Estimating CHU-9D Utility Scores from the WAItE: A Mapping Algorithm for Economic Evaluation

Tomos Robinson, PhD and Yemi Oluboyede, PhD

Health Economics Group, Institute of Health and Society, Newcastle University, Newcastle-upon-Tyne, UK

Keywords: adolescence, obesity, WAItE, CHU-9D, mapping, health related quality of life

Text pages: 18

Tables: 5

Figures: 1

Word Count: 3992

Acknowledgements: This study was funded by the Health Foundation, project reference C0150R4035. The publication of the study results was not contingent on the sponsor's approval or censorship of the manuscript. This study was approved by the Faculty of Medical Sciences Research Ethics Committee, part of Newcastle University's Research Ethics Committee (project reference 1262/12643). Both authors report no conflict of interests. Data and Stata do-files available from the authors on request.

Abstract

Background: The **W**eight-Specific **A**dolescent **I**nstrumen**t** for Economic **E**valuation (WAItE) is a new condition-specific patient reported outcome measure that incorporates the views of adolescents in assessing the impact of above healthy weight status on key aspects of their lives. Presently it is not possible to use the WAItE to calculate quality adjusted life years (QALYs) for cost-utility analysis (CUA), given that utility scores are not available for health states described by the WAItE.

Objective: This paper examines different regression models for estimating Child Health Utility 9 Dimension (CHU-9D) utility scores from the WAItE for the purpose of calculating QALYs to inform CUA.

Methods: The WAItE and CHU-9D were completed by a sample of 975 adolescents. Nine regression models were estimated: Ordinary Least Squares, Tobit, Censored Least Absolute Deviations, Two-Part, Generalised Linear Model, robust MM-estimator, Beta-Binomial, Finite Mixture Models, and Ordered Logistic Regression. The mean absolute error (MAE) and mean squared error (MSE) were used to assess the predictive ability of the models.

Results: The robust MM-estimator with stepwise-selected WAItE item scores as explanatory variables had the best predictive accuracy.

Conclusions: Condition-specific tools have been shown to be more sensitive to changes that are important to the population for which they have been developed for. The mapping algorithm developed in this study facilitates the estimation of health-state utilities necessary for undertaking CUA within clinical studies that have only collected the WAItE.

1. Introduction

Increasingly, health benefits from public health interventions are being captured as health-related quality of life (HRQoL), a multidimensional construct that measures the impact of health or disease on physical or psychological functioning [1]. HRQoL can be measured directly (through a Standard Gamble or Time Trade Off exercise) or indirectly through 'generic' or 'condition-specific' HRQoL instruments. HRQoL instruments can have several advantages over direct valuation, such as being more reliable and less complex to administer [2]. 'Generic' instruments can be seen to assess the core dimensions of health that are relevant to all conditions, and allow comparisons of health benefit across both interventions and conditions. In contrast, 'condition-specific' instruments focus on the most important domains of HRQoL affected by the specific condition, and are usually more sensitive to disease-specific improvements in HRQoL.

HRQoL instruments can be further classified as 'preference based' or 'non-preference based'. Preference-based measures differ from non-preference-based measures in the way the scoring algorithms are derived, in that they are estimated from the values that people place on different aspects of health rather than a simple summative scoring procedure [3]. Only preference based instruments can be used to generate quality adjusted life years (QALYs) for use in CUA [4].

Within paediatric medicine, a number of weight-specific HRQoL instruments have been developed for use in adolescence, including the KINDL-obesity module [5], the Impact of Weight on Quality of Life-Kids [6], the Moorehead-Ardelt Quality of Life Questionnaire [7], Sizing Me Up [8], Youth Quality of Life-Weight [9] and Oxford Paediatric Obesity Instrument [10]. However, none of these instruments can be considered preference-based, nor were they designed for this purpose. In response to this, a weight-specific HRQoL instrument for adolescents, the **W**eight-Specific **A**dolescent **I**nstrumen**t** for Economic **E**valuation (WAItE), was developed. The WAItE is a brief 7-item measure incorporating the views and experiences of adolescents aged 11-18 years [11]. The WAItE was developed to be suitable for undertaking a valuation study, in order to operationalise the calculation of QALYs [4].

