT), two-part (TP) models, beta-binomial (BB) estimator, multivariable fractional polynomials (MFP), finite mixture models (FMM) and generalized additive (GA) models. For the response mapping, possibilities were multinomial logistic, ordinal logistic regression and generalised ordered probit models [30]. Theoretically, some models are seen to be better suited to deal with problems associated with utility scores such as skewness (e.g. GLM, BB, MFP), heteroscedasticity (e.g. CLAD, GLM, MM), ceiling

effects (e.g. CLAD, BB, Tobit, TP) and the potential presence of outliers (e.g. MM). Despite some models having theoretical advantages over others, the performance of these models in the literature has been mixed [29]. Compared to the CLAD, for instance, the OLS performed better in some instances [31] but worse in others [32]. As it was not possible to compare all models identified in the literature in our study, a decision was made to use a wide spectrum of models that account for the theoretical considerations associated with the distribution of the CHU9D and outlined against each model below. In total, seven statistical methods were adopted in this study (the first six employed direct mapping while the seventh used indirect/response mapping).

- Ordinary least squares (OLS) [33] is the most widely used mapping model reported in literature [33]. Although theoretically it does not perform well in the presence of heteroscedasticity and non-normality of residuals, in practice it has been found to have good performance in the mapping literature [34].
- Generalised linear modelling (GLM) [35], was included in the study as it allows for skewed distributions of the dependent variable such as is the case in this study (see Kernel density plots Figure 1). Modified Parks tests suggested by Manning were used to guide the choice of the GLM distribution and link [27].
- The Robust MM-estimator (MM) [34] is designed to deal with some limitations of traditional regression methods including heteroscedasticity and the presence of outliers [34]. It was firstly introduced into the mapping literature by Chen and colleagues for both adolescents and adult sample and found to have good performance [34, 8].
- Multivariable fractional polynomials (MFP) [36] are useful for modelling non-linear relationship between the independent and dependent variables while preserving the continuous nature of the covariates [36]. At each step of the algorithm, MFP

constructs a fractional polynomial transformation for each continuous covariate while fixing the current functional forms of the other covariates. We used the GLM method to fit a MFP model [36].

- The Beta-binomial (BB) estimator [37, 38] is robust to skewness and can estimate both unimodal and bimodal utilities. It is in the class of Beta regressions shown to be superior to alternative regression strategies such as the OLS [39]. The BB estimator has a potential limitation: it restricts utilities to a 0 to 1 range [24, 25]. However, only one observation in our dataset had a CHU9D score of -0.046 which was subsequently set equal to 0 as has been done in previous mapping studies [40].

- Finite mixture models (FMM) combine two or more probability density functions making them capable of approximating any arbitrary distribution [41, 42]. Because of its flexibility, the FMM is able to handle complex and multimodal distributions that often characterise health-related quality of life data such as the EQ-5D [43]. As a visual inspection of the distribution of the CHU9D did not clearly reveal how many components the CHU9D had, models with up to six components were estimated and, following guidance in the literature [44], the model with the smallest Akaike's or Schwarz's Bayesian information criteria (AIC or BIC) was chosen as the final one.
- Multinomial logistic regression (MLOGIT) was used for the response mapping onto CHU9D dimensions [45]. In this model, the probability of obtaining a particular level on each CHU9D dimension was estimated and then converted into utilities using the ‘expected value approach’ [46]. An advantage of the indirect response mapping is that it allows for health state descriptions to be determined after which the appropriate country- or population-specific CHU9D tariff can be applied [29].

2.4 Assessing Model Performance

The goodness-of-fit of the estimated models was assessed using a number of metrics as was appropriate: linktest (OLS, GLM, MFP); Hausman's specification test (OLS, GLM, MLOGIT); coefficient of determination - R^2 (OLS, MM, BB) and, lastly, AIC and BIC (GLM, MFP, MLOGIT, BB). The predictive ability of the models was mainly assessed using two measures of predictive accuracy, the mean absolute error (MAE) and the mean squared error (MSE). Four additional criteria, estimated using the validation sample were used to further assess the predictive ability of the models. These criteria were: (a) exactness of the predicted mean, (b) the range (the difference between the upper and lower limits) of predictions, (c) the proportion of predicted utilities deviating from observed values by absolute error <0.03 and <0.05 and, (d) the intraclass correlation coefficients [47] depicting the level of absolute agreement between predicted and observed CHU9D scores. The MAE was calculated as the mean of the absolute values of the difference between the predicted and actual observed CHU9D values. The MSE was computed as the mean squared differences between the actual and predicted CHU9D utilities. A lower MAE or MSE for a predictive model indicates higher accuracy in predictions and therefore a better performing model.

More weight was put on the MAE because the MSE is more sensitive to potential outliers compared to the MAE[48]. Further, it has been suggested that the MSE has low reliability and may therefore produce different results depending on different fractions of data [49], as was the case in this study where a number of different samples of data were used for the estimation and validation exercises.

As an external validation dataset is currently not available, predictive model performance was assessed using in-sample datasets which were generated using two methods. In validation

method 1 (the k-fold method), the original sample was randomly divided into five equal-sized subgroups using random number generation algorithms. In each iteration, four of the groups (80% of the dataset) were combined and then allocated to the "estimation sample" while the remaining group (20% of the dataset) was used as the "validation sample". This process was repeated five times, so as to make certain that each of the five subgroups was used in the estimation and validation iterations. Thereafter, the validation results were pooled together and model performance based on the pooled estimated goodness of fit statistics (MSE and MAE) was assessed. This validation technique is also referred to as cross validation [50, 51]. Validation method 2 involved generating three random samples from the original data sample with sample sizes of n=100 (Sample 1), n=300 (Sample 2) and n=500 (sample 3). Predictive models estimated using the entire dataset were validated on each of these three samples. The models that performed the best in both validation techniques were chosen as optimal.

3 Results

3.1 Study Sample Characteristics

A total of 755 adolescents completed both the CHU9D and the PedsQL. Table 1 presents the descriptive characteristics of respondents. The mean (± SD) of the CHU9D utility and PedsQL total scores were 0.724 (± 0.218) and 72.861 (± 16.562), respectively. Fifty three percent of respondents were female; 59% were from families with high socioeconomic status; 94% reported themselves as having excellent, very good or good health; and 10% reported having had a long term disability, illness or medical condition.

Figure 1 shows the Kernel density plot of the CHU9D utilities and the PedsQL total scores. Both the CHU9D utility scores and PedsQL total scores appeared non-normally distributed (negatively skewed). Further investigation using the Shapiro Wilks test of normality led to

the formal rejection of the null hypothesis that the CHU9D and the PedsQL were normally distributed. Figure 2 presents a scatterplot depicting the linear relationship between the CHU9D utility and the PedsQL total scores. A positive association was observed from the scatterplot and this result was corroborated by a moderately strong statistically significant correlation (Spearman's correlation Coefficient (r) = 0.632; p<.0001). At the dimension level (results available from authors upon request), the correlation between the CHU9D utility score and the PedsQL dimensions listed in parentheses ranged from r = 0.432, (social functioning) to r = 0.592, (emotional functioning). The correlation between the CHU9D utility score and the physical and school functioning dimensions were r = 0.463 and r = 0.532, respectively. All correlations were statistically significant (p-value <0.05). These results are in alignment with a qualitative exercise conducted as a separate exercise that provided an empirical assessment of the conceptual overlap between dimensions of the CHU9D and those of the PedsQL (Supplementary Material Table 1). Following discussion amongst authors of the present study and a review of the literature, correlations between particular dimensions of the two instruments were hypothesised and thereafter estimated. The resultant correlation analysis suggested moderate to strong association between the two instruments as each of the PedsQL dimensions (with the exception of 'social functioning' dimension) could be matched conceptually to at least two CHU9D dimensions. These hypotheses were confirmed by the results, to be presented in a separate paper, of a quantitative analysis undertaken by the authors of the current study on the same sample assessing convergence between the dimensions of the two instruments (results available from the authors upon request).

