Mapping the Paediatric Quality of Life Inventory (PedsQL™) Generic Core Scales onto the Child Health Utility Index–9 Dimension (CHU-9D) Score for Economic Evaluation in Children

Tosin Lambe[1] • Emma Frew[1] • Natalie J. Ives[3] • Rebecca L. Woolley[3] • Carole Cummins[4] • Elizabeth A. Brettell[3] • Emma N. Barsoum[3] • Nicholas J. A. Webb[2] • On behalf of the PREDNOS Trial Team

1 Health Economics Unit, Institute of Applied Health Research, University of Birmingham, Edgbaston, Birmingham, UK

2 Department of Paediatric Nephrology and NIHR Manchester Clinical Research Facility, Manchester Academic Health Science Centre, Royal Manchester Children's Hospital, Manchester, UK

3 Birmingham Clinical Trials Unit, University of Birmingham, Birmingham, UK

4 Institute of Applied Health Research, University of Birmingham, Birmingham, UK

Abstract

Background The Paediatric Quality of Life Inventory (PedsQL™) questionnaire is a widely used, generic instrument designed for measuring health-related quality of life (HRQoL); however, it is not preference-based and therefore not suitable for cost–utility analysis. The Child Health Utility Index–9 Dimension (CHU-9D), however, is a preference-based instrument that has been primarily developed to support cost–utility analysis.

Objective This paper presents a method for estimating CHU-9D index scores from responses to the PedsQL™ using data from a randomised controlled trial of prednisolone therapy for treatment of childhood corticosteroidsensitive nephrotic syndrome.

Methods HRQoL data were collected from children at randomisation, week 16, and months 12, 18, 24, 36 and 48. Observations on children aged 5 years and older were pooled across all data collection timepoints and were then randomised into an estimation (n = 279) and validation (n = 284) sample. A number of models were developed using the estimation data before internal validation. The best model was chosen using multi-stage selection criteria. Results Most of the models developed accurately predicted the CHU-9D mean index score. The best performing model was a generalised linear model (mean absolute error = 0.0408; mean square error = 0.0035). The proportion of index scores deviating from the observed scores by< 0.03 was 53%.

Conclusions The mapping algorithm provides an empirical tool for estimating CHU-9D index scores and for conducting cost–utility analyses within clinical studies that have only collected PedsQL™ data. It is valid for children aged 5 years or older. Caution should be exercised when using this with children younger than 5 years, older adolescents (> 13 years) or patient groups with particularly poor quality of life.

Key Points

The Paediatric Quality of Life Inventory (PedsQL^{TM}) is a widely used tool/questionnaire for measuring health-related quality of life in children and adolescents but it is unsuitable for calculating quality-adjusted life-years (QALYs), which are often required for use in economic evaluation.

The algorithm produced in this study now permits the estimation of CHU-9D scores from PedsQL^{TM} responses in children as young as 5 years

1.Introduction

Cost-effectiveness analysis is a comparative assessment of both costs and outcomes linked to healthcare interventions. Evidence of healthcare benefits are often synthesised from clinical trials, public health studies or from other forms of health research. These health benefits are increasingly captured as health-related quality of life (HRQoL) using either 'condition-specific' or 'generic' survey instruments. Condition-specific instruments focus on health dimensions relevant to a particular disease whereas generic instruments assess core dimensions of health that are relevant to all conditions [1]. Clinical trials tend to use condition-specific instruments as a primary or secondary measure of benefit because these instruments focus on the specific domains of quality of life affected by a condition and are therefore sensitive to treatment effect in these domains. On the other hand, generic instruments measure a broader HRQoL construct [2]; therefore, they allow comparisons of treatment benefit across a wide range of interventions across multiple conditions.

Generic instruments can be further classed as either 'preference' or 'non-preference-based'. Preference-based generic instruments attach weights to the domains of health to reflect a stronger preference for one aspect of HRQoL over another, in order to generate a single index/score of HRQoL (also termed a utility score) [3]. In contrast, most non-preference-based instruments simply sum the scores from all the health domains and thus assume an equal weighting. Some non-preference-based instruments may apply complex weighting systems, but these weights are not preference-based [4]. For cost–utility analysis, preference based generic instruments are required to measure utility-based quality-of-life scores and, unfortunately, the majority of generic instruments used in clinical trials are non-preference-based [5] and are consequently of limited use for measuring and comparing the cost effectiveness of diverse interventions on a common scale [6, 7].

To capture both length and quality of life from treatment, quality-adjusted life-years (QALYs) are often applied [8, 9], whereby cost effectiveness is expressed as cost per additional QALY gained. Within paediatric medicine, however, most HRQoL instruments developed for children and adolescents are non-preference-based [10] and therefore cannot be used for economic evaluation [11] where QALYs are the desired outcome. A prediction algorithm/mapping function can, however, be used to predict utility scores from responses to a non-preferencebased instrument [5]. This algorithm reflects the statistical relationship between the preference and the non-preference-based instrument, using

responses from a prior population whose responses to both instruments have been collected.

The Paediatric Quality of Life Inventory (PedsQL^{TM}) is a generic non-preference-based instrument that provides a modular approach for measuring HRQoL in healthy children and adolescents and those with acute or chronic health conditions. PedsQL^{TM} is commonly used due to its simple computational system and its validity for a wide age range of 2- to 18-year-olds [12]. In a review of paediatric qualityof-life measures, PedsQL^{TM} fulfilled basic psychometric criteria, was suitable for completion in the clinic and could be recommended for use in clinical trials [13]. It also has the advantage of having both proxy and patient-completed versions, and has additional modules measuring some disease-specific quality of life. A viable preference-based alternative to the PedsQL^{TM} is the Child Health Utility–9 Dimension (CHU-9D), which has been specifically developed for economic evaluation in children aged 5 years and older [14]. In situations where only PedsQL^{TM} data are available, CHU-9D utility scores can be predicted from the PedsQL^{TM} using a mapping algorithm. Only one study has mapped the PedsQL^{TM} onto CHU-9D [15], in which the Short Form 15-item (SF-15) version of the PedsQL^{TM} was used in place of the standard 23-item questionnaire. Data for that study [15] were obtained from Australian older adolescents only (15–17 years old), and the CHU-9D responses were scored using the Australian value set [16].

This study mapped responses from the 23-item generic core scale version of the PedsQL^{TM} onto CHU-9D index scores in children aged 5–13 years who were participants in a randomised controlled trial (RCT) of different corticosteroid regimens in childhood corticosteroid-sensitive nephrotic syndrome [17

2 Methods

2.1 Data

The data for this study were obtained from the PREDNOS (PREDnisolone in NephrOtic Syndrome) study, a UK-based double-blind placebo-controlled RCT designed to evaluate the clinical and cost effectiveness of an extended corticosteroid (prednisolone) treatment over 16 weeks compared with the standard 8-week treatment regimen in children with corticosteroid-sensitive nephrotic syndrome. Participants were recruited from general hospitals and tertiary paediatric nephrology units across the UK, and were followed up for at least 24 months up to a maximum of 48 months; the study closed when the last participant had completed 24 months of follow-up.

In accordance with the study protocol, the proxy-reported version of the PedsQL^{TM} and the CHU-9D were used to collect HRQoL data at baseline, week 16, and at months 12, 24, 36 and 48 for children in both study arms. PedsQL^{TM} was completed for children across all age groups (2–18 years) using the appropriate age-specific module, whilst the CHU-9D was completed for children who were 5 years and older [17]. In order to optimise the sample size, data on children who had completed both instruments across all timepoints were considered relevant for the mapping exercise. The sample was split into either an estimation or a validation sample. The estimation sample was used to develop the models while the validation sample was used for internal cross-validation of the mapping models.

Two approaches were available for selecting the estimation and the validation sample. The first was to randomise children at baseline into either the estimation or validation sample, and then account for the panel nature of the data in the regression equations. This approach results in a dataset that ensures all observations from individual children are either contained within the estimation or the validation sample and never within both. This, however, also results in the variation being reduced, as fewer children are contained within each sample. Therefore, an alternative approach was chosen for this mapping study whereby the entire sample of time-variant observations were randomised into either the estimation or the validation sample; and a clustering variable was included to account for having multiple observations from the same child. A preliminary analysis was conducted to explore the impact on the predicted CHU-9D index score of having the same participants but different observations in the estimation and validation sample (see Electronic Supplementary Material Appendix). Randomising the entire sample into an estimation and a validation sample limits the issue of overfitting when selecting the final model; a 3:1 split was used [18]. Of the entire sample, 50% was randomised to the validation sample in an attempt to rigorously avoid over-fitting. The estimation and validation samples contained only observations with valid CHU-9D and PedsQL^{TM} index scores, after excluding missing items.

2.2 Outcome Measures

The CHU-9D was initially designed for children aged 7–11 years; however, further research has now extended its use to children as young as 5 years [19–21] and to adolescents up to age 17 years [22]. The use of the instrument in 5-year-olds is currently being trialled [23]. The self- and proxy-reported versions of the CHU-9D questionnaire each consist of nine dimensions: sad, worried, annoyed, tired, sleep, pain, school, routine, and activity. Each dimension contains five severity levels, resulting in a possible 1953125 unique health states associated with the measure. Responses from the CHU-9D instrument were transformed into quality-of-life (utility) weights derived from a UK general population sample using an algorithm developed by Stevens [14]. Applying these weights produces a utility value set of between 0.33 (worst health state) and 1 (best health state), and a utility score of zero denotes death on the CHU-9D scale.