Although specifically designed to be preference based, a valuation study for the WAItE has yet to be carried out. In its current form, the WAItE can be used to undertake a cost-effectiveness analysis of weight management interventions, as WAItE total scores can be used to assess the difference in the average scores between two or more groups. However,

measuring outcomes in this way makes it difficult to compare outcomes between different conditions. Overcoming this comparability problem involves generating a scoring tariff for the health states described by the WAItE. This can be done directly via a valuation study or indirectly using a mapping algorithm. Directly assigning utility values to WAItE health states through a stand-alone valuation study is considered the 'gold standard' method [4], however such a valuation study can be costly to carry out. Alternatively, a mapping algorithm may be used to predict utility scores from responses to a non-preference-based instrument. This algorithm can be seen to reflect the statistical relationship between the preference and non-preference-based instruments, using responses from a population whose responses to both instruments have been collected simultaneously [12].

Although previous studies have conducted a mapping exercise in the adolescent population [13] and mapped weight-specific measures of HRQoL to preference based measures in the adult population [14], no study has mapped a weight-specific measure of HRQoL onto a generic preference based measure in the adolescent population. Given this, the aim of this study was to develop a mapping algorithm from the WAItE to the Child Health Utility 9D (CHU-9D), a generic preference-based measure of HRQoL in the paediatric population [15], therefore facilitating weight-specific adolescent HRQoL to be measured in the context of CUA.

2 Methods

2.1 Study Design

To develop a mapping algorithm from the WAItE to the CHU-9D, an online survey was developed with the company Survey Sampling International (SSI) for administration to a sample of adolescents aged 11-18 years residing in the United Kingdom (UK). The first section of the survey comprised of a series of sociodemographic questions, including age, gender, ethnicity, self-reported height and weight and self-assessed health. The second section comprised the WAItE and CHU-9D.

The target sample was 1000, similar to the median sample (1167) found in a recent review of the mapping literature [16]. All UK adults from the SSI panel with children between the ages of 11-15 (around 15,000) were identified and approached to complete the survey from SSI's participant panel. The 11–15 year old participants were then able to complete the survey given the consent of their guardian. Furthermore, around 2,500 16 to 18 year olds were directly invited to complete the survey by SSI. The survey was left open until 1,000

participants had completed the survey. Respondents to SSI surveys receive an average of £0.30 per 5 minute interview. The median time to complete the survey was 6 minutes. A quota on weight status was stipulated initially, with an aim of having the 1,000 respondents split equally between three weight groups: 'normal', 'overweight' and 'obese'. This initial quota was done to generate a sample more representative of the individuals who may benefit from a weight management interventions, who in this case can be considered the population of interest. The survey was approved by Newcastle University's Faculty of Medical Sciences Research Ethics Committee (project reference 1262/12643).

2.2 Outcome Measures

WAItE

The WAItE has seven individual items (relating to tiredness, walking, participation in sports, concentration, embarrassment, unhappiness and being treated differently), with a five-point Likert scale representing the increasing degrees of severity (ranging from 'never' to 'always'). The WAItE total score is calculated by summing the answers of the seven items, and is scored between 7 and 35 [11]. In analysis, the WAItE total score was reverse coded so that a higher WAItE total score indicated a higher quality of life, in line with the CHU-9D.

CHU-9D

The CHU-9D has nine individual items (related to being worried, sad, in pain, tired, annoyed, schoolwork, sleep, daily routine, and ability to join in activities), each with a five-point Likert scale representing increasing degrees of severity (ranging from, for example, 'I don't feel worried today' to 'I feel very worried today'). The instrument has been validated for use in both younger children and adolescent populations [15]. We used the scoring algorithm based on the preferences of the UK adult general population, meaning that the utility scores are bound between 0.33 and 1.