3.2 Prediction of CHU9D utility scores

Seven statistical methods and three model specifications were assessed separately using the estimation and validation samples based on validation methods I (k-fold) and II (three random samples). Table 2 presents the results summarising the key goodness-of-fit statistics for different model and method combinations based on the full estimation sample. All estimators tended to overestimate the lower limit of the utility score with the exception of the MM-1 lower limit = -0.210 and MM-2 = -0.136 which under predicted the observed minimum score. The closest estimate to the observed value was the OLS estimate (model 2) with lower limit = 0.017. On the upper boundary, the least absolute difference was shown for GLM-3 and MFP-3 (0.972 vs 1 for the predicted and observed CHU9D values). In terms of the MAE and MSE, the MM-estimate (model 3) had the lowest MAE (0.1251) and model (2) had the second lowest MAE (0.1261). With reference to the MSE, we observed some mixed results. The OLS-3 estimator had the lowest MSE (0.0256) and the second lowest MSE was GLM-2, MFP-2 and BB-3 with a score of (0.0259). Based on the results in Table 2 we conclude that the mapping algorithm using the MM-estimator models 2 and 3 are the two best models using the MAE criterion.

3.3 Validation

The primary models obtained using the estimation samples were validated on random samples using validation methods I (k-fold) and II (three random samples). Table 3 shows the validation analysis results based on both validation methods. In terms of accurately predicting the mean CHU9D utility score, the MM and OLS estimators were able to accurately predict the mean CHU9D utility score in most instances. None of the models were able to predict negative utility scores. All models were assessed for goodness of fit using the MAE and MSE and a consistent pattern was seen from this assessment. In particular, dimension level models (model 2) invariably performed better in both the MAE and MSE than item level or total

score models. The MAE ranged from 0.1169 (MM-2 in validation method II) to 0.1348 (MM-1 in validation II) while the MSE ranged from 0.0212 (BB-2 in validation II) to 0.0293 (OLS-1 and BB-1 in validation II). The results in Table 3 indicate that, when all regression models were categorised based on the MAE, the study key selection criteria, the MM model (2) exhibited the best predictive ability while the BB model (2) was best on the MSE.

3.4 Best performing models

The best performing models were selected on the basis of their performance in all four validation samples with more weight put on the MAE following guidance in the literature [48, 49]. For model specification 1, the MM performed best on the MAE the most times (in the pooled sample of validation I and random sample I of validation II) and was also best on the MSE in random sample 1 of validation II. For model specification 2, the MM again performed best the most times when assessed against the MAE (in the pooled sample of validation I and random samples 1 and 3 of validation II). In terms of model specification 3, the BB performed best the most times when gauged against the MAE (random samples 1 and 3 of validation II). The BB also performed was best on the MSE in all random samples of validation II. On the basis on these results, the MM (1), MM (2) and BB (3) were picked as the best performing models and their detailed performance statistics are presented in Table 4. Figure 3 shows that the residuals for these models were similar.

When the three best performing models were compared in terms of the MAE, MM (2) had on average the lowest predictive error followed by BB (3). A similar pattern was seen for the MSE. There was mixed performance by these models in terms of the comparison between predicted and observed utilities with all models under predicted the mean CHU9D utilities in

some validation samples and over predicting in others. However, predicted utilities from MM (2) had the highest agreement (ICC) to the observed utilities followed by MM (1). Table 4 also shows that MM (2) had the highest proportion of predicted utilities deviating from observed values by <0.03 (21%) followed by MM (1) (18%). The distribution of the MAE and MSE for all three models across the range of observed CHU9D utility scores was also examined (Table 5). A similar pattern was seen for all three models with the biggest errors generated for the 0-0.2 and 0.2-0.4 ranges and the lowest errors associated with the 0.6-0.8 range. In general, predictive performance for all models was better at higher CHU9D ranges. On the basis of these results, we propose using MM model with dimension scores as explanatory variables (MM 2) to predict CHU9D utility scores from the PedsQL.

3.5 Mapping equations

The regression model coefficients for predicting the CHU9D utility scores using MM (1) and BB (3) are presented together with those for best model (MM 2) in Table 5. Age and gender were consistently insignificant when included in all the models and therefore were not included in the final models. In MM (1), both the PedsQL total score and its squared term were robustly significant ($p<0.05$) suggesting the existence of a non-linear relationship between the CHU9D and PedsQL. A nonlinear relationship was again seen in MM (2) and BB (3) where the squared term of the emotional functioning dimension in the former and the squared term of the 'it is hard for me to walk for more than one block' item in the physical functioning dimension of the latter were statistically significant. All the coefficients for the statistically significant PedsQL dimensions and items were positive meaning less functional impairment is associated with a higher quality of life. Potential multicollinearity was assessed using the variance inflation factor (VIF). None of the selected PedsQL dimensions and items

had a VIF>10 and were therefore not omitted from the final regression models. Based on findings in section 3.4, optimal equation (for MM 2) would be expressed as follows:

$$CHU9D\ utility\ score$$
$$= -0.135516 + 0.264648 * PedsQL\ Physical\ Functioning + 1.196678$$
$$* PedsQL\ Emotional\ Functioning + 0.203405\ PedsQL\ School\ Functioning$$
$$- 0.572612 * PedsQL\ Emotional\ Functioning\ Square$$

The variance-covariance matrix for this model (MM 2) is reported in the supplementary appendix.

4 Discussion

Accurate measurement and valuation of HRQoL within paediatric and adolescent populations is an important component of economic evaluations of health care interventions targeted for children. The PedsQL and the CHU9D have both been demonstrated as practical and valid instruments for the measurement of HRQoL in children and adolescents. However and in contrast with the CHU9D, the PedsQL is not currently preference-based. This study has developed a mapping algorithm that can reasonably predict CHU9D utility scores from the PedsQL for the purpose of conducting cost utility analysis when health state utility data is not collected. In accordance with guidelines for mapping studies [52], a series of statistical methods were considered for this exercise including: OLS, GLM, MM, MFP, BB, FMM, and MLOGIT. The MM model with PedsQL dimension scores (model 2) was found to be the best in terms of predictive accuracy as assessed by MAE, the range, and proportions of the predicted mean CHU9D utilities deviating from the observed values by <0.03 and <0.05. These mapping algorithms are not intended to be a substitute for validated utility based

measures but rather as an alternative technique that can be employed when preference-based measures are not available or cannot be used.

Previous studies have shown conceptual overlap between the target and initial instruments to be an important determinant of successful mapping analysis [7, 53, 29]. In our study, both instruments measure similar constructs and as such have conceptual overlap between each other. This commonality was corroborated by the moderate to strong correlation (r = 0.632) observed between the CHU9D utility and the PedsQL total score. As anticipated, three of the four PedsQL dimensions, including physical (r = 0.463), emotional (r = 0.592) and school (r = 0.532) functioning that are covered in the CHU9D dimensions had moderate correlation with the CHU9D. In contrast social functioning which is not covered in the CHU9D dimension had the smallest but significant correlation (r = 0.432) with the CHU9D. These results are in agreement with those obtained from a qualitative assessment of the conceptual overlap between the two instruments where dimensions of the PedsQL were pitted against comparable CHU9D dimensions based on discussions between authors of this paper (Supplementary Material Table 1).

In this dataset, the MM performed better than other models considered in terms of predictive accuracy, producing a wider range of predicted CHU9D utility scores. The values of the MAE and MSE statistics obtained for the MM models (0.1169 to 0.1348 and 0.0213 to 0.0311, respectively) were all on the lower end of the range reported in the literature for mapping studies [7]. As no comparable mapping studies between the PedsQL and the CHU9D have been previously reported, it was not possible to compare our findings with previous mapping algorithms. The heterogeneous performance of the MM across the range of observed CHU9D utility scores has been shown elsewhere [10, 54, 55]. As our sample had a

lower proportion of respondents with low CHU9D utility scores (i.e. only 71 participants or 9% of the sample had scores ≤ 0.4), we recommend replicating this analysis on a sample with a bigger representation of individuals with lower quality of life.