The PedsQL^{TM} generic core scale is a well-validated non-preference-based measure developed for toddlers, school-age children and adolescents. The self-reported version of the questionnaire has been validated in 5- to 18-year-olds while the parent- or proxy-reported version is valid for use in 2- to 18-year-olds [12, 24]. Both versions of the instrument have the same number of items across the four subscales or domains of health for each age-specific module: toddlers (2–4 years), young children (5–7 years), older children (8–12 years) and adolescents (13–18 years). The number of items within the health domain varies for some modules. The physical functioning (PF) domain has eight items, and both the emotional functioning (EF) and the social functioning (SF) domains have five items each. School functioning (FU) has five items for all age groups except toddlers where there are three items. Similar to the CHU-9D instrument, responses to each of the 23 items are on a 5-point scale of increasing severity from 0 to 4: never a problem; almost never a problem; sometimes a problem; often a problem; and almost always a problem. Responses are then reverse scored and linearly transformed (0 = 100, 1 = 75, 2 = 50, 3 = 25 and 4 = 0). The total PedsQL^{TM} score is the mean of the transformed score from all items answered. The total score is expressed on a 0 to 100 scale with 100 reflecting best possible health state

2.3 Analysis

Characteristics of participants in the study were summarised as means and standard deviation (SD)

for continuous variables, and frequency (%) for categorical variables. The conceptual overlap between the two instruments across the whole sample was explored using Spearman correlation coefficients.

The prediction mapping exercise regressed the CHU-9D utility scores (dependent variable) against the PedsQL^{TM} total, subscale or item scores (independent variables) to generate an algorithm that could then be subsequently used to predict the CHU-9D values. In order to select the model with a good prediction accuracy, three 'functional forms' or estimators were explored since it was not pragmatic to compare all mapping functions that are available. The estimators were chosen based on their perceived theoretical advantage and their performance in previous mapping studies.

The first was the ordinary least squares (OLS) regression with predicted utility scores censored at the value of 1. Whilst the OLS regression minimises the sum of squared errors, and represents the most common method within mapping studies [18], it has been shown to not cope well with multi-modal distributions [25] and does not always predict a perfect health. Despite its limitation, the OLS often gives good prediction accuracy in mapping.

The generalised linear model (GLM) [26] was chosen as the second functional form because it accommodates skewness and heteroscedasticity in the estimation sample. The GLM requires specification of a distribution 'family' that captures the relationship between the mean and variance, and a link function between the linear part and the mean. The Modified Park test was applied to identify the preferred 'family' based on the lowest Chi squared (ν^2) value, and the Hosmer–Lemeshow and Pearson correlation tests [27, 28] were used to select the link function, defined as fitting well if both tests yielded non-significant P-values. The third form chosen for the prediction function was the Tobit model, a censored regression that accommodates both the lower and upper limit utility scores [29]. Tobit models have been suggested for mapping despite concerns about inconsistencies in the presence of non-normality and heteroscedasticity [30].
As well as the three functional forms chosen, other models have been used for mapping such as the beta-binomial estimator and finite mixture models used to accommodate skewed distributions [15]. However, neither of these models have been shown to be better than GLM or OLS when predicting utility value at near perfect health state [31, 32]. Furthermore, the MM-estimator [33] has the potential to cope with heteroscedasticity and the undesired effect of outliers within the estimation sample, and has been shown to have the lowest predictive error in a previous paper that mapped PedsQL^{TM} onto CHU-9D in an older population [15]. Unfortunately, however, the MM-estimator does not permit the use of cluster variables, which are required given the nature of this mapping sample. In addition, there are alternatives to the Tobit estimator for handling ceiling effects such as the multi-variable fractional polynomials (MFP) and the censored least absolute deviation (CLAD) but, again, neither of these estimators have been convincingly shown to be better than OLS [34].

In summary, six model specifications (covariates) were developed based on the OLS, Tobit and the GLM 'functional forms', thus generating 18 models in total. The modes specification/covariates of these models are as follows:

Model-1 PedsQL^{TM} total scale score Model-2 Model-1, age and sex Model-3
PedsQL^{TM} subscale scores Model-4 Model-3, age and sex
Model-5 PedsQL^{TM} subscale score square terms and
interaction terms
Model-6 Model-5, age and sex

The PREDNOS data are a longitudinal dataset that can be viewed as having a two-stage structure, where the data collection timepoints (level 1 units) are nested within subjects/patients (level 2 unit). A random-intercept mixed-effect model is often used to account for multi-level hierarchical data

structures, but was not considered appropriate in our mapping context because the relationship between CHU-9D and PedsQL™ should be the same regardless of when the questionnaire was administered. In practice, neither the CHU-9D utility score nor PedsQL™ total score computation depend on follow-up timepoints within studies. Therefore, the PREDNOS data were considered to have only one hierarchical level, which is at the patient level. The within-patient correlation was taken into account by including the 'clustering' option for each of the 18 model specifications. For example, Model-1 specification is as follows:

regress [CHU-9D score] [PedsQL™ score], vce (cluster, [patient ID]).
where [patient ID] was a unique patient identifier.

2.4 Assessing Model Performance

The following selection criteria were applied to shortlist the models.

Step 1 The models were assessed on the exactness of their mean prediction in the estimation sample [35]. Models that accurately predicted the mean CHU-9D score up to one-ten-thousandth of a QALY were shortlisted for the next step

Step 2 One model from each functional form was selected based on their combined prediction accuracy in the estimation and validation sample. The indicators of prediction accuracy were the mean absolute error (MAE) and the mean square error (MSE). The MAE is the mean absolute difference between the observed and the predicted values, while MSE is the mean squared difference between the observed and the predicted CHU-9D utility score. Larger MAE and MSE values indicate poorer fit, and vice versa. MAE was prioritised over MSE as the primary criterion because it has been shown to be less sensitive to outliers [36], which are often found with utility data

Step 3 To assess and compare the shortlisted models from step 2, a series of assessments were applied. First, the distribution of the predicted and the observed CHU-9D scores were plotted to examine how well the predicted scores matched the observed. Second, the proportion of predictions deviating from observed values by < 0.03, 0.05 and < 0.1 were calculated as a representation of how often the models produce reliable predictions. Lastly, the MAEs were presented for different CHU-9D ranges to assess how well the models perform at the top and lower ranges of index scores

All analysis described in Sect. 2.3 follows the Mapping onto Preference-based measures reporting Standards (MAPS) [37]. The Akaike information criterion (AIC), Bayesian information criterion (BIC) and R-square for selected models were presented for the final model but these were not used as model-selection criteria. The purpose of a mapping function is to predict utility values, not on explanatory power or fit of the function.

3 Results

3.1 Sample Characteristics

There were 643 observations across the five data collection timepoints from children who were aged 5 years or older presenting with first-episode corticosteroid-sensitive nephrotic syndrome. These observations were randomised into groups A (n = 321) and B (n = 322). The longitudinal nature of the study meant that the number of missing items in the two groups varied across the data collection timepoints. After removing observations where either CHU-9D or PedsQL™ index score could not be computed, the remaining 279 observations with pairs of valid PedsQL™ and CHU-9D index scores in the first group formed the estimation sample, while the 284 observations in the second group formed the validation sample. The estimation and validation samples constituted the total mapping sample (N = 563).

The randomisation yielded a balanced distribution of demographic characteristics between the estimation and the validation sample (Table 1). The mean CHU-9D utility score was 0.93742 (SD = 0.07897) and 0.94094 (SD = 0.07173) for all observations within the estimation and validation sample, respectively. The mean PedsQL™ score was 80.93 (SD = 16.76) within the estimation sample and 80.31 (SD = 17.79) within the validation sample. Within each sample, the mean PedsQL™ total score was lower than the mean CHU-9D utility score when both scores were standardised on a 100-point scale. Although both HRQoL measures were negatively skewed (Fig. 1), the ceiling effect was more prominent with the CHU-9D. Level 1 or 'no problem' always had the highest proportion of responses. For more details on the CHU-9D responses please refer to the Electronic Supplementary Material.

There was a moderate statistical correlation between the CHU-9D utility scores and PedsQL™ total scores (Spearman's rho = 0.530; P<0.0001). The correlations between the CHU-9D utility score and PF, EF, FU and SF were 0.438, 0.585, 0.377 and 0.422, respectively. The Spearman correlation coefficient between the CHU-9D dimensions and PedsQL™ subscale scores/functions ranged from -0.0672 to - 0.4523. All correlations were statistically significant (P<0.0001).

3.2 Performance and Validation

Table 2 summarises the performance measures for all the model specifications, for both the estimation and validation sample. Within the estimation sample, the models were able to reasonably predict the mean CHU-9D value (0.93742; SD = 0.07898). Of the 18 models, 12 were able to predict the precise mean value by up to one-ten-thousandth of a utility value, and were therefore shortlisted for the next selection process. The exceptions were the six Tobit models. Within the validation sample, however, the models were less able to predict the mean CHU-9D score (0.94094; SD = 0.07174). The GLM_2 had the lowest mean predicted value (0.93409) while Tobit_3 had the highest mean predicted value (0.96575), giving a difference between the observed and predicted mean values of 0.0069 and 0.0245, respectively. These differences were below the threshold of 0.03—generally considered to be a minimally important difference [38, 39]. A further observation was that some OLS models and all of the Tobit models had maximum predicted values beyond the upper limit of the CHU-9D utility scale (0.33–1.00). However, none of the models predicted a utility value below the lower limit of the CHU-9D utility scale.