2.3 Statistical and Econometric Analysis

Participant characteristics were summarised as means for continuous variables and frequencies for categorical variables. Normality of distribution was tested using the Shapiro-Wilks test [17], and the correlation between the WAItE and the CHU-9D was estimated using the Spearman correlation coefficient [18]. This study was conducted in accordance with the MAPS (MApping onto Preference-based measures reporting Standards) checklist [19] and the recently published ISPOR good practice for outcomes research task report [20]. We first

estimated direct mapping models by regressing the WAItE total and item scores directly onto the CHU-9D utility scores. An indirect mapping approach was further considered, in which the response levels of each CHU-9D item was estimated, before being summed together. This method is seen to preserve the main design features of the target instrument [21]. To improve predictive performance, age and gender were also included in the econometric specifications, as well as quadratic terms to control for non-linear relationships. The four specifications can be displayed as:

$$\mathrm{CHU9D_{Utility}} = \alpha_0 + \beta_1 * \mathrm{WAItE_{Total}} + \beta_2 * \mathrm{WAItE_{Total}}^2 + \delta_1 * \mathrm{AGE} + \delta_2 * \mathrm{GEN} + \varepsilon_i \quad (1)$$

$$\mathrm{CHU9D_{Utility}} = \alpha_0 + \sum_{i=1}^{k} \beta_i * \mathrm{WAItE_{Item}} + \sum_{i=1}^{k} * \mathrm{WAItE_{Item}}^2 + \delta_1 * \mathrm{AGE} + \delta_2 * \mathrm{GEN} + \varepsilon_i \quad (2)$$

$$\mathrm{CHU9D_{Item}} = \alpha_0 + \beta_1 * \mathrm{WAItE_{Total}} + \beta_2 * \mathrm{WAItE_{Total}}^2 + \delta_1 * \mathrm{AGE} + \delta_2 * \mathrm{GEN} + \varepsilon_i \quad (3)$$

$$\mathrm{CHU9D_{Item}} = \alpha_0 + \sum_{i=1}^{k} \beta_i * \mathrm{WAItE_{Item}} + \sum_{i=1}^{k} * \mathrm{WAItE_{Item}}^2 + \delta_1 * \mathrm{AGE} + \delta_2 * \mathrm{GEN} + \varepsilon_i \quad (4)$$

$\mathrm{CHU9D_{Utility}}$ represents the total CHU-9D utility score and $\mathrm{CHU9D_{Item}}$ represents one of the nine CHU-9D items. Similarly, $\mathrm{WAItE_{Total}}$ represents the total WAItE score, and $\mathrm{WAItE_{Item}}$ represents the selected WAItE items based using the stepwise regression technique, with *k* representing the number of selected items. The significance level for statistical inference was 5%.

Numerous econometric techniques have been used to estimate direct mapping models, with the most common being Ordinary Least Squares (OLS), which has been found to work well in several previous mapping studies [22, 23]. However, when the distribution of the target instrument is non-normal, other regression models are seen as being more appropriate [24].

The Tobit model takes into account the fact that there may be a mass of observations at either the upper or lower bound of the distribution [25], while the Censored Least Absolute Deviations (CLAD) model [26] takes into account both these bounding issues and heteroscedasticity. A two-part logit-OLS regression can also be used to account for the large proportion of respondents reporting a utility value of 1. First, a logistic regression model is estimated to predict which of the participants have a utility value of 1. Second, an OLS model predicting CHU-9D utility scores is estimated for those participants who have a utility score below 1. Utility predictions for the two-part model are then estimated using the expected value method, which can be displayed as:

$$\Pr(U = 1) + (1 - \Pr(U = 1)) * U \quad (5)$$

$\Pr(U = 1)$ represents the predicted probability of being in full health from the logit model, and U represents the predicted utility conditional on having imperfect health, calculated from the OLS model.