This study has some limitations. Firstly, no external sample dataset is currently available and therefore in-sample validation of mapping algorithms, effectively used in several previous mapping studies [29, 56-58] was utilised in this study. Secondly, our mapping study sample was comprised largely of healthy adolescents with a mean CHU9D score of 0.72. Therefore, it is possible that the models assessed in this study will be more applicable to adolescent populations with similar socio-demographic characteristics and HRQoL statuses as the study population. Further research should be conducted to assess the performance of the mapping algorithm developed here in adolescents living in the community with health conditions and in adolescent patient populations. Thirdly, our data were obtained through a web based survey. The advantages of a web based mode of administration for a survey of this nature include its increasing familiarity, particularly for young people and its ability to engage large numbers of community based adolescents who would otherwise be more difficult to reach. Potential disadvantages include concerns about data quality and that participants may not provide accurate information. However, appropriate data checks were applied, to effectively deal with this limitation, including the wording of the questions, the order in which the questions are asked, question type and design. Fourthly, it was not possible to obtain detailed data from the online panel company that conducted the survey on the sampling frame from which our study sample was drawn or on the specific sampling procedures applied. However, our study sample was broadly representative of a similar cohort of 15-17 adolescents from the Australian general population in terms of gender split (53% female in our study versus 49% in the general population [59]) and responses to the general health question (92% of

adolescents aged 15 to 17 years in the current survey reported themselves to be in "excellent," "very good," or "good," health as opposed to 93% of adolescents aged 15-24 years in the general population who participated in the National Health Survey [60]). Similarly, our sample was of similar self-reported disability status (10% of our study participants had a long term disability, illness or medical condition compared to 7% of 15-24 years olds in the general population [60]. Finally it is important to note that the mapping algorithms have been developed on the basis of a sample of 15-17 year olds and therefore may not be applicable in samples of younger children. However, age was found to be insignificant in deriving mapping algorithms on adolescents sample in this study and elsewhere [8, 10]. Applying the reported mapping algorithm on a younger adolescent sample should be used with caution.

5 Conclusion

To our knowledge this is the first empirical study internationally that has derived CHU9D utility scores from SF-15 item version of the PedsQL^{TM}. Our results show that it is possible to predict CHU9D utility scores from the PedsQL SF-15 with best results obtained when MM model (2) (with dimension scores as explanatory variables) is used. This mapping algorithm maybe usefully applied for the prediction of CHU9D utilities from the PedsQL thereby facilitating the calculation of QALYs for assessing the relative cost-effectiveness of new health care technologies and pharmaceuticals targeted at young people. Future research should replicate this analysis on samples aged below 15 years to test the robustness of these algorithms on younger populations in which the PedsQL can be applied. Development of preference weights for the family of PedsQL instruments is another alternative that would enable its future use in economic evaluation.

Data Availability Statement

The datasets generated during and/or analysed during the current study are not publicly available as analysis is still ongoing, but are available from the corresponding author on reasonable request**.**

Compliance with Ethical Standards

Acknowledgements

We would like to thank all the study participants who generously gave up their time to participate in this study.

Author contributions

C. Mpundu-Kaambwa - analysed the data, interpreted the results, wrote the first draft and will act as a guarantor for the work.

G. Chen - formulated the idea for the study, oversaw the design and collection of data, analysed the data, interpreted the results and made critical revisions to the manuscript.

R. Russo - interpreted the results and made critical revisions to the manuscript.

K. Stevens - formulated the idea for the study, oversaw the design and collection of data, interpreted the results and made critical revisions to the manuscript.

K. Dam Petersen - interpreted the results and made critical revisions to the manuscript.

J. Ratcliffe – formulated the idea for the study, oversaw the design and collection of data, interpreted the results and made critical revisions to the manuscript.

All authors approved the final draft.

Competing interests: CMK, GC, RR, KS, KDP and JR declare that they have no conflict of interest.

Funding: This study was funded by an Australian NHMRC Project Grant (grant number 1021899 entitled 'Adolescent values for the economic evaluation of adolescent health care treatment and preventive programs'.)

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

Ethical approval for this study was obtained from the Social and Behavioural Research Ethics Committee, Flinders University (project number 5508).

References

1. Brazier J, Ratcliffe J, Tsuchiya A, Salomon J. Measuring and Valuing Health Benefits for Economic Evaluation. Oxford: Oxford University Press; 2007.

2. Fontaine KR, Barofsky I. Obesity and health-related quality of life. Obes Rev. 2001;2(3):173-82.

3. Drummond MF, Sculpher MJ, Torrance GW, O'Brien BJ, Stoddart GL. Methods for the Economic Evaluation of Health Care Programmes. Oxford: Oxford University Press; 2005.

4. Harris A, Bulfone L. Getting value for money: "The Australian experience". . In: International M-H, editor. Health Care Coverage Determinations: An International Comparative Study. Jost, Timothy S ed. Maidenhead: Open University Press; 2004.

5. National Institute for Health and Care Excellence. Guide to the Methods of Technology Appraisal. National Health Service. 2010.

6. Varni JW, Burwinkle TM, Seid M, Skarr D. The PedsQL 4.0 as a pediatric population health measure: feasibility, reliability, and validity. Ambul Pediatr. 2003;3(6):329-41.

7. Brazier JE, Yang Y, Tsuchiya A, Rowen DL. 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.

8. Chen G, Stevens K, Rowen D, Ratcliffe J. From KIDSCREEN-10 to CHU9D: creating a unique mapping algorithm for application in economic evaluation. Health Qual Life Outcomes. 2014;12:134. doi:10.1186/s12955-014-0134-z.

9. Furber G, Segal L, Leach M, Cocks J. Mapping scores from the Strengths and Difficulties Questionnaire (SDQ) to preference-based utility values. Qual Life Res. 2014;23(2):403-11. doi:10.1007/s11136-013-0494-6.

10. Khan KA, Petrou S, Rivero-Arias O, Walters SJ, Boyle SE. Mapping EQ-5D utility scores from the PedsQL generic core scales. Pharmacoeconomics. 2014;32(7):693-706. doi:10.1007/s40273-014-0153-y.

11. Payakachat N, Tilford JM, Kuhlthau KA, van Exel NJ, Kovacs E, Bellando J et al. Predicting health utilities for children with autism spectrum disorders. Autism Res. 2014;7(6):649-63. doi:10.1002/aur.1409.

12. Varni JW, Seid M, Kurtin PS. PedsQL 4.0: reliability and validity of the Pediatric Quality of Life Inventory version 4.0 generic core scales in healthy and patient populations. Med Care. 2001;39(8):800-12.

13. Stevens K. Assessing the performance of a new generic measure of health-related quality of life for children and refining it for use in health state valuation. Appl Health Econ Health Policy. 2011;9(3):157-69. doi:10.2165/11587350-000000000-00000.

14. Stevens K. Valuation of the Child Health Utility 9D Index. Pharmacoeconomics. 2012;30(8):729-47. doi:10.2165/11599120-000000000-00000.

15. 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. doi:10.2165/11597900-000000000-00000.

16. Ratcliffe J, Huynh E, Chen G, Stevens K, Swait J, Brazier J et al. Valuing the Child Health Utility 9D: Using profile case best worst scaling methods to develop a new adolescent specific scoring algorithm. Social Science & Medicine. 2016;157:48-59. doi:http://dx.doi.org/10.1016/j.socscimed.2016.03.042.

17. Boyce W, Torsheim T, Currie C, Zambon A. The Family Affluence Scale as a measure of national wealth: validation of an adolescent self-report measure. Social Indicator Research. 2006;78(3):473-87.

18. Stevens K, Ratcliffe J. Measuring and valuing health benefits for economic evaluation in adolescence: an assessment of the practicality and validity of the child health utility 9D in the Australian adolescent population. Value Health. 2012;15(8):1092-9. doi:10.1016/j.jval.2012.07.011.