All Tobit models were excluded from further analysis after the first selection criteria. For step 2, GLM_6 and OLS_3 had the 'best' performance in terms of MAE in the estimation and validation sample for their respective functional forms, and were therefore selected for a final comparison: step 3. GLM_6 performed in the estimation sample, but the reverse was observed in the validation sample.

Table 3 contains the model performance results. For the final models in step 3, the distribution of the predicted score was also examined (Fig. 2). GLM_6 had a wide range of predicted CHU-9D scores compared with OLS_3 (Fig. 2). Approximately 56% of the predicted values from GLM_6 in the validation sample had absolute errors lower than the minimally important difference value of 0.03; the corresponding value for the OLS3 was 53%. GLM_6 remained the preferred model specification when the error threshold was extended to 0.05.

Although the prediction accuracy of the mean scores were similar in both models, the accuracy level was not uniform across the CHU-9D utility range, as shown in Table 4. The number of observations with a utility score of <0.7 was small; therefore the comparison between the best two models were restricted to observations with higher ultility values. GLM_6 was superior to OLS_3 in the estimation sample; however, in the validation sample there were diverging results. OLS_3 had a better prediction accuracy when utility values were higher than 0.8 but less than full health, whilst the GLM_6 was superior at predicting full health and utility values between 0.7 and 0.8. So, although OLS_3 had a better prediction accuracy in the validation sample overall, it was found to be only marginally better than GLM_6.

In summary, relative to GLM_6, OLS_3 lacked the ability to predict the wider range of CHU-9D values (0.71), and a higher proportion of its predicted values had absolute errors above the minimally important difference. Furthermore, it was less able to predict full health; this is particularly important for utity data, which tends to have ceiling effects. Taking all these factors into account, the GLM_6 model was selected as the preferred model for mapping from PedsQL^{TM}to CHU-9D. Table 5 shows the coefficients for generating deterministic and probabilistic
ty predictions using the GLM_6 model. Coefficients for OLS_3 have also been presented in situations where this might be desired. Using GLM_6 as an example, the CHU-9D utility score can be calculated from the following coefficients:

CHU - 9D utility score

$$
\begin{aligned}
&= 0.7135215 + (\text{PedsQL PF})^2\,(1.62 * 10^{-4}) \\
&\quad + (\text{PedsQL EF})^2\,(4.77 * 10^{-4}) \\
&\quad - (\text{PedsQL SF})^2\,(4.0 \times 10^{-5}) \\
&\quad - (\text{PedsQL FU})^2\,(1.646 \times 10^{-4}) \\
&\quad - (\text{PedsQL PF})\,(\text{PedsQL EF})\,(1.10 \times 10^{-4}) \\
&\quad - (\text{PedsQL PF})\,(\text{PedsQL SF})\,(1.14 \times 10^{-4}) \\
&\quad + (\text{PedsQL PF})\,(\text{PedsQL FU})\,(3.7 \times 10^{-5}) \\
&\quad - (\text{PedsQL EF})\,(\text{PedsQL SF})\,(2.46 \times 10^{-4}) \\
&\quad - (\text{PedsQL EF})\,(\text{PedsQL FU})\,(1.16 \times 10^{-4}) \\
&\quad + (\text{PedsQL SF})\,(\text{PedsQL FU})\,(4.36 \times 10^{-4}) + \text{Age}\,(2.79 \times 10^{-2}) - (5.46 \times 10^{-4})\text{if femal}
\end{aligned}
$$

Discussion

Whilst complying with current guidance for conducting and reporting mapping analyses [37], the results of this analysis show that CHU-9D utility scores can be estimated from PedsQL^{TM} subscale scores with sufficient accuracy.

A total of 18 models were explored: six model specifications, each with three functional forms. All the models produced reasonably similar predictions of the mean utility scores. The preferred algorithm for mapping PedsQL^{TM} onto CHU-9D was selected using a three-stage elimination process. The GLM and OLS models outperformed the Tobit models in terms of MAE, with GLM_6 and OLS_3 being shortlisted for the third and final selection process. GLM_6 was chosen as the preferred mapping model because it was able to predict a wider range of CHU-9D utility scores and had a higher proportion of predicted values, deviating from the observed values by less than 0.03.

The GLM_6 model predicted the CHU-9D utility scores with more accuracy than other similar published studies (MAE = 0.04078; MSE = 0.00353). For example, in one study that looked at the relationship between the CHU-9D and the SDQ (Strengths and Difficulties Questionnaire; a behavioural questionnaire), the MSE was 0.124 [40, 41], while another study that estimated CHU-9D utility scores from the KIDSCREEN questionnaire had an MAE of 0.095 [42]. The GLM_6 produced from this analysis also performed better than a previous model that had predicted EQ–5D–Y (EuroQoL–youth version) utility scores from PedsQL^{TM} (MAE = 0.115) [35]. Furthermore, Mpundu-Kaambwa and colleagues [15] mapped the PedsQL^{TM} onto the CHU-9D (MAE = 0.1169; MSE = 0.0213) using an Australian value set and data derived from 15- to 17-yearolds. The mapping algorithm reported here has been derived from more observations, has a wider age range and used a UK value set.

Despite these strengths, there are some limitations. The sample size was small compared with other mapping studies [5], thereby limiting the ability to robustly demonstrate the relationship between CHU-9D and PedsQL^{TM}. A larger sample size may have reduced the prediction error of the model. Response mapping is an alternative to mapping summary scores whereby the domain scores from each instrument are mapped. However, a recent review [5] concluded that this more complex approach did not offer anything superior over and above a simpler mapping of summary scores, a finding that is supported by a recent applied study that mapped from PedsQL^{TM} onto CHU-9D [15]. For these reasons,

this mapping exercise focused on summary score mapping only. Another caveat was the ceiling effect. A wider spectrum of health profiles was lacking as a considerable number of children had perfect or near perfect health, with none having utility scores below 0.5 in the estimation sample. This was reflected in the distribution of scores across the five response levels for each of the CHU-9D domains and each PedsQL^{TM} subscale score. The implication is that caution is advised if using the algorithm for a less healthy paediatric population. Additionally, the age range of the sample reflected the natural history of the condition, with few older adolescents. Future research can focus on refining this mapping algorithm should data for children with more severe health states across a wider age range become available.

A final methodological concern was the proxy completion of both HRQoL instruments in this study given that self-completion is the gold standard approach for measuring HRQoL, at least in adults [43, 44]. Proxy-reported responses are not directly interchangeable because of proxy biases [45]. However,

for some very young children the use of proxies is unavoidable.

Mapping is not a substitute for direct utility estimation. It is therefore advised that, where possible, preference-based outcomes be collected for the measurement of cost effectiveness. However, if direct estimation is not feasible, the algorithm presented provides a valuable and scientifically robust approach to predicting CHU-9D utility values.

The standard errors for the coefficients are reported, making it possible for future users of the algorithm to account for the uncertainty around the predicted values.

5 Conclusion

This study builds on a previous study that mapped PedsQL™ onto CHU-9D scores within a sample of 15- to 17-year-olds using the Australian value set [15]. Our findings show that CHU9D scores can be estimated from PedsQL™ generic core scale values with good prediction accuracy. The CHU-9D index score for this study was derived using the UK value set. Therefore, this algorithm can be used to generate QALYs for evaluating the cost effectiveness of interventions targeting children. Future research should consider validating this algorithm further in children with lower CHU-9D utility index scores than those observed in the PREDNOS study.

Data Availability The data for this analysis were from the PREDNOS study and are not yet available because the study results have not been published. At the discretion of the funder (NIHR) of the study, the data may become publicly available at a later date. The methods section of the paper explains the randomisation and regression analysis underpinning this study. STATA® do-files are, however, available from the corresponding author on request.

Acknowledgements

The PREDNOS investigators would like to thank the National Institute for

Health Research (NIHR) Health Technology Assessment Programme for funding this study and Kidney Research UK and Kid's Kidney Research, who funded the pilot study. We thank all the parents and participants who agreed to enter the study, the investigators who contributed to the trial and the NIHR Clinical Research Network: Children for their support with study setup and recruitment. We also thank the Service Users Group from the UK Nephrotic Syndrome Trust (NeST) and the Renal Patient Support Group, in particular Mrs Wendy Cook, Mrs Samantha Davies-Abbott and Dr Shahid Muhammad, who provided valuable input regarding trial design and

conduct and reviewed the plain English summary. The support of the British Association for Paediatric Nephrology and the Royal College of Paediatrics and Child Health is much appreciated. We also value and appreciate constructive comments received at the Health Economists' Study Group meeting in January 2017. We thank Stavros Petrous in particular, who discussed an earlier version of this paper at the meeting.