The Generalised Linear Model (GLM) allows for skewed distributions of the dependent variable. The Pregibon link test and Hosmer-Lemeshow test were used to guide the choice of the most appropriate GLM distribution and link function [27]. The robust MM-estimator is designed to account for heteroscedasticity and the presence of outliers, and has been found to have good performance in the context of mapping studies [28]. The Beta-Binomial (BB) model [29] is robust to skewness and can estimate both unimodal and bimodal utilities, and has been shown to be superior to OLS regression in the context of mapping [30]. Finite Mixture Models (FMMs) are able to combine two or more probability distribution functions, and can therefore handle the multimodal distributions of data that commonly characterise HRQoL data. Finally, the ordered logit model (OLOGIT) can be used in response mapping models to predict the responses to the individual CHU-9D items.

In line with the previous literature [12, 31], mean squared error (MSE) and mean absolute error (MAE) were used to measure the goodness of fit of the models. The MSE is equal to the mean of squared errors between the observed CHU-9D utility score and the CHU-9D utility score predicted from the model, whereas the MAE is equal to the mean of the absolute differences between observed and predicted utility scores. Given the guidance from the previous literature [32], more weight was put on the MAE than the MSE when the two

statistics conflicted. We also calculated the percentage of the response within 0.03 and 0.05 of the true utility score.

2.4 Validation

Model performance was assessed using an internal dataset and two methods. First, we implemented the cross-validation (or '*k*-fold') method, which has successfully been used in several other studies in the mapping literature [33, 34]. This approach first involves randomly dividing the sample into five groups. In each instance, 80% of the sample were assigned to the estimation sample, with the remaining 20% being assigned to the validation sample. Each time, the estimation sample was used to generate the algorithm, which was then used to estimate utility values in the validation sample. This process was repeated five times. Second, we implemented the 'random samples' validation method, which has also been used successfully in the mapping literature [31]. In this case, predictive models estimated using the entire dataset were validated on random samples of 100, 300 and 500 individuals from the sample. The regression model that performed the best in both validation techniques was chosen as optimal.

3. Results

3.1 Sample Characteristics

1000 participants completed the survey. 25 individuals were excluded from analysis due to infeasible BMI values, meaning the final sample size was 975. Table 1 summarises the characteristics of the respondents. The mean age of the respondents was 15.4 years, and 50.6% of the respondents were female. From the calculated BMI values, 36.2% of the sample were classified as having a 'normal' weight, 22.7% of the sample were classified as 'overweight' and 41.1% of the sample were classified as 'obese'. 57.4% of the adolescents reported themselves as having some form illness or disability. Gender was not statistically significant in any regression model, and was left out of the final model specification.

[TABLE 1 ABOUT HERE]

The mean WAItE score was 25.39, while the median score was 26. The mean utility of the CHU-9D was 0.81, while the median utility was 0.84. The distributions of both the WAItE

and the CHU-9D are displayed in the supplementary materials. The distribution of the CHU-9D utilities was negatively skewed and showed evidence of a ceiling effect, with around 20% of the respondents having a utility of over 0.95 and 10.1% of the respondents reporting a utility value of 1. As indicated by the Shapiro-Wilk test, both instruments were non-normally distributed. Figure 1 displays a scatter plot displaying the relationship between the WAItE total score and the CHU-9D utilities. The positive association observed from the scatter plot, was corroborated by a strong correlation between the measures (Spearman rank correlation coefficient of 0.729).

[FIGURE 1 ABOUT HERE]

3.2 Prediction of CHU-9D Utility Scores

Table 2 presents the results from the nine regression models and two model specifications in the full estimation sample. The Pearson Correlation test [36], Pregibon link test [37] and modified Hosmer-Lemeshow test [38] indicated that the most appropriate GLM distribution was the gamma distribution with a log link. As it was not obvious from a visual inspection of the CHU-9D utility values how many modal components the CHU-9D had, FMMs with 2, 3 and 4 components were estimated, with the model with the smallest Bayesian Information Criteria chosen as the final model.