19. Ratcliffe J, Stevens K, Flynn T, Brazier J, Sawyer M. An assessment of the construct validity of the CHU9D in the Australian adolescent general population. Qual Life Res. 2012;21(4):717-25. doi:10.1007/s11136-011-9971-y.

20. Chen G, Flynn T, Stevens K, Brazier J, Huynh E, Sawyer M et al. Assessing the Health-Related Quality of Life of Australian Adolescents: An Empirical Comparison of the Child Health Utility 9D and EQ-5D-Y Instruments. Value Health. 2015;18(4):432-8. doi:10.1016/j.jval.2015.02.014.

21. Petrou S, Rivero-Arias O, Dakin H, Longworth L, Oppe M, Froud R et al. The MAPS Reporting Statement for Studies Mapping onto Generic Preference-Based Outcome Measures: Explanation and Elaboration. PharmacoEconomics. 2015;33(10):993-1011. doi:10.1007/s40273-015-0312-9.

22. Boers M, Verhoeven AC, Markusse HM, van de Laar MA, Westhovens R, van Denderen JC et al. Randomised comparison of combined step-down prednisolone, methotrexate and sulphasalazine with sulphasalazine alone in early rheumatoid arthritis. Lancet. 1997;350(9074):309-18. doi:10.1016/s0140-6736(97)01300-7.

23. International CLL-IPI working group. An international prognostic index for patients with chronic lymphocytic leukaemia (CLL-IPI): a meta-analysis of individual patient data. Lancet Oncol. 2016;17(6):779-90. doi:10.1016/s1470-2045(16)30029-8.

24. Chappell LC, Seed PT, Myers J, Taylor RS, Kenny LC, Dekker GA et al. Exploration and confirmation of factors associated with uncomplicated pregnancy in nulliparous women: prospective cohort study. Bmj. 2013;347:f6398. doi:10.1136/bmj.f6398.

25. Kuk D, Varadhan R. Model selection in competing risks regression. Statistics in Medicine. 2013;32(18):3077-88. doi:10.1002/sim.5762.

26. Allen LA, Yager JE, Funk MJ, Levy WC, Tulsky JA, Bowers MT et al. Discordance between patient-predicted and model-predicted life expectancy among ambulatory heart failure patients. JAMA : the journal of the American Medical Association. 2008;299(21):2533-42. doi:10.1001/jama.299.21.2533.

27. StataCorp. Stata Statistical Software: Release 14. 2015.

28. Dakin H. Review of studies mapping from quality of life or clinical measures to EQ-5D: an online database. Health Qual Life Outcomes. 2013;11:151. doi:10.1186/1477-7525-11-151.

29. Longworth L, Rowen D. Mapping to obtain EQ-5D utility values for use in NICE health technology assessments. Value Health. 2013;16(1):202-10. doi:10.1016/j.jval.2012.10.010.

30. Hernandez Alava M, Wailoo A, Wolfe F, Michaud K. A comparison of direct and indirect methods for the estimation of health utilities from clinical outcomes. Med Decis Making. 2014;34(7):919-30. doi:10.1177/0272989x13500720.

31. Huang IC, Frangakis C, Atkinson MJ, Willke RJ, Leite WL, Vogel WB et al. Addressing ceiling effects in health status measures: a comparison of techniques applied to measures for people with HIV disease. Health Serv Res. 2008;43(1 Pt 1):327-39. doi:10.1111/j.1475-6773.2007.00745.x.

32. Payakachat N, Summers KH, Pleil AM, Murawski MM, Thomas J, 3rd, Jennings K et al. Predicting EQ-5D utility scores from the 25-item National Eye Institute Vision Function Questionnaire (NEI-VFQ 25) in patients with age-related macular degeneration. Qual Life Res. 2009;18(7):801-13. doi:10.1007/s11136-009-9499-6.

33. Gujarati DN. Basic Econometrics. McGraw-Hill; 2003.

34. Chen G, Khan MA, Iezzi A, Ratcliffe J, Richardson J. Mapping between 6 Multiattribute Utility Instruments. Med Decis Making. 2016;36(2):160-75. doi:10.1177/0272989x15578127.

35. McCullagh P, Nelder JA. Generalized linear models. 2nd ed. London: Chapman & Hall; 1989.

36. Royston P, Sauerbrei W. Multivariable modeling with cubic regression splines: A principled approach. The Stata Journal. 2007;7(1):45-70.

37. Briggs A, Sculpher M, Claxton K. Decision modelling for health economic evaluation. OUP Oxford; 2006.

38. Ospina R, Ferrari SL. A general class of zero-or-one inflated beta regression models. Computational Statistics & Data Analysis. 2012;56(6):1609-23.

39. Basu A, Manca A. Regression estimators for generic health-related quality of life and quality-adjusted life years. Med Decis Making. 2012;32(1):56-69. doi:10.1177/0272989x11416988.

40. Khan I, Morris S. A non-linear beta-binomial regression model for mapping EORTC QLQ- C30 to the EQ-5D-3L in lung cancer patients: a comparison with existing approaches. Health Qual Life Outcomes. 2014;12(1):1-16. doi:10.1186/s12955-014-0163-7.

41. Everitt B, Hand D. Finite mixture distributions. London and New York: Chapman and Hall; 1981.

42. McLachlan G, Peel D. Finite Mixture Models. New York: Wiley; 2000.

43. Kent S, Gray A, Schlackow I, Jenkinson C, McIntosh E. Mapping from the Parkinson's Disease Questionnaire PDQ-39 to the Generic EuroQol EQ-5D-3L: The Value of Mixture Models. Med Decis Making. 2015;35(7):902-11. doi:10.1177/0272989x15584921.

44. Deb P. Finite Mixture Models. 2008. http://repec.org/snasug08/deb_fmm_slides.pdf. Accessed 11 September 2016.

45. 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.

46. Le QA, Doctor JN. 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. doi:10.1097/MLR.0b013e318207e9a8.

47. Koch G. “Intraclass correlation coefficient.”. In: Johnson SKaNL, editor. Encyclopedia of Statistical Sciences, Ed. New York: John Wiley & Sons, Inc; 1983. p. pp. 213–7.

48. Hyndman RJ, Koehler AB. Another look at measures of forecast accuracy. International Journal of Forecasting. 2006;22(4):679-88. doi:http://dx.doi.org/10.1016/j.ijforecast.2006.03.001.

49. Shcherbakov MV, Brebels B, Shcherbakova NL, Tyukov AP, Janovsky TA, Kamae VA. A Survey of Forecast Error Measures. World Applied Sciences Journal 24 (Information Technologies in Modern Industry, Education & Society). 2013;24(24):171-6.

50. Wong CK, Lam CL, Rowen D, McGhee SM, Ma KP, Law WL et al. Mapping the Functional Assessment of Cancer Therapy-general or -Colorectal to SF-6D in Chinese patients with colorectal neoplasm. Value Health. 2012;15(3):495-503. doi:10.1016/j.jval.2011.12.009.

51. Wu EQ, Mulani P, Farrell MH, Sleep D. Mapping FACT-P and EORTC QLQ-C30 to patient health status measured by EQ-5D in metastatic hormone-refractory prostate cancer patients. Value Health. 2007;10(5):408-14. doi:10.1111/j.1524-4733.2007.00195.x.

52. Petrou S, Rivero-Arias O, Dakin H, Longworth L, Oppe M, Froud R et al. Preferred reporting items for studies mapping onto preference-based outcome measures: the MAPS statement. Qual Life Res. 2016;25(2):275-81. doi:10.1007/s11136-015-1082-8.

53. Chuang LH, Whitehead SJ. Mapping for economic evaluation. Br Med Bull. 2012;101:1-15. doi:10.1093/bmb/ldr049.

54. Pinedo-Villanueva RA, Turner D, Judge A, Raftery JP, Arden NK. Mapping the Oxford hip score onto the EQ-5D utility index. Qual Life Res. 2013;22(3):665-75. doi:10.1007/s11136-012-0174-y.