List of Trial Co-Investigators: Abertawe Bro Morgannwg University Health Board, Morriston Hospital: Dr M. James-Ellison, H.M. Williams. Airedale NHS Foundation Trust, Airedale General Hospital: Dr P. Bala. Alder Hey Children's NHS Foundation Trust, Alder Hey Children's Hospital: Dr C. Jones, Dr R. Holt, Dr H. Morgan, E. Bailey, L. Flanagan. Ashford and St Peter's Hospitals NHS Foundation Trust, St Peter's Hospital: Dr T. Bhatti, Dr S. Bahl, L. Walding. Barnsley Hospital NHS Foundation Trust, Barnsley Hospital: Dr E. Gouta, Dr R. Gupta, Dr D. Kerrin. Barts Health NHS Trust, Newham University Hospital, The Royal London Hospital, Whipps Cross University Hospital: Dr A. Parikh, Dr A. Duthie, Dr N. Gadong, Dr B. Anjum, B. Crone, H. De Jesus, I. Lancoma-Malcolm, N. Plaatjies. Basildon and Thurrock University Hospitals NHS Foundation Trust, Basildon University Hospital: Dr K. Khalifa. Belfast Health and Social Care Trust, The Royal Belfast Hospital for Sick Children: Dr K. McKeever, Dr G. McCall, P. McCreesh, M. Millar. Betsi Cadwaladr University Health Board, Glan Clwyd Hospital, Ysbyty Gwynedd: Dr M. Hesseling, Dr M. Kubwalo, A. Bolger, L. Hobson. Birmingham Women's and Children's NHS Foundation Trust, Birmingham Children's Hospital: Dr D. Milford, Dr L. Kerecuk, C. Norton, N. Rahania, H. Williamson, F. Bibi, T. Gazeley, J. Williams. Bolton NHS Foundation Trust, Royal Bolton Hospital: Dr F. Watson, C. Abbot. Bradford Teaching Hospitals NHS Foundation Trust, Bradford Royal Infirmary: Dr S. Frazer, L. Akeroyd. Burton Hospitals NHS Foundation Trust, Queen's Hospital: Dr M. Ahmed, Dr D. Muogbo, C. Backhouse, S. Boswell. Calderdale and Huddersfield NHS Foundation Trust, Calderdale Royal Hospital, Huddersfield Royal Infirmary: Dr E. Crosbie. Cambridge University Hospitals NHS Foundation Trust, Addenbrooke's Hospital: Dr P. Heinz, Dr B. Ulbrich. Cardiff and Vales University Health Board, University Hospital of Wales: Dr S. Hegde, P. Jones. Central Manchester University Hospitals NHS Foundation Trust, Royal Manchester Children's Hospital: Professor N.J.A. Webb, Dr R. Lennon, Dr M. Lewis, Dr N. Plant, Dr M. Shenoy, Dr S. Sukthankar, C. Bryant, S. Douglas, H. Sumner, J. Howell. Chelsea and Westminster Hospital NHS Foundation Trust, West Middlesex University Hospital: Dr N. Elhadi, I. Bishton. Chesterfield Royal Hospital NHS Foundation Trust, Chesterfield Royal Hospital: Dr O. Ayonrinde. Colchester Hospital University NHS Foundation Trust, Colchester General Hospital: Dr A. Turner, Dr J. Campbell, A. Turner. Countess of Chester Hospital NHS Foundation Trust, Countess of Chester Hospital: Dr J. Gibbs, Dr E. Newby, Dr S. Saladi, Dr A. Timmis, C. Burchett, S. De-Beger. County Durham and Darlington NHS Foundation Trust, Darlington Memorial Hospital, University Hospital of North Durham: Dr A. Nair, Dr A. Banerjee. Croydon Health Services NHS Trust. Croydon University Hospital: Dr T. Fenton. Dartford and Gravesham NHS Trust, Darent Valley Hospital: Dr S. D'Costa. Derby Teaching Hospitals NHS Foundation Trust, Royal Derby Hospital: Dr R. Bowker, C. Smith, V. Unsworth. Doncaster and Bassetlaw Teaching Hospitals NHS Foundation Trust, Bassetlaw Hospital, Doncaster Royal Infirmary: Dr A. Natarajan, Dr L. Men Wong. Dorset County Hospital NHS Foundation Trust, Dorset County Hospital: Dr P. Parslow. East and North Hertfordshire NHS Trust, Lister Hospital: Dr V. Kandala, Dr S. Imran. East Cheshire NHS Trust, Macclesfield District General Hospital: Dr S. Chandrasekaran, Dr S.-A. Ho, Dr I. Losa, Dr K. Marinaki, N. Keenan, J. Nichols, J. Shippey. East Lancashire Hospitals NHS Trust, Royal Blackburn Hospital: Dr J. Iqbal. Epsom and St Helier University Hospitals NHS Trust, Epsom Hospital: Dr K. Watts.

Frimley Health NHS Foundation Trust, Wexham Park Hospital: Dr A. Yannoulias, Dr Z. Huma. Gloucestershire Hospitals NHS Foundation Trust, Gloucestershire Royal Hospital: Dr A. Sambo, Dr L. Jadresic, S. Beames. Great Ormond Street Hospital for Children NHS Foundation Trust, Great Ormond Street Hospital: Dr D. Bockenhauer, Dr D. Hothi. Great Western Hospitals NHS Foundation Trust, Great Western Hospital: Dr N. West, Dr H. Price. Guy's and St Thomas' NHS Foundation Trust, Evelina London Children's Hospital: Dr M. Sinha, E. Neal, E. Reus, P. Sofocleous. Harrogate and District NHS Foundation Trust, Harrogate District Hospital: Dr I. Cannings. Homerton University Hospital NHS Foundation Trust, Homerton University Hospital: Dr R. Sood. Hull and East Yorkshire Hospitals NHS Trust, Hull Royal Infirmary: Dr V. Mathew, Dr E. Sharif. Kettering General Hospital NHS Foundation Trust, Kettering General Hospital: Dr H. Bilolikar, S. White. Kingston Hospital NHS Foundation Trust, Kingston Hospital: Dr T. Ayeni. Lancashire Teaching Hospitals NHS Foundation Trust, Royal Preston Hospital: Dr R. O'Connor. Luton and Dunstable Hospital NHS Foundation Trust, Luton and Dunstable University Hospital: Dr M. Eisenhut, Dr T. Rajkowski, S. Clough, R. Dixon, K. Reep. Maidstone and Tunbridge Wells NHS Trust, Maidstone Hospital, Tunbridge Wells Hospital: Dr K. Balasubramanian, Dr H. Kisat. Medway NHS Foundation Trust, Medway Maritime Hospital: Dr J. Cansick. Mid Cheshire Hospitals NHS Foundation, Trust Leighton Hospital: Dr J. Ellison, S. Smith, S. Tapscott. NHS Ayrshire and Arran, University Hospital Crosshouse: Dr B. Oates, C. Bell. NHS Dumfries & Galloway, Dumfries and Galloway Royal Infirmary: Dr R. Shyam. NHS Fife, Victoria Hospital: Dr E. Menzies, P. Cruickshanks. NHS Forth Valley, Forth Valley Royal Hospital: Dr Schulga. NHS Grampian, Royal Aberdeen Children's Hospital: Dr C. Oxley. NHS Greater Glasgow and Clyde, Royal Alexandra Hospital, Royal Hospital for Children: Dr D. Hughes, Dr A. Sharma, Dr L. Krischock, Dr J. Houston, D. Carroll, S. Goddard, E. Waxman. NHS Highland, Raigmore Hospital: Dr A. Webb. NHS Lanarkshire, Wishaw General Hospital: Dr T.T. Saing, Dr C. Cunningham, A. Brown, V. Causer. NHS Lothian, Royal Hospital for Sick Children: Dr D. Hughes, Dr S. Taheri, A. Macdonald, T. McGregor, E. Carson,

Riding. Norfolk and Norwich University Hospitals NHS Foundation Trust, Norfolk and Norwich University Hospital: Dr N. Thalange, Dr C. Upton, L. Fear. North Cumbria University Hospitals NHS Trust, Cumberland Infirmary: Dr G. Jones. North Tees and Hartlepool NHS Foundation Trust, University Hospital of North Tees: Dr V. Tandle, G. Osborne. North West Anglia NHS Foundation Trust, Peterborough City Hospital: Dr M. Aslam. Northampton General Hospital NHS Trust, Northampton General Hospital: Dr I. Norton. Northern Devon Healthcare NHS Trust, North Devon District Hospital: Dr D. Dalton, Dr K. Thattakkat. Northern Lincolnshire and Goole NHS Foundation Trust, Diana, Princess of Wales Hospital: Dr B. Wobi, Dr B. Etuwewe, T. Walker, J. Browning, M. Carling, E. Killingbeck, D. Ryan, C. Tsvangirai-Mahachi. Nottingham University Hospitals NHS Trust, Nottingham Children's Hospital: Dr A. Lunn, Dr M. Christian. Oxford University Hospitals NHS Foundation Trust, John Radcliffe Hospital: Dr J. Craze. Plymouth Hospitals NHS Trust, Derriford Hospital: Dr C. Derry, Dr R. Jones, S.-J. Sharman. Poole Hospital NHS Foundation Trust, Poole Hospital: Dr S. Wadams, H. Barham, S. Power. Portsmouth Hospitals NHS Trust, Queen Alexandra Hospital: Dr C. Tuffrey, Dr S. Birch, Dr J. Scanlan, A. Gribbin, S. McCready. Royal Berkshire NHS Foundation Trust, Royal Berkshire Hospital: Dr A. Gordon, Dr G. Boden, S. Hallett. Royal Cornwall Hospitals NHS Trust, Royal Cornwall Hospital: Dr C. Williams, G. Craig. Royal Devon and Exeter Hospital NHS Foundation Trust, Royal Devon and Exeter-Wonford Hospital: Dr H. Cottis, Dr V. Lewis, Dr C. McMillan, Dr N. Osborne, C. Harrill, S. Wilkins. Royal Free London NHS Foundation Trust, Royal Free Hospital (non-recruiting centre): Professor R. Kleta. Royal United Hospitals Bath NHS Foundation Trust, Royal United Hospital: Dr L. Diskin, Dr A. Billson, A. Barratt. Sheffield Children's NHS Foundation Trust, Sheffield Children's Hospital: Dr G. Moss. Sherwood Forest Hospitals NHS Foundation Trust, King's Mill Hospital: Dr S. Rhodes, A. Coupe, C. Moulds. Southern Health and Social Care Trust,