Table 2 also presents the results from the key goodness-of-fit statistics for the different regression models in the full estimation sample. Although the OLS, GLM and Two-Part models in both specifications accurately predicted the mean utility score (0.8056), the majority of the remaining models overestimated both the mean utility score. All models overestimated the lower limit of the utility score (0.3454).

[TABLE 2 ABOUT HERE]

In Model Specification 1 (which used the WAItE total score to predict the CHU-9D utilities), the robust MM-estimator had the lowest MAE (0.0765), while the OLS, Two-Part, GLM and FMM had the joint lowest MSE (0.0102). The CLAD model had both the most absolute

differences < 0.03 (29.8%), and the most absolute differences < 0.05 (44.9%). In Model Specification 2 (which used the individual WAItE items to predict CHU-9D utilities), the CLAD model had the lowest MAE (0.0751), while the OLS, Two-Part, GLM and FMM had the joint lowest MSE (0.0100). The robust MM-estimator had both the most absolute differences < 0.03 (30.4%), and the most absolute differences < 0.05 (48.1%).

3.3 Validation

The results from the the full estimation sample were validated using the *k*-fold and random samples validation methods described in sub-section 2.4. Tables 3 and 4 show the results from this validation analysis.

[TABLES 3 and 4 ABOUT HERE]

In Model Specification 1, the GLM model accurately predicted the mean CHU-9D utility score the most times, while in Model Specification 2 the Two-Part model accurately predicted the mean CHU-9D utility scores the most times. Overall, the item level specifications (Model Specification 2) performed better than the total score specifications (Model Specification 1) in terms of both MAE and MSE. The MAE ranged from 0.0680 (robust MM-estimator in Model Specification 2 and the third random sample) to 0.1201 (OLOGIT model in Model Specification 1 and first random sample), while the MSE ranged from 0.0088 (Two-Part, GLM and BB models in Model Specification 2 and third random sample) to 0.0261 (OLOGIT model in Model Specification 1 and the first random sample). The results in Tables 3 and 4 indicate that when all regression models were assessed using MAE, the robust MM-estimator in Model Specification 2 had the best predictive ability (best performing model in all three random samples), while the Two-Part model in Model Specification 2 was the best in terms of MSE (best performing model in the *k*-fold validation and the first and second random samples).

3.4 Best Performing Models

On the basis of their performance in the full estimation sample and all four validation samples, the Two-Part Model in Model Specification 2 and the robust MM-estimator in

Model Specification 2 were chosen as the two best performing models. Detailed performance statistics for these models are presented in Tables 5.

[TABLE 5 ABOUT HERE]

When the two best performing models were compared in terms of the MAE, the robust MM-estimator had a lowest value in the full sample and all four validation samples. When compared in terms of the MSE, the Two-Part model had the lowest value in the full sample, the *k*-fold validation sample and random sample 1. The robust MM-estimator had the lowest MSE value in the random sample 3. In random sample 2, both estimators had a MSE of 0.0093. The Two-Part model predicted the observed mean CHU-9D utility better than the robust MM-estimator in the full sample, *k*-fold validation sample and random sample 1, whereas the robust MM-estimator predicted the observed mean CHU-9D utility better in random sample 2 and random sample 3. The robust MM-estimator had the highest number of absolute differences between 0.03 and 0.05 of the mean in the full sample and all four validation samples. On the basis of these results, we propose using the robust MM-estimator with individual WAItE item scores as explanatory variables to predict CHU-9D utility scores from the WAItE.

The distribution of the MAE and MSE for the two best performing models across the range of CHU-9D utility scores was also examined, with these results displayed in the supplementary materials. For the Two-Part model, the smallest error in terms of MAE and MSE was found in the 0.7-0.8 range. For the robust MM-estimator, the smallest error was found in the 0.9-1 range. For both the Two-Part and robust MM-estimators, the largest errors were found in the sub-sample of individuals with a CHU-9D utility score below 0.7.