55. Tsuchiya A, Brazier JE, McColl E, Parkin D. Deriving preference-based single indices from non-preference based condition-specific instruments: Converting AQLQ into EQ5D indices Sheffield Health Economics Group Discussion Paper Series. 2002;Ref 02/1.

56. Brennan DS, Spencer AJ. Mapping oral health related quality of life to generic health state values. BMC Health Serv Res. 2006;6:96. doi:10.1186/1472-6963-6-96.

57. Sauerland S, Weiner S, Dolezalova K, Angrisani L, Noguera CM, Garcia-Caballero M et al. Mapping utility scores from a disease-specific quality-of-life measure in bariatric surgery patients. Value Health. 2009;12(2):364-70. doi:10.1111/j.1524-4733.2008.00442.x.

58. Bansback N, Marra C, Tsuchiya A, Anis A, Guh D, Hammond T et al. Using the health assessment questionnaire to estimate preference-based single indices in patients with rheumatoid arthritis. Arthritis Rheum. 2007;57(6):963-71. doi:10.1002/art.22885.

59. Australian Demographic Statistics, 2015, 'Table 8: Estimated resident population, by age and sex – at 30 June 2015', data cube: Excel spreadsheet, cat. no. 31010do002_201512 [database on the Internet]. Australian Bureau of Statistics 2015. Available from:

http://www.abs.gov.au/AUSSTATS/abs@.nsf/DetailsPage/3101.0Dec%202015?OpenDocument. Accessed: 6 September 2016

60. Australian Institute of Health and Welfare. Young Australians: Their Health and Wellbeing 2011 (Cat. no. PHE 140). Canberra, Australia: Australian Institute of Health and Welfare; 2011.

Table 1 Participant characteristics (n = 755)

Characteristic	
CHU9D utility score	
Mean (SD)	0.724 (0.218)
Median, (IQR)	0.769 (0.587 – 0.888)
PedsQL score	
Mean (SD)	72.861 (16.562)
Median, (IQR)	75.00 (63.333 - 85.000)
Age (year)	
Mean (SD)	15.8 (0.8)
Median (IQR)	16.0 (15.0-16.0)
Age band N (%)	
15	299 (39.6)
16	286 (37.9)
17	170 (22.5)
Gender	
Males	358 (47.4)
Females	397 (52.6)
Family Affluence Scale, N (%)	
High (scores 6-9)	447 (59.2)
Medium (scores 4-5)	267 (35.4)
Low (scores 0-3)	41 (5.4)
Self –reported health, N (%)	
Excellent	236 (31.3)
Very good	339 (44.9)
Good	138 (18.3)
Fair	36 (4.8)
Poor	6 (0.8)
Long term disability, illness or medical condition, N (%)	
Yes	76 (10.1)
No	679 (89.9)

CHU9D Child Health Utility 9D; PedsQL Pediatric Quality of Life Inventory; SD Standard Deviation; IQR Inter Quartile Range.

Table 2 Goodness of fit results from full estimation sample (n =755)

Model specification	Mean (SD)	Min	Max	MAE	MSE	Abs diff. < 0.03 (%)	Abs diff. < 0.05 (%)
Observed	0.724 (0.218)	-0.046	1.000	-	-		
Model (1)							
OLS	0.724 (0.142)	0.101	0.956	0.1302	**0.0272**	15	26
GLM	0.724 (0.142)	0.190	0.953	0.1302	**0.0272**	15	25
MM	0.730 (0.162)	-0.210	0.954	**0.1296**	0.0281	14	26
MFP	0.724 (0.142)	0.190	0.953	0.1302	**0.0272**	15	25
BB	0.728 (0.146)	0.097	0.912	0.1301	0.0273	13	24
FMM	0.730 (0.127)	0.228	0.947	0.1317	0.0275	14	25
MLOGIT	0.726 (0.141)	0.164	0.928	0.1303	**0.0272**	15	24
Model (2)							
OLS	0.724 (0.146)	0.017	0.953	0.1271	0.0260	15	26
GLM	0.724 (0.148)	0.182	0.955	0.1270	0.0259	14	26
MM	0.733 (0.160)	-0.136	0.957	**0.1261**	0.0264	16	26
MFP	0.724 (0.148)	0.182	0.955	0.1270	0.0259	14	26
BB	0.728 (0.148)	0.110	0.924	0.1268	0.0260	14	24
FMM	0.730 (0.130)	0.151	0.947	0.1289	0.0263	15	24
MLOGIT	0.726 (0.146)	0.119	0.924	0.1269	**0.0258**	14	24
Model (3)							
OLS	0.724 (0.148)	0.113	0.952	0.1262	**0.0256**	14	26
GLM	0.725 (0.153)	0.244	0.972	0.1272	0.0261	17	26
MM	0.734 (0.161)	0.025	0.954	**0.1251**	0.0263	17	29
MFP	0.725 (0.153)	0.244	0.972	0.1272	0.0261	17	26
BB	0.728 (0.150)	0.136	0.920	0.1263	0.0259	15	24
FMM	0.729 (0.130)	0.227	0.929	0.1289	0.0263	13	24
MLOGIT	0.726 (0.146)	0.710	0.909	0.1271	0.0260	15	23

Dependent variable: CHU9D utility score. Best results in each model are in bold type.

Independent variable(s): Model (1) PedsQL Total score and squared terms; Model (2) PedsQL Dimension Scores and squared terms; (3) PedsQL items scores and squared terms.

CHU9D Child Health Utility 9D; BB beta-binomial; FMM Finite Mixture Models; GLM generalised linear model; MAE mean absolute error; MFP multivariate factorial polynomials; MM Robust MM-estimator; MSE mean squared error; MLOGIT Multinomial logistic regression; Min Minimum; Max Maximum; PedsQL Pediatric Quality of Life Inventory; SD standard deviation; Abs diff. < 0.03 (0.05) % proportion of predicted utilities whose absolute values deviate from the mean of the observed utility values by less than 0.03 (0.05).