Craigavon Area Hospital: M. Smith, S. Gilpin, J. Ratcliffe. Southport and Ormskirk Hospital NHS Trust, Ormskirk and District General Hospital: Dr S. Gardner. St Helens and Knowsley Teaching Hospitals NHS Trust, Whiston Hospital: Dr L. Chilukuri. Stockport NHS Foundation Trust, Stepping Hill Hospital: Dr C. Cooper, S. Bennett. Tameside and Glossop Integrated Care NHS Foundation Trust, Tameside Hospital: Dr A. Petkar, Dr A. Ul-Haq, G. Waring. Taunton and Somerset NHS Foundation Trust, Musgrove Park Hospital: Dr R. Mann, N. Thorne. The Dudley Group NHS Foundation Trust, Russells Hall Hospital: Dr Z. Ibrahim. The Ipswich Hospital NHS Trust, Ipswich Hospital: Dr J. Buck, D. Beeby. The Leeds Teaching Hospitals NHS Trust, Leeds General Infirmary: Dr E. Finlay, M. Allen, R. Mottram. The Mid Yorkshire Hospitals NHS Trust, Dewsbury and District Hospital, Pinderfields Hospital: Dr K. Deakin. The Newcastle upon Tyne Hospitals NHS Foundation Trust, Great North Children's Hospital: Dr S. Johnson, Dr Y. Tse, K. Bell, D. Chisholm. The Pennine Acute Hospitals NHS Trust, North Manchester General Hospital, The Royal Oldham Hospital: Dr B. Padmakumar, Dr T. Magadevan (clinician, North Manchester General Hospital). The Queen Elizabeth Hospital King's Lynn NHS Foundation Trust, Queen Elizabeth Hospital: Dr M. Hughes, Dr S. Rubin, A. Frary. The Rotherham NHS Foundation Trust, Rotherham Hospital: Dr S. Suri, Dr S. El-Refee, Dr C. Harrison. The Royal Wolverhampton NHS Trust, Cannock Chase Hospital, New Cross Hospital: Dr K. Davies, S. Kempson. The Shrewsbury and Telford Hospital NHS Trust, Princess Royal Hospital, Royal Shrewsbury Hospital: Dr N. Ayub. United Lincolnshire Hospitals NHS Trust, Lincoln County Hospital, Pilgrim Hospital: Dr S. Mukhopadhyay, Dr M. Crawford, Dr M. Pervez, Dr D. Thomas. University Hospital of South Manchester NHS Foundation Trust, Wythenshawe Hospital: Dr G. Vemuri, Dr R. Puttha, Dr A. Setti, Dr F. Al-Zidgali, C. Holliday, L. Woodhead. University Hospital Southampton NHS Foundation Trust, Southampton Children's Hospital: R. Gilbert. University Hospitals Bristol NHS Foundation Trust, Bristol Royal Hospital for Children: Professor M. Saleem, M. Ross, H. Smee. University Hospitals Coventry and Warwickshire NHS Trust, University Hospital: Dr N. Coad. University Hospitals of Leicester NHS Trust, Leicester Royal Infirmary: Dr A. Hall, Dr P. Houtman. University Hospitals of Morecambe Bay NHS Foundation Trust, Furness General Hospital, Royal Lancaster Infirmary: Dr N.P. Nardeosingh, Dr M.-Y. Formosa, Dr P. Ward. University Hospitals of North Midlands NHS Trust, County Hospital, Royal Stoke University Hospital: Dr Y. Slater, Dr S. Raja, Dr K. Tewary, Dr S. Shankar. Warrington and Halton Hospitals NHS Foundation Trust, Warrington Hospital: Dr R.D. Webb, N. Rogers. West Suffolk NHS Foundation Trust, West Suffolk Hospital: Dr R. Lakshman, Dr K. Cesar. Western Health and Social Care Trust, Altnagelvin Area Hospital: Dr D. Armstrong, J. Brown. Western Sussex Hospitals NHS Foundation Trust, St Richard's Hospital, Worthing Hospital: Dr N. Brennan, Dr S. Nicholls. Wirral University Teaching Hospital NHS Foundation Trust, Arrowe Park Hospital: Dr E. Breen, S. Hughes, L. Lewis. Worcestershire Acute Hospitals NHS Trust, Alexandra Hospital, Worcestershire Royal Hospital: Dr J. Scanlon, Dr M. Ahmed, Dr A. Gallagher. Wrightington, Wigan and Leigh NHS Foundation Trust, Royal Albert Edward Infirmary: Dr V. Joshi, Dr M. Mukherjee, Dr S. Velmurugan. Wye Valley NHS Trust, The County Hospital: Dr S. Meyrick. Yeovil District Hospital NHS Foundation Trust, Yeovil District Hospital: Dr M. Fernando, Dr C. Zaborowski, A. Stannett. York Teaching Hospital NHS Foundation Trust, Scarborough Hospital, The York Hospital: Dr U. Venkatesh, Dr G. Millman, S. Dyer, K. Elliot. University of Manchester (non-recruiting centre): Professor D. Ray.

Author contributions Tosin Lambe (TL): conducted the mapping analysis and wrote the paper. Emma Frew (EF): developed the mapping protocol, supervised the mapping analysis and wrote the paper. Natalie J. Ives (NJI): supervised the statistical aspects of the study and edited the paper. Rebecca L. Woolley (RLW): supervised the statistical aspects and edited the paper. Carole

Cummins (CC): major input into the trial design that generated the data for this mapping study, and edited the paper. Elizabeth A. Brettell (EAB): overall co-ordination, management and oversight of the PREDNOS study that generated the data for this mapping paper and edited the paper. Emma N. Barsoum (ENB): co-ordination of the PREDNOS trial, significant input into the study set-up, monitoring and ensuring execution of the PREDNOS study that generated data for this mapping paper, and edited this paper. Professor Nicholas J.A. Webb (NJAW): Chief Investigator of the PREDNOS study, led the protocol design, recruited participants and edited this mapping paper.

Funding This study was funded by the NIHR Health Technology Assessment (HTA) programme (HTA 08/53/31).

Compliance with Ethical Standards

Ethical Approval The study was approved by the North West–GM Central Research Ethics Committee (reference: 12/NW/0766).

Department of Health Disclaimer The views expressed are those of the authors and not necessarily those of the National Health Service (NHS), NIHR or the Department of Health.

Conflict of Interest NJAW has served on Advisory Boards within the past 5 years for Abbvie, Alexion, AMAG, Astellas, Raptor, Takeda and UCB. These have been related to the design and conduct of early-phase trials in childhood kidney disease. None has been related to the treatment of corticosteroid-sensitive nephrotic syndrome. TL, EF, NJI, RLW, EAB, ENB, and CC declare no conflict of interest.

Informed Consent Written informed consent was obtained from parents/guardians of participants and written assent was obtained from participants of appropriate age.

References

1. Patrick DL, Deyo RA. Generic and disease-specific measures in assessing health status and quality of life. Med Care. 1989:27:S217–32.
2. Coons SJ, Rao S, Keininger DL, Hays RD. A comparative review of generic quality-of-life instruments. Pharmacoeconomics. 2000;17:13–35.
3. Mehrez A, Gafni A. Quality-adjusted life years, utility theory, and healthy-years equivalents. Med Decis Making. 1989;9:142–9.
4. Jones PW, Quirk F, Baveystock C. The St George's respiratory questionnaire. Respir Med. 1991;85:25–31.
5. 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;11:215–25.
6. Feeny D, Krahn M, Prosser LA, Salomon J. Valuing health outcomes. In: Neumann PJ, Sanders GD, Russell LB, Siegel JE, Ganiats TG, editors. Cost-effectiveness in health and

medicine. 2nd ed. New York: Oxford University; 2017. p. 167–200.