3.5 Mapping Equations

The regression model coefficients for predicting the CHU-9D utility scores from the WAItE using the two best performing regression models are displayed in the supplementary materials. Based on the findings from the results from section 3.4, the optimal algorithm for calculating CHU-9D utility scores from the can be shown algebraically as:

$$0.4798967 + (0.0107895 * WAItE\_Tired) + (0.015773 * WAItE\_Walking) + (0.0231428 * WAItE\_Concentration) + (0.0131602 * WAItE\_Embarrassed) + (0.0288246 * WAItE\_Unhappy) + (0.0244563 * WAItE\_Treated\ Different) + (-0.006123 * Age)$$

For example, an 11 year-old adolescent who reports 'never' to all six of the included WAItE items would have an estimated CHU-9D utility value of 0.9932757:

$$0.4798967 + (0.0107895 * 5) + (0.015773 * 5) + (0.0231428 * 5) + (0.0131602 * 5) + (0.0288246 * 5) + (0.0244563 * 5) + (-0.006123 * 11) = 0.9932757$$

Whereas, an 18 year old adolescent who reports 'always' to all six of the included WAItE items would have an estimated CHU-9D utility value of 0.4858994:

$$0.4798967 + (0.0107895 * 1) + (0.015773 * 1) + (0.0231428 * 1) + (0.0131602 * 1) + (0.0288246 * 1) + (0.0244563 * 1) + (-0.006123 * 18) = 0.4858994$$

The variance-covariance matrix for this model and a plot of the predicted CHU-9D utilities against the observed utilities can be found in the supplementary materials.

4. Discussion

The purpose of this study was to develop a mapping algorithm to estimate CHU-9D utility scores from the WAItE, enabling the derivation of weight-specific health state utilities for adolescents. This algorithm allows the costs and benefits of weight management interventions for adolescents to be compared to interventions undertaken in different settings, and therefore promotes allocative efficiency in health care decision making.

In accordance with recent guidance [19, 20], numerous regression models were estimated to establish the most appropriate algorithm for predicating CHU-9D utilities from the WAItE. The robust MM-estimator using stepwise selected WAItE item scores as explanatory variables was found to be the most accurate. The values of the MAE and MSE statistics obtained from this preferred model can be considered low compared to similar studies in the literature [31, 39, 40]. However, it should be noted that several other regression models, in particular the Two-Part, GLM and OLS models, generated similar estimates to those from the robust MM-estimator. The response mapping models (using ordered logits) performed poorly, with this poor performance likely due to the small number of observations in several CHU-9D and WAItE items.

Similar to several other studies in the literature, almost all the estimators overestimated the lower bound of the CHU-9D utility, and some over predicted the upper bound of 1. Although this over prediction is a difficult issue to circumvent, several studies [41, 42] have argued that predicted values outside the theoretical limit can be dealt with by truncating them to the boundary value. However, further analysis (available on request) displayed that assigning values above 1 a value of 1 made no difference to the results.

There are several strengths to this study. First, we estimated an exhaustive number of regression models to determine the most appropriate algorithm. Second, as displayed in Figure 1, the target and source instruments overlap adequately in the overall concepts measured. A strong correlation between the target and source instruments has previously been found to be an important determinant of a successful mapping analysis [43].

However, there are also some potential shortcomings. First, our data was gathered was from an online survey. Although this allowed us to achieve a higher sample size that would have been possible using face to face interviews, it is possible that the children did not complete the survey themselves. However, given the increasing use of internet surveys and the relatively large sample size, we would assume that this recruitment method is acceptable, and that all recruitment methods have individual shortcomings. Second, alike several other mapping studies [33, 34], an external dataset was unavailable for use, and therefore we used in-sample validation. Future research should test the algorithm generated in this study in an external dataset to ensure validity.

Third, a quota was applied to the estimation sample to oversample those adolescents whose BMI fell into the ‘overweight’ and ‘obese’ weight status categories, and therefore we did not

have an estimation sample that was nationally representative according to weight status. Due to the purposive nature of the sampling, the mapping algorithm developed in this study is specifically aimed to represent the 'normal' 'overweight' and 'obese' weight categories equally, in order to be more representative of those individuals who may be expected to benefit from a weight management intervention.