Table 3 Goodness-of-fit results from validation analyses

	Validation Method 1 (k-fold)					Validation Method 2 (Three random samples)														
	Pooled sample (n=755)					Random sample 1 (N = 100)					Random sample 2 (N = 300)					Random sample 3 (N = 500)				
	Mean utility	MSE	MAE	Abs diff. <0.03	Abs diff. <0.05	Mean utility	MSE	MAE	Abs diff. <0.03	Abs diff. <0.05	Mean utility	MSE	MAE	<0.03	<0.05	Mean utility	MSE	MAE	Abs diff. <0.03	Abs diff. <0.05
Observed	0.724	-	-			0.734	-	-			0.720	-	-			0.734	-	-		
Model 1																				
OLS	**0.724**	**0.0274**	0.1305	15	26	0.726	0.0233	0.1228	18	27	0.715	0.0293	0.1345	13	26	0.727	0.0256	**0.1250**	15	27
GLM	**0.724**	**0.0274**	0.1305	16	25	0.726	0.0232	0.1225	20	26	0.715	**0.0292**	0.1346	13	25	0.727	**0.0255**	**0.1250**	16	27
MM	0.730	0.0283	**0.1303**	14	25	**0.734**	**0.0224**	**0.1185**	18	32	**0.720**	0.0311	0.1348	15	25	**0.733**	0.0268	0.1253	14	27
MFP	0.725	0.0279	0.1318	17	25	0.726	0.0232	0.1225	20	26	0.715	**0.0292**	0.1346	13	25	0.727	**0.0255**	**0.1250**	16	27
BB	0.728	0.0275	0.1304	14	23	0.730	0.0228	0.1213	16	25	0.719	0.0293	**0.1336**	13	24	0.731	0.0258	0.1257	14	24
FMM	0.729	0.0276	0.1319	14	25	0.732	0.0243	0.1255	20	27	0.722	0.0290	0.1347	13	25	0.732	0.0257	0.1262	14	25
MLOGIT	0.720	0.0292	0.1366	14	23	0.728	0.0232	0.1223	19	27	0.717	0.0291	0.1342	13	24	0.729	**0.0255**	0.1252	16	24
Model 2																				
OLS	**0.724**	0.0263	0.1277	15	25	0.724	0.0220	0.1205	18	29	0.715	0.0276	0.1303	13	26	0.727	0.0246	0.1226	14	27
GLM	**0.724**	**0.0262**	0.1275	16	24	0.726	0.0221	0.1204	17	26	0.715	0.0276	0.1304	13	25	0.727	0.0246	0.1225	15	27
MM	0.732	0.0266	**0.1270**	15	25	**0.734**	0.0213	**0.1169**	21	30	0.723	0.0285	0.1297	15	26	**0.735**	0.0252	**0.1220**	15	27
MFP	**0.724**	**0.0262**	0.1275	16	24	0.726	0.0221	0.1204	17	26	0.715	0.0276	0.1304	13	25	0.727	0.0246	0.1225	15	27
BB	0.728	**0.0262**	0.1271	15	24	0.728	**0.0212**	0.1181	17	28	**0.720**	**0.0274**	**0.1293**	13	23	0.732	0.0245	0.1225	14	24
FMM	0.729	0.0265	0.1291	14	25	0.729	0.0230	0.1237	19	27	0.723	0.0276	0.1317	14	25	0.733	0.0247	0.1237	14	25
MLOGIT	**0.717**	0.0274	0.1327	13	22	0.724	0.0215	0.1196	15	27	0.718	0.0272	0.1295	14	24	0.729	**0.0243**	0.1228	15	24
Model 3																				
OLS	**0.724**	**0.0262**	0.1278	14	24	0.719	0.0224	0.1235	17	26	0.717	0.0273	0.1294	15	25	0.729	0.0245	0.1227	13	27
GLM	**0.724**	0.0268	0.1291	14	25	0.718	0.0234	0.1252	18	27	0.718	0.0284	0.1317	17	24	0.730	0.0253	0.1241	16	31
MM	0.733	0.0270	**0.1271**	16	28	**0.728**	0.0225	0.1214	18	30	0.727	0.0285	0.1290	18	29	0.739	0.0249	**0.1209**	16	31
MFP	**0.724**	0.0268	0.1291	14	25	0.718	0.0234	0.1252	18	27	0.718	0.0284	0.1317	17	24	0.730	0.0253	0.1241	19	26
BB	0.728	0.0264	0.1274	15	24	0.724	**0.0219**	**0.1205**	16	26	**0.721**	**0.0270**	**0.1276**	16	25	**0.733**	**0.0242**	0.1217	14	24
FMM	0.728	0.0267	0.1296	13	23	0.722	0.0241	0.1291	13	24	0.723	0.0275	0.1314	14	24	0.733	0.0250	0.1249	14	24
MLOGIT	0.721	0.0274	0.1310	14	22	0.721	0.0226	0.1242	15	21	0.719	0.0273	0.1292	16	24	0.729	0.0250	0.1241	15	23

Dependent variable: CHU9D utility score. Best results in each model are in bold type.

Independent variable (s): Model (1) PedsQL Total score and PedsQL Total score squared terms; Model (2) PedsQL Dimension Scores and PedsQL Dimension Scores squared terms; (3) PedsQL items scores and PedsQL items scores squared terms.

CHU9D Child Health Utility 9D; BB beta-binomial; FMM Finite Mixture Models; GLM generalised linear model; MAE mean absolute error; MFP multivariate factorial polynomials; MM Robust MM-estimator; MSE mean squared error; MLOGIT Multinomial logistic regression; Min Minimum; Max Maximum; PedsQL Pediatric Quality of Life Inventory; Abs diff. < 0.03 (0.05) % proportion of predicted utilities whose absolute values deviate from the mean of the observed utility values by less than 0.03 (0.05).

Table 4 Model performance for best fitting models

Estimation method	Observed CHU9D Mean (SD)	Mean (SD)	Min	P.25	Median	P.75	Max	MAE	MSE	Abs diff. < 0.03 (%)	Abs diff < 0.05 (%)	ICC
Model 1 (MM-estimator)												
Validation I (pooled sample)	0.724 (0.218)	0.730 (0.162)	0.060	0.643	0.758	0.853	0.954	0.1303	0.0283	14	25	0.765
Validation II (random samples)												
Random Sample 1	0.734 (0.224)	0.734 (0.156)	0.134	0.636	0.765	0.855	0.954	0.1185	0.0224	18	32	0.822
Random Sample 2	0.720 (0.214)	0.720 (0.171)	-0.210	0.636	0.750	0.848	0.954	0.1348	0.0311	15	25	0.737
Random Sample 3	0.734 (0.212)	0.733 (0.165)	-0.210	0.653	0.765	0.848	0.954	0.1253	0.0268	14	27	0.772
Model 2 (MM-estimator)												
Validation I (pooled sample)	0.724 (0.218)	0.732 (0.160)	0.093	0.647	0.749	0.848	0.956	0.1270	0.0266	15	25	0.779
Validation II (random samples)												
Random Sample 1	0.734 (0.224)	0.734 (0.153)	0.175	0.627	0.741	0.854	0.957	**0.1169**	**0.0213**	**21**	30	**0.830**
Random Sample 2	0.720 (0.214)	0.723 (0.168)	-0.136	0.635	0.741	0.840	0.957	0.1297	0.0285	15	26	0.759
Random Sample 3	0.734 (0.212)	0.735 (0.162)	-0.136	0.651	0.754	0.852	0.957	0.1220	0.0252	15	27	0.785
Model 3 (BB-estimator)												
Validation I (pooled sample)	0.724 (0.218)	0.728 (0.151)	0.231	0.643	0.756	0.842	0.920	0.1274	0.0264	15	24	0.772
Validation II (random samples)												
Random Sample 1	0.734 (0.224)	0.724 (0.147)	0.273	0.616	0.751	0.845	0.920	0.1205	0.0219	16	26	0.819
Random Sample 2	0.720 (0.214)	0.721 (0.155)	0.136	0.635	0.747	0.842	0.920	0.1276	0.0270	16	25	0.758
Random Sample 3	0.734 (0.212)	0.732 (0.150)	0.136	0.652	0.760	0.847	0.920	0.1217	0.0242	14	24	0.782

Dependent variable = CHU9D utility score. Best results are in bold type.

Independent variable (s): Model (1) PedsQL Total score and PedsQL Total score squared terms; Model (2) PedsQL Dimension Scores and PedsQL Dimension Scores squared terms; (3) PedsQL items scores and PedsQL items scores squared terms.

BB beta-binomial; MM robust MM-estimator; MAE mean absolute error; MSE mean squared error; Min Minimum; Max Maximum; PedsQL Pediatric Quality of Life Inventory; SD Standard deviation; Min minimum; P.25 25th percentile; P.75 75th percentile; Max maximum; MSE mean squared error; MAE mean absolute error; Abs diff. < 0.03 (0.05) % proportion of predicted utilities whose absolute values deviate from the mean of the observed utility values by less than 0.03 (0.05).

Table 5 Mapping equations from PedsQL to Child Health Utility 9D utility scores[a]

Independent variables	MM Model 1 [MM (1)]	MM Model 2 [MM (1)]	BB Model 3 [BB (3)]
PedsQL Total	1.707043***		
	(0.405060)		
PedsQL Total Square	-0.543056**		
	(0.265408)		
PedsQL Dimen PF		0.264648***	
		(0.055167)	
PedsQL Dimen EF		1.196678***	
		(0.183097)	
PedsQL Dimen SF		0.203405***	
		(0.036426)	
PedsQL Dimen EF Square		-0.572612***	
		(0.124107)	
PedsQL item EF2			1.053955***
			(0.222367)
PedsQL item EF4			0.991873***
			(0.173220)
PedsQL item SchF2			0.590129***
			(0.195802)
PedsQL item SchF3			0.772430***
			(0.175686)
PedsQL item PF3			0.455649***
			(0.155322)
PedsQL item PF1 Square			0.436041**
			(0.174260)
Constant	-0.210178	-0.135516**	-1.852265***
	(0.150083)	(0.060300)	(0.173271)

Standard errors in parentheses. *** p<0.01, p<**0.05

BB beta-binomial; MM robust MM-estimator

PedsQL Pediatric Quality of Life Inventory; PedsQL Total, PedsQL total score; PedsQL Total Square, PedsQL total score squared; PF Physical functioning; EF Emotional functioning; SF Social functioning; EF2 "Sad or blue", EF4 "Worry about what will happen to me"; SchF2 "Forget things"; SchF3 "Have trouble keeping up with my schoolwork"; PF*1* "it is hard for me to walk for more than one block"; PF3 "hard for me to do sports activity or exercise"; item, PedsQl item scores; Dimen PedsQL dimension scores.