7. Brazier JE, Ratcliffe J, Salomon JA, Tsuchiya A. Measuring and valuing health benefits for economic evaluation. 2nd ed. Oxford: Oxford University; 2016.
8. Gerard K, Mooney G. QALY league tables: handle with care. Health Econ. 1993;2:59–64.
9. Nord E. The QALY—a measure of social value rather than individual utility? Health Econ. 1994;3:89–93.
10. Stevens KJ. Working with children to develop dimensions for a preference-based, generic, pediatric, health-related quality-of-life measure. Qual Health Res. 2010;20:340–51.
11. 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:e600–14.
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;39:800–12.
13. Eiser C, Morse R. A review of measures of quality of life for children with chronic illness. Arch Dis Child. 2001;84:205–11.
14. Stevens K. Valuation of the Child Health Utility 9D Index. Pharmacoeconomics. 2012;30:729–47.
15. Mpundu-Kaambwa C, Chen G, Russo R, Stevens K, Petersen KD, Ratcliffe J. Mapping CHU9D Utility Scores from the PedsQL™ 4.0 SF-15. Pharmacoeconomics. 2017;35(4):453–67.
16. Ratcliffe J, Huynh E, Chen G, Stevens K, Swait J, Brazier J, Sawyer M, Roberts R, Flynn T. Valuing the Child Health Utility 9D: using profile case best worst scaling methods to develop a new adolescent specific scoring algorithm. Soc Sci Med. 2016;157:48–59.
17. Webb N, Trompeter R, Cummins C, Wheatley K, Frew E, Kaambwa B, et al. Long-term tapering versus standard prednisolone (steroid) therapy for the treatment of the initial episode of childhood nephrotic syndrome: national multicentre randomised double blind trial. National Institute for Health Research Health Technology Assessment programme. 2014; ISRCTN16 645249. https://www.birmingham.ac.uk/Documents/college-mds/trials/bctu/prednos/PREDNOSProtocolV22.pdf. Accessed 13 Dec 2017.
18. Dakin H. Review of studies mapping from quality of life or clinical measures to EQ-5D: an online database. Health Qual Life Outcomes. 2013;11:1.
19. Frew EJ, Pallan M, Lancashire E, Hemming K, Adab P. Is utility-based quality of life associated with overweight in children? Evidence from the UK WAVES randomised controlled study. BMC Pediatr. 2015;15:211.
20. Adab P, Pallan MJ, Lancashire ER, Hemming K, Frew E, Griffin T, et al. A cluster-randomised controlled trial to assess the effectiveness and cost-effectiveness of a childhood obesity prevention programme delivered through schools, targeting 6–7 year old children: the WAVES study protocol. BMC Public Health. 2015;15:488.
21. Canaway AG, Frew EJ. Measuring preference-based quality of life in children aged 6–7 years: a comparison of the performance of the CHU-9D and EQ-5D-Y—the WAVES Pilot Study. Qual Life Res. 2013;22:173–83.
22. 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;15:1092–9.
23. University of Sheffield, School of Health and Related Research (ScHARR). A brief overview

of Child Health Utility 9D (CHU9D). 2017. https://www.sheffield.ac.uk/scharr/sections/heds/mvh/paediatric/about-chu9d. Accessed 27 Sep 2017.

24. 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;3:329–41.
25. Golub GH, Van Loan CF. An analysis of the total least squares problem. SIAM J Numer Anal. 1980;17:883–93.
26. Crawley MJ. Generalized linear models. In: The R book. 2nd ed. Chichester: Wiley; 2007.
27. Kramer AA, Zimmerman JE. Assessing the calibration of mortality benchmarks in critical care: the Hosmer-Lemeshow test revisited. Crit Care Med. 2007;35:2052–6.
28. Hosmer DW Jr, Lemeshow S. Applied logistic regression. New York: Wiley; 2004.
29. Tobin J. Estimation of relationships for limited dependent variables. Econometrica. 1958:26:24–36.
30. Reynolds A, Shonkwiler J. Testing and correcting for distributional misspecifications in the Tobit model: an application of the information matrix test. Empir Econ. 1991;16:313–23.
31. 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;12:163.
32. Crott R. Direct mapping of the QLQ-C30 to EQ-5D preferences: a comparison of regression methods. Pharmacoeconomics. 2017. https://doi.org/10.1007/s41669-017-0049-9.
33. Chen G, Khan MA, Iezzi A, Ratcliffe J, Richardson J. Mapping between 6 multiattribute utility instruments. Med Decis Making. 2016;36:160–75.
34. Payakachat N, Summers KH, Pleil AM, Murawski MM, Thomas J, 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;18:801–13.
35. Khan KA, Petrou S, Rivero-Arias O, Walters SJ, Boyle SE. Mapping EQ-5D utility scores from the

PedsQL™ generic core scales. Pharmacoeconomics. 2014;32:693–706.
36. Hyndman RJ, Koehler AB. Another look at measures of forecast accuracy. Int J Forecast. 2006;22:679–88.
37. Petrou S, Rivero-Arias O, Dakin H, Longworth L, Oppe M, Froud R, Gray A. The MAPS reporting statement for studies mapping onto generic preference-based outcome measures: explanation and elaboration. Pharmacoeconomics. 2015;33:993–1011.
38. Drummond M. Introducing economic and quality of life measurements into clinical studies. Ann Med. 2001;33:344–9.
39. Marra CA, Woolcott JC, Kopec JA, Shojania K, Offer R, Brazier JE, et al. A comparison of generic, indirect utility measures (the HUI2, HUI3, SF-6D, and the EQ-5D) and disease-specific instruments (the RAQoL and the HAQ) in rheumatoid arthritis. Soc Sci Med. 2005;60:1571–82.
40. Boyer NR, Miller S, Connolly P, McIntosh E. Paving the way for the use of the SDQ in economic evaluations of school-based population health interventions: an empirical analysis of the external validity of SDQ mapping algorithms to the CHU9D in an educational setting. Qual Life Res. 2016;25(4):913–23.

41. 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.
42. 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.
43. NICE. Guide to the methods of technology appraisal 2013. NICE guideline. PMG9. 2013. https://www.nice.org.uk/process/pmg9. Accessed 12 Dec 2017.
44. Lv Essen. Proxy ratings of patient quality of life factors related to patient–proxy agreement. Acta Oncol. 2004;43:229–34.
45. Rand S, Caiels J. Using proxies to assess quality of life: a review of the issues and challenges. Canterbury: QORU, University of Kent; 2015.

Table 1 Demographic characteristics of estimation and validation sample by data collection timepoint

Demographic characteristic	Timepoint					
	Baseline	4 months	12 months	24 months	36 months	48 months
Estimation sample						
n	55	47	58	54	39	26
Sex						
Male [n (%)]	35 (63.6)	33 (70.2)	36 (62.1)	32 (59.2)	23 (58.9)	18 (69.2)
Age (years)						
Mean (SD)	7 (2.1)	7.6 (2.1)	7.4 (2.1)	7.2 (1.9)	7.3 (1.9)	8.1 (2.0)
Median (IQR)	6 (3)	7 (4)	7 (3)	7 (2)	7 (3)	8 (3)
Range	5–12	5–12	5–12	5–12	5–12	5–12
CHU-9D						
Mean (SD)	0.940 (0.063)	0.929 (0.103)	0.941 (0.080)	0.950 (0.068)	0.922 (0.081)	0.937 (0.077)
Median (IQR)	0.952 (0.106)	0.952 (0.100)	0.952 (0.081)	0.968 (0.073)	0.931 (0.108)	0.967 (0.107)
Range	0.786–1.000	0.534–1.000	0.509–1.000	0.68–1.000	0.702–1.000	0.697–1.000
PedsQL™						
Mean (SD)	77.11 (16.16)	82.4 (16.8)	81.94 (15.91)	84.24 (14.31)	78.49 (20.58)	80.85 (17.72)
Median (IQR)	79.35 (28.26)	89.13 (29.35)	87.5 (20.65)	88.04 (18.48)	82.61 (30.43)	82.61 (29.35)
Range	40.22–100.00	45.65–100.00	41.3–100.00	43.48–100.00	31.52–100.00	39.13–100.00
Validation sample						
n	36	46	50	70	56	26
Sex						
Male [n (%)]	25 (69.4)	30 (65.2)	32 (64.0)	44 (62.9)	29 (51.8)	15 (57.7)
Age						
Mean (SD)	6.9 (1.8)	7.1 (1.9)	7.3 (2.0)	7.6 (2.2)	7.4 (2.2)	8 (1.9)
Median (IQR)	7 (3)	7 (2)	7 (3)	7 (3)	7 (3)	8 (2)
Range	5–11	5–12	5–12	5–12	5–13	5–13
CHU-9D						
Mean (SD)	0.924 (0.081)	0.945 (0.067)	0.941 (0.075)	0.938 (0.076)	0.951 (0.067)	0.945 (0.06)
Median (IQR)	0.952 (0.1)	0.96 (0.079)	0.96 (0.081)	0.952 (0.102)	0.967 (0.071)	0.959 (0.097)
Range	0.711–1	0.69–1	0.739–1	0.65–1	0.712–1	0.828–1
PedsQL™						
Mean (SD)	75.88 (16.91)	81.35 (14.53)	78.28 (19.01)	80.6 (17.25)	83.13 (19.11)	81.68 (20.3)
Median (IQR)	77.72 (27.36)	83.7 (18.48)	83.7 (28.26)	86.96 (27.17)	91.85 (27.17)	90.76 (20.65)
Range	42.39–97.83	41.3–100	21.74–100	33.7–100	40.22–100	29.35–100

CHU-9D Child Health Utility Index–9 Dimension, IQR interquartile range, PedsQL™ Paediatric Quality of Life Inventory generic core scale, SD standard deviation

Table 2 Performance of Paediatric Quality of Life Inventory generic core scale (PedsQL^{TM}) to Child Health Utility Index–9 Dimension (CHU-9D) index score models in the estimation and validation samples