[a] All PedsQL scores included in the regression were rescaled onto the 0-1 scale by dividing raw scores by 100.

Table 6 Distribution of errors according to selected ranges in the CHU9D utility score

		MM Model 1 [MM (1)]		MM Model 2 [MM (1)]		BB Model 3 [BB (3)]	
		MSE	MAE	MSE	MAE	MSE	MAE
CHU9D utility score range	n						
0 < CHU9D ≤ 0.2	16	0.0839	0.2532	0.0862	0.2534	0.0901	0.2708
0.2 < CHU9D ≤ 0.4	55	0.0615	0.2228	0.0647	0.2278	0.0648	0.2230
0.4 < CHU9D ≤ 0.6	123	0.0451	0.1820	0.0407	0.1731	0.0395	0.1689
0.6 < CHU9D ≤ 0.8	232	0.0134	0.0927	0.0135	0.0927	0.0129	0.0923
0.8 < CHU9D ≤ 1.0	329	0.0237	0.1144	0.0209	0.1090	0.0204	0.1111

MAE mean absolute error; MSE mean squared error

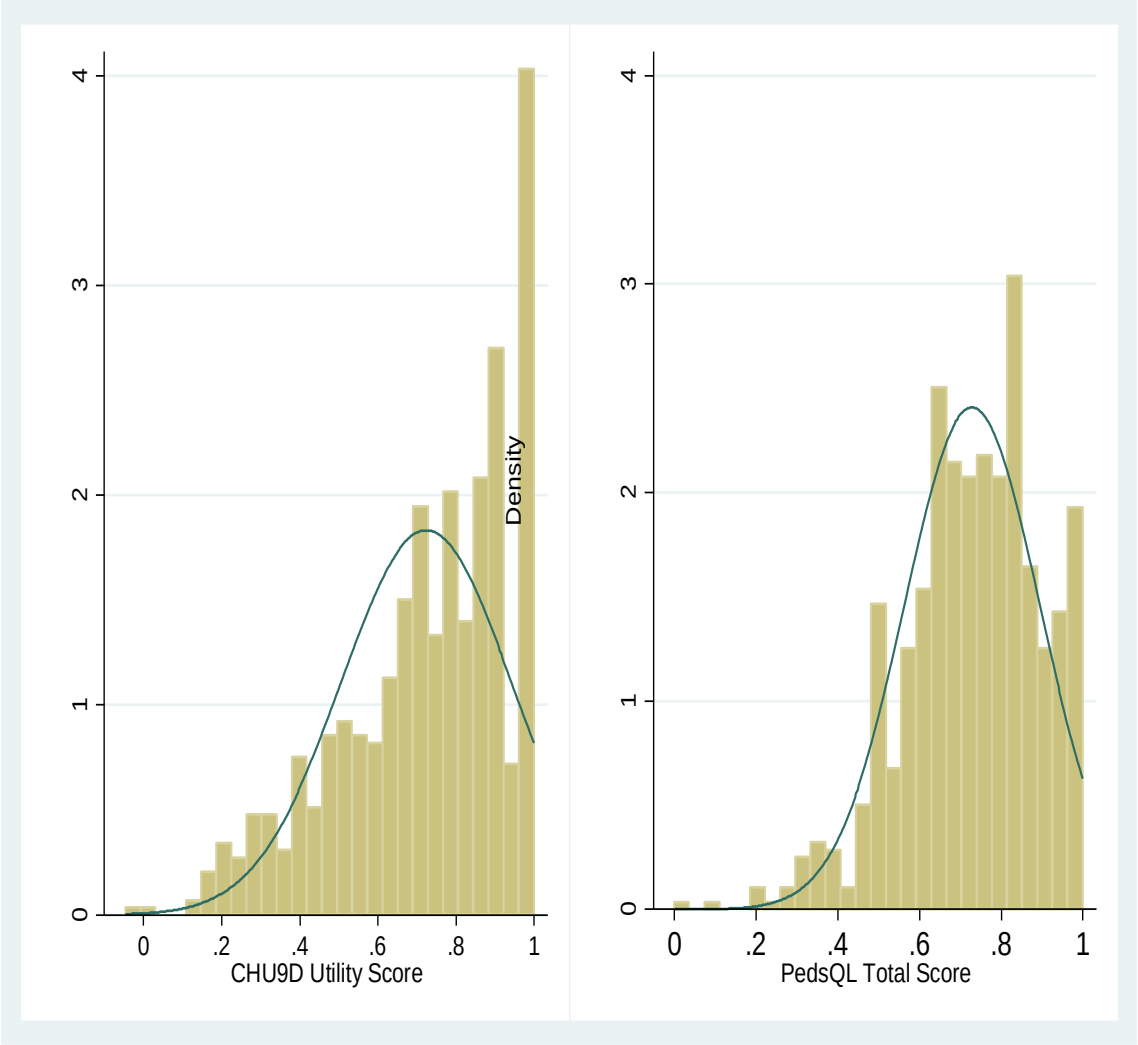

Fig 1: Distribution of CHU9D utility scores and PedsQL total score

Figure 1 presents Kernel density plots of the Pediatric Quality of Life Inventory (PedsQL) total scores and the Child Health Utility 9 Dimension (CHU9D) utilities

Note: For comparability between the two instruments the PedsQL total scores were rescaled onto the 0-1 scale by dividing raw scores by 100