	Estimation sample							Validation sample					Average MAE across samples
	Mean	Min.	Max.	MSE	MAE	AIC	BIC	Mean	Min.	Max.	MSE	MAE	
Observed	0.93742 (0.07898)	0.50940	1.00000					0.94094 (0.07174)	0.65000	1.00000			
GLM_1	0.93742 (0.04433)	0.74143	0.97345	0.00466	0.04789	98.12	105.38	0.93462 (0.05026)	0.66582	0.97345	0.00366	0.04525	0.04657
GLM_2	0.93742 (0.04535)	0.72021	0.98126	0.00446	0.04704	101.78	116.30	0.93409 (0.05250)	0.66769	0.98006	0.00372	0.04579	0.04642
GLM_3	0.93742 (0.04978)	0.72956	0.98071	0.00403	0.04313	101.91	120.06	0.93936 (0.05009)	0.64983	0.97949	0.00326	0.04046	0.04180
GLM_4	0.93742 (0.05019)	0.73330	0.98309	0.00393	0.04254	105.77	131.19	0.93907 (0.05061)	0.65902	0.98256	0.00324	0.04060	0.04157
GLM_5	0.93742 (0.05176)	0.65715	0.98975	0.00356	0.04109	112.75	152.70	0.93756 (0.05431)	0.71233	0.98512	0.00344	0.04172	0.04141
GLM_6	0.93742 (0.05193)	0.66093	0.98935	0.00353	0.04078	116.70	163.90	0.93761 (0.05476)	0.70516	0.98550	0.00345	0.04182	0.04130
OLS_1	0.93742 (0.04266)	0.81166	0.98597	0.00440	0.04595	− 718.03	− 710.77	0.93586 (0.04530)	0.78676	0.98597	0.00348	0.04429	0.04512
OLS_2	0.93742 (0.04338)	0.80481	1.00366	0.00434	0.04575	− 717.97	− 703.44	0.93579 (0.04651)	0.78818	1.00054	0.00348	0.04460	0.04518
OLS_3	0.93742 (0.04732)	0.81207	0.99522	0.00398	0.04245	− 739.82	− 721.67	0.93902 (0.04632)	0.78872	0.99389	0.00310	0.03981	0.04113
OLS_4	0.93742 (0.04762)	0.81562	1.00483	0.00396	0.04236	− 737.82	− 712.40	0.93884 (0.04693)	0.79050	1.00305	0.00310	0.03989	0.04113
OLS_5	0.93742 (0.04924)	0.76241	1.01474	0.00380	0.04218	− 741.10	− 701.16	0.93777 (0.05063)	0.77988	1.01377	0.00327	0.04050	0.04134
OLS_6	0.93742 (0.04935)	0.76394	1.01301	0.00379	0.04219	− 737.88	− 690.67	0.93778 (0.05071)	0.77576	1.01234	0.00326	0.04052	0.04136
Tobit_1	0.96369 (0.05818)	0.79220	1.02990	0.00533	0.05285	− 185.28	− 174.39	0.96156 (0.06177)	0.75824	1.02990	0.00428	0.05003	0.05144
Tobit_2	0.96348 (0.05855)	0.78748	1.04827	0.00526	0.05242	− 183.12	− 164.97	0.96136 (0.06271)	0.75878	1.04466	0.00431	0.05063	0.05153
Tobit_3	0.96319 (0.06452)	0.79299	1.04910	0.00496	0.05195	− 205.65	− 183.86	0.96575 (0.06269)	0.76089	1.04170	0.00405	0.04816	0.05006
Tobit_4	0.96307 (0.06456)	0.79396	1.05047	0.00492	0.05159	− 202.25	− 173.20	0.96549 (0.06296)	0.76257	1.04897	0.00403	0.04806	0.04983
Tobit_5	0.96304 (0.06735)	0.74322	1.07926	0.00482	0.05284	− 205.81	− 162.23	0.96387 (0.06842)	0.76384	1.04872	0.00434	0.05107	0.05196
Tobit_6	0.96300 (0.06734)	0.74480	1.07843	0.00481	0.05270	− 201.93	− 151.09	0.96385 (0.06838)	0.76169	1.05209	0.00433	0.05099	0.05185

A model had the best prediction accuracy for its functional form if it had the lowest MAE across the estimation and validation sample

AIC Akaike information criterion, *BIC* Bayesian information criterion, *GLM* generalised linear model, *MAE* mean absolute error, *Max.* maximum value, *Min.* minimum value, *MSE* mean square error, *OLS* ordinary least squares

Table 3 Model performance of the two best-fitting models

	Estimation sample			Validation sample		
	Observed	GLM_6	OLS_3	Observed	GLM_6	OLS_3
Mean	0.937419	0.937419	0.937419	0.940941	0.937612	0.939018
SD	0.078978	0.051926	0.047318	0.071737	0.054762	0.046323
CV	0.084251	0.055393	0.050477	0.076240	0.058406	0.049331
Min.	0.509400	0.660930	0.812068	0.650000	0.705160	0.788717
P_{25}	0.907600	0.910639	0.900076	0.912300	0.914908	0.905183
P_{50}	0.952100	0.957229	0.946303	0.952100	0.958496	0.950063
P_{75}	1.000000	0.978902	0.980433	1.000000	0.977276	0.977413
Max.	1.000000	0.989350	0.995221	1.000000	0.985504	0.993891
MSE		0.00353	0.00398		0.00345	0.00310
MAE		0.04078	0.04245		0.04182	0.03981
<0.03 AE (%)		53.40	51.61		55.89	53.17
<0.05 AE (%)		72.04	70.25		73.23	70.77
<0.10 AE (%)		92.27	90.32		91.20	93.31

CV coefficient of variation, GLM generalised linear model, MAE mean absolute error, Max. maximum value, Min. minimum value, MSE mean squared error, OLS ordinary least squares, P_{25} 25th percentile, P_{50} 50th percentile, P_{75} 75th percentile, SD standard deviation,<0.03 AE (%) percentage with absolute error below 0.03,<0.05 AE (%) percentage with absolute error below 0.05,<0.10 AE (%) percentage with absolute error below 0.10

Table 4 Distribution of errors by observed Child Health Utility Index–9 Dimension (CHU-9D) range

CHU-9D range	n	GLM_6		OLS_3	
		MSE	MAE	MSE	MAE
Estimation					
0.5 < value<0.6	3	0.09443	0.30095	0.11168	0.33058
0.6 < value<0.7	6	0.01873	0.12096	0.02823	0.16497
0.7 < value<0.8	11	0.00726	0.07674	0.00988	0.09602
0.8 < value<0.9	49	0.00301	0.04441	0.00259	0.04034
0.9 < value<1.0	111	0.00154	0.02853	0.00155	0.03015
Full health	102	0.00242	0.03847	0.00279	0.03899
Validation					
0.6 < value<0.7	3	0.05502	0.23049	0.05277	0.22693
0.7 < value<0.8	12	0.01185	0.09583	0.01376	0.10958
0.8 < value<0.9	47	0.00468	0.05691	0.00329	0.04316
0.9 < value<1.0	115	0.00187	0.03057	0.00131	0.02942
Full health	107	0.00223	0.03593	0.00237	0.03643

GLM generalised linear model, MAE mean absolute error, MSE mean squared error, OLS ordinary least squares

Table 5 Coefficients for the two best-fitting models

	GLM_6		OLS_3	
	Coefficient	SE	Coefficient	SE
PedsQL™ PF squared	0.000162	0.000103		
PedsQL™ EF squared	0.000477***	0.000127		
PedsQL™ SF squared	- 0.000040	0.000145		
PedsQL™ FU squared	- 0.0001646	0.000101		
PedsQL™ PF 9 EF	- 0.000110	0.000147		
PedsQL™ PF 9 SF	- 0.000114	0.000173		
PedsQL™ PF 9 FU	0.000037	0.000143		
PedsQL™ EF 9 SF	- 0.000246	0.000209		
PedsQL™ EF 9 FU	- 0.000116	0.000167		
PedsQL™ SF 9 FU	0.000436***	0.000130		
PedsQL™ PF			0.0007133*	0.000297
PedsQL™ EF			0.0016477***	0.000228
PedsQL™ SF			- 0.00011	0.000383
PedsQL™ FU			0.000261	0.000276
Age	0.0279345	0.039717		
Sex	- 0.0546336	0.146341		
Constant	0.7135215	0.399623	0.7422337***	0.028841

EF emotional functioning, FU school functioning, GLM generalised linear model, OLS ordinary least squares, PedsQL™ Paediatric Quality of Life Inventory generic core scale, PF physical functioning, SE standard error, SF social functioning

* P<0.05, *** P<0.001

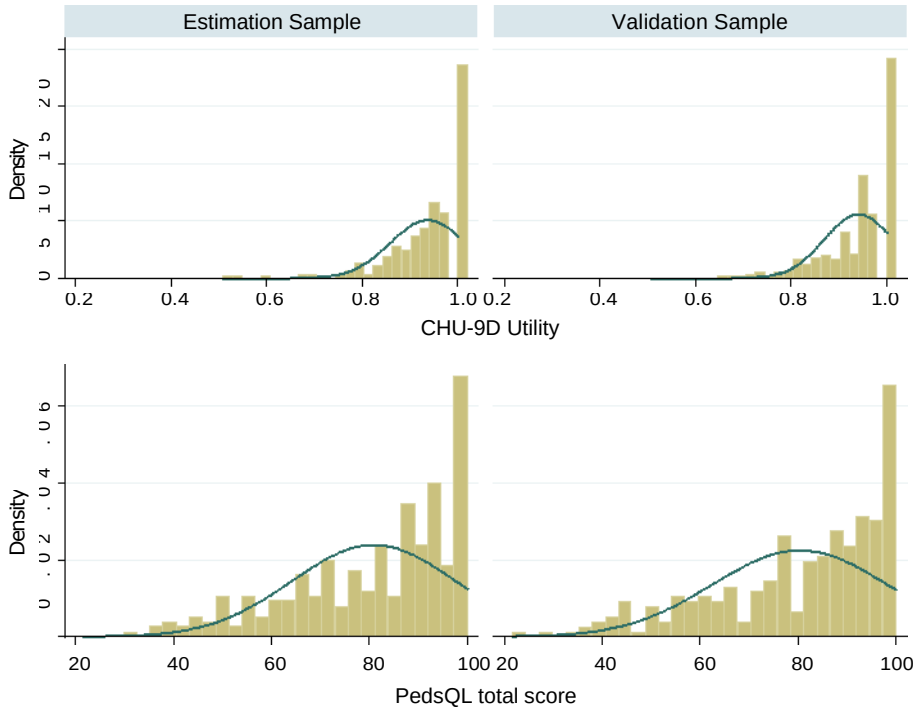
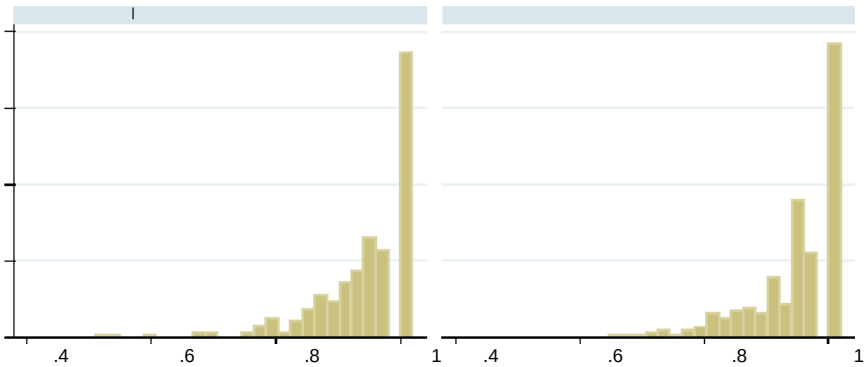
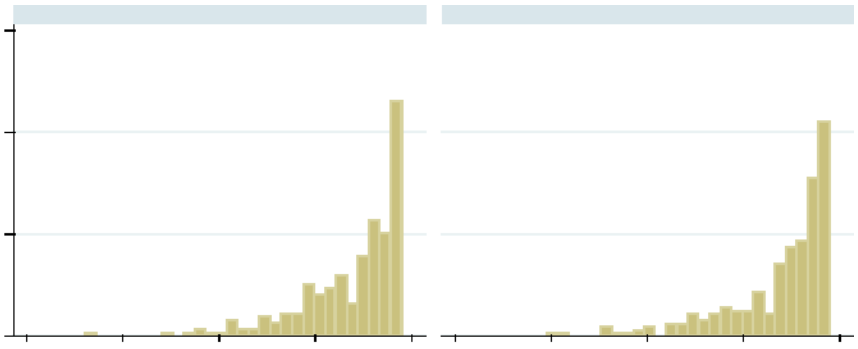
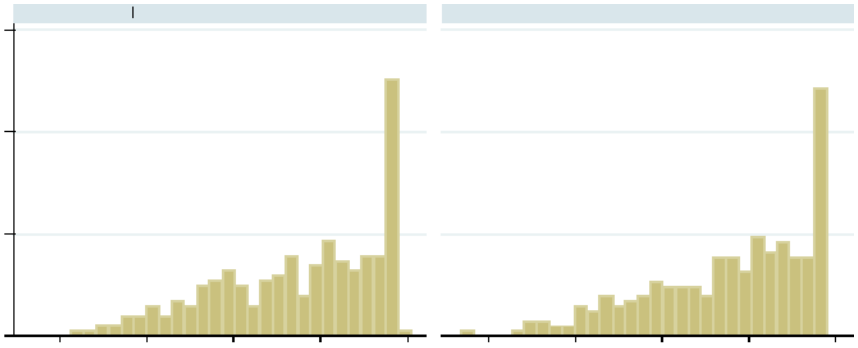


Fig. 1 Kernel density plots of Child Health Utility Index–9 Dimension (CHU-9D) utilities and Paediatric Quality of Life Inventory generic core scale (PedsQL™) total scores for the estimation and validation data



Distribution of observed CHU-9D score
Distribution of predicted CHU9D score (GLM_6)





Distribution of predicted CHU-9D score (OLS_3)

Table 2 Performance of Paediatric Quality of Life Inventory generic core scale (PedsQL^{TM}) to Child Health Utility Index–9 Dimension (CHU-9D) index score models in the estimation and validation samples

	Estimation sample							Validation sample					Average MAE across samples
	Mean	Min.	Max.	MSE	MAE	AIC	BIC	Mean	Min.	Max.	MSE	MAE	
Observed	0.93742 (0.07898)	0.50940	1.00000					0.94094 (0.07174)	0.65000	1.00000			
GLM_1	0.93742 (0.04433)	0.74143	0.97345	0.00466	0.04789	98.12	105.38	0.93462 (0.05026)	0.66582	0.97345	0.00366	0.04525	0.04657
GLM_2	0.93742 (0.04535)	0.72021	0.98126	0.00446	0.04704	101.78	116.30	0.93409 (0.05250)	0.66769	0.98006	0.00372	0.04579	0.04642
GLM_3	0.93742 (0.04978)	0.72956	0.98071	0.00403	0.04313	101.91	120.06	0.93936 (0.05009)	0.64983	0.97949	0.00326	0.04046	0.04180
GLM_4	0.93742 (0.05019)	0.73330	0.98309	0.00393	0.04254	105.77	131.19	0.93907 (0.05061)	0.65902	0.98256	0.00324	0.04060	0.04157
GLM_5	0.93742 (0.05176)	0.65715	0.98975	0.00356	0.04109	112.75	152.70	0.93756 (0.05431)	0.71233	0.98512	0.00344	0.04172	0.04141
GLM_6	0.93742 (0.05193)	0.66093	0.98935	0.00353	0.04078	116.70	163.90	0.93761 (0.05476)	0.70516	0.98550	0.00345	0.04182	0.04130
OLS_1	0.93742 (0.04266)	0.81166	0.98597	0.00440	0.04595	− 718.03	− 710.77	0.93586 (0.04530)	0.78676	0.98597	0.00348	0.04429	0.04512
OLS_2	0.93742 (0.04338)	0.80481	1.00366	0.00434	0.04575	− 717.97	− 703.44	0.93579 (0.04651)	0.78818	1.00054	0.00348	0.04460	0.04518
OLS_3	0.93742 (0.04732)	0.81207	0.99522	0.00398	0.04245	− 739.82	− 721.67	0.93902 (0.04632)	0.78872	0.99389	0.00310	0.03981	0.04113
OLS_4	0.93742 (0.04762)	0.81562	1.00483	0.00396	0.04236	− 737.82	− 712.40	0.93884 (0.04693)	0.79050	1.00305	0.00310	0.03989	0.04113
OLS_5	0.93742 (0.04924)	0.76241	1.01474	0.00380	0.04218	− 741.10	− 701.16	0.93777 (0.05063)	0.77988	1.01377	0.00327	0.04050	0.04134
OLS_6	0.93742 (0.04935)	0.76394	1.01301	0.00379	0.04219	− 737.88	− 690.67	0.93778 (0.05071)	0.77576	1.01234	0.00326	0.04052	0.04136
Tobit_1	0.96369 (0.05818)	0.79220	1.02990	0.00533	0.05285	− 185.28	− 174.39	0.96156 (0.06177)	0.75824	1.02990	0.00428	0.05003	0.05144
Tobit_2	0.96348 (0.05855)	0.78748	1.04827	0.00526	0.05242	− 183.12	− 164.97	0.96136 (0.06271)	0.75878	1.04466	0.00431	0.05063	0.05153
Tobit_3	0.96319 (0.06452)	0.79299	1.04910	0.00496	0.05195	− 205.65	− 183.86	0.96575 (0.06269)	0.76089	1.04170	0.00405	0.04816	0.05006
Tobit_4	0.96307 (0.06456)	0.79396	1.05047	0.00492	0.05159	− 202.25	− 173.20	0.96549 (0.06296)	0.76257	1.04897	0.00403	0.04806	0.04983
Tobit_5	0.96304 (0.06735)	0.74322	1.07926	0.00482	0.05284	− 205.81	− 162.23	0.96387 (0.06842)	0.76384	1.04872	0.00434	0.05107	0.05196
Tobit_6	0.96300 (0.06734)	0.74480	1.07843	0.00481	0.05270	− 201.93	− 151.09	0.96385 (0.06838)	0.76169	1.05209	0.00433	0.05099	0.05185

A model had the best prediction accuracy for its functional form if it had the lowest MAE across the estimation and validation sample

AIC Akaike information criterion, *BIC* Bayesian information criterion, *GLM* generalised linear model, *MAE* mean absolute error, *Max.* maximum value, *Min.* minimum value, *MSE* mean square error, *OLS* ordinary least squares