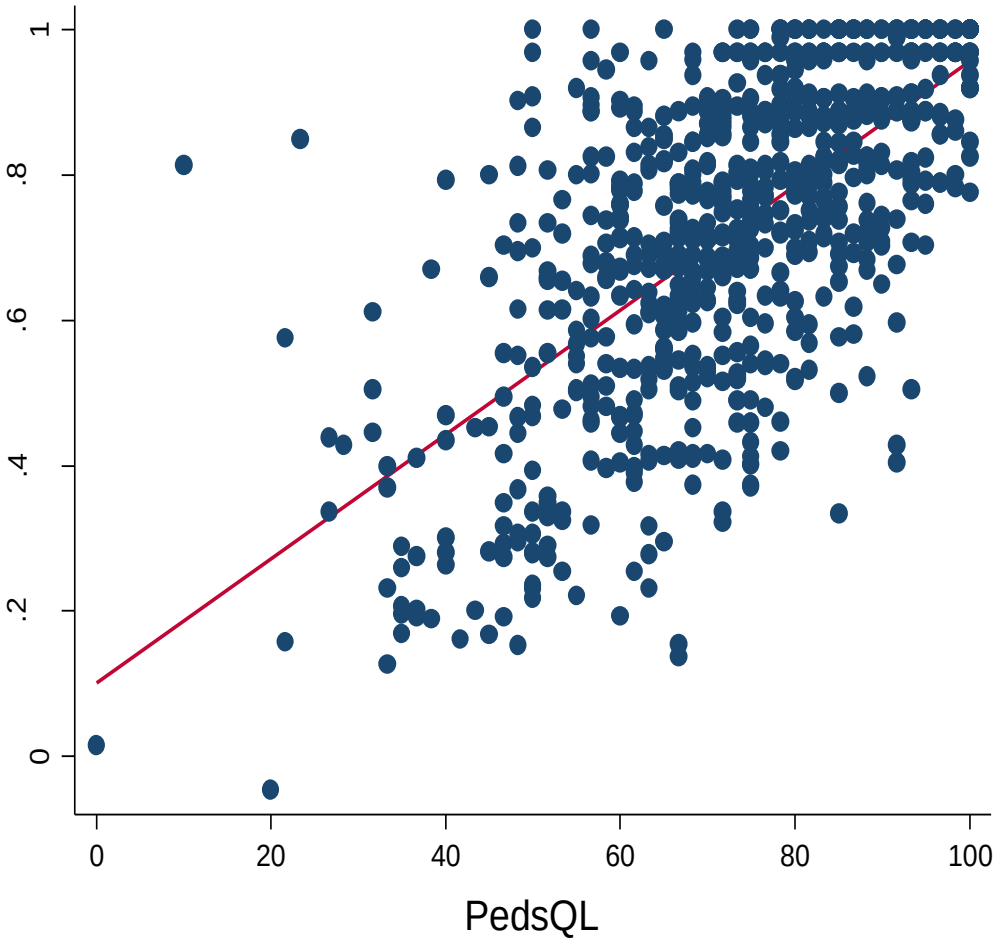

Fig 2: Scatter plots between Instruments

Figure 2 presents scatter plots between the Pediatric Quality of Life Inventory (PedsQL) total scores and the Child Health Utility 9 Dimension (CHU9D) utilities.

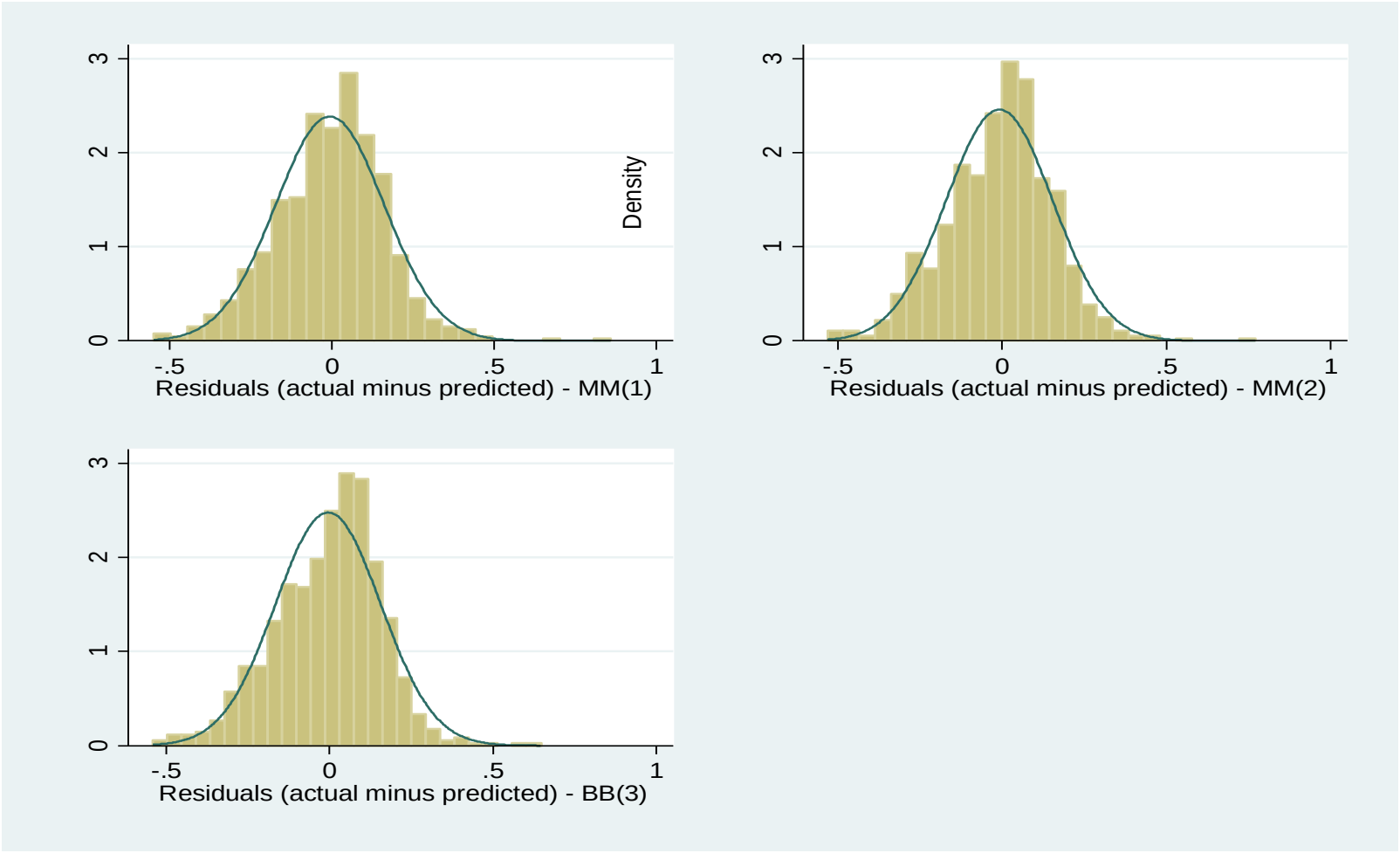

Fig 3: Plot of residuals (actual minus predicted) for the three best fitting models Figure 3 presents plots of residuals for the three best fitting models (robust MM estimator-[model 1], robust MM estimator [model 2] and Beta-binomial [model 3]

SUPPLEMENTARY DOCUMENTS

Supplementary Table 1: Comparability of PedsQL and CHU9D dimensions[a]

PedsQL Dimensions	**Comparable CHU9D dimensions**
Physical Health	Tired, Daily routine, Activities
Psychosocial Health	Worried, Sad, Annoyed, Activities, Daily routine
Emotional function	Worried, Sad, Annoyed
Social function	-
School function	School work/homework, Daily routine

[a] This qualitative assessment is based on discussions within the team

Mapping Algorithm for mapping CHU9D utility scores from the PedsQL™ 4.0 SF-15

****

** This program generates CHU9D utility scores from PedsQL™ 4.0 SF-15 scores based on the Robust MM-estimator (i.e. MM Model 2 [MM2] in Table 5) through Stata Software

** Note: 'PedsQL_Dimen_PF_raw' is the raw score of 'PedsQL Physical Functioning', 'PedsQL_Dimen_EF_raw' is the raw score of 'PedsQL Emotional Functioning', 'PedsQL_Dimen_SF_raw' is the raw score of 'PedsQL School Functioning'; all raw scores had a range of 0-100.

****

```
gen PedsQL_Dimen_PF = PedsQL_Dimen_PF_raw/100

gen PedsQL_Dimen_EF = PedsQL_Dimen_EF_raw/100

gen PedsQL_Dimen_SF = PedsQL_Dimen_SF_raw/100

gen PedsQL_Dimen_EF2 = PedsQL_Dimen_EF^2


gen CHU9D_MM2 = -0.135516 + 0.264648*PedsQL_Dimen_PF +
1.196678*PedsQL_Dimen_EF + 0.203405*PedsQL_Dimen_SF -
0.572612*PedsQL_Dimen_EF2


label var CHU9D_MM2 "CHU9D utility, predicted from PedsQL 4.0 SF-15
dimensions (robust MM-estimator Model 2)"


sum CHU9D_MM2


drop PedsQL_Dimen_PF- PedsQL_Dimen_EF2
```

**********The End**********

Variance Covariance Matrix from Mapping Algorithm

	PedsQL_Dimen_PF	PedsQL_Dimen_EF	PedsQL_Dimen_SF	PedsQL_Dimen_EF2	constant
PedsQL_Dimen_PF	0.003043				
PedsQL_Dimen_EF	-0.003696	0.033524			
PedsQL_Dimen_SF	-0.000222	-0.000769	0.001327		
PedsQL_Dimen_EF2	0.002070	-0.022034	-0.000212	0.015403	
constant	-0.001037	-0.007931	-0.000035	0.005452	0.003636