Fourth, potentially due to the nature of the sample, an unusually high number of respondents (57.2%) reported themselves as having some form illness or disability, compared to a national average of around 12%, as reported in the 2015 Labour Force Survey [44]. Finally, it must be noted that the generated algorithm had a higher level of error for those adolescents with a lower HRQOL.

5. Conclusion

When a preference based instrument is not included in a study, mapping from a non-preference based instruments to obtain health state utilities can be seen as the second best alternative. This study developed an algorithm to map CHU-9D utility scores from the WAItE, a newly developed measure of weight-specific HRQoL for use in adolescents, and is the first study to map from a weight-specific HRQoL measure to a generic preference based measure in the adolescent population. Our results show that it is possible to predict CHU-9D utility scores from the WAItE, with the best results obtained when utilising the robust MM-estimator with stepwise selected WAItE item scores as predictors. This algorithm may be applied for the prediction of CHU-9D utilities, thereby facilitating the calculation of QALYs for assessing the relative cost-effectiveness of public health interventions and weight loss programmes specifically targeted at the adolescent population.

References

1. Fontaine KR, Barofsky I. Obesity and health-related quality of life. Obesity Reviews. 2001;2(3):173-82.
2. Brazier, J., Deverill, M., & Green, C. A review of the use of health status measures in economic evaluation. Journal Health Serv Res Policy. 1999; 4(3): 174-184.
3. Neumann PJ, Goldie SJ, Weinstein MC. Preference-based measures in economic evaluation in health care. Anu Rev Public Health. 2000;21(1):587-611.
4. Brazier, J., Ratcliffe, J., Saloman, J., Tsuchiya, A. Measuring and valuing health benefits for economic evaluation (2nd ed). Oxford University Press, 2017
5. Ravens-Sieberer U, Redegeld M, Bullinger M. Quality of life after in-patient rehabilitation in children with obesity. Int J Obes Relat Metab Disord . 2001;25 (S1):S63
6. Kolotkin RL, Zeller M, Modi AC, Samsa GP, Quinlan NP, Yanovski JA, et al. Assessing weight-related quality of life in adolescents. Obesity. 2006;14(3):448-57.
7. Moorehead MK, Ardelt-Gattinger E, Lechner H, Oria HE. The validation of the Moorehead-Ardelt quality of life questionnaire II. Obesity Surgery. 2003;13(5):684-92.
8. Zeller MH, Modi AC. Development and Initial Validation of an Obesity-specific Quality-of-life Measure for Children: Sizing Me Up. Obesity. 2009;17(6):1171-7.
9. Morales LS, Edwards TC, Flores Y, Barr L, Patrick DL. Measurement properties of a multicultural weight-specific quality-of-life instrument for children and adolescents. Qual Life Res . 2011;20(2):215-24.
10. Doyle S. Development of a new health related quality of life instrument for use in paediatric obesity. Obesity Reviews. 2011;12:77-8.
11. Oluboyede Y, Hulme C, Hill A. Development and refinement of the WAItE: a new obesity-specific quality of life measure for adolescents. Qual Life Res. 2017:26: 2025-2039.

12. Lambe T, Frew E, Ives NJ, et al. 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. PharmacoEconomics. 2017:1-15.
13. 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
14. Brazier JE, Kolotkin RL, Crosby RD, Williams GR. Estimating a preference-based single index for the Impact of Weight on Quality of Life-Lite (IWQOL-Lite) instrument from the SF-6D. Value Health. 2004;7(4):490-8.
15. Stevens K. Valuation of the child health utility 9D index. Pharmacoeconomics. 2012;30(8):729-47.
16. 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):151.
17. Shapiro SS, Wilk MB. An analysis of variance test for normality (complete samples). Biometrika. 1965;52(3/4):591-611.
18. Spearman C. The proof and measurement of association between two things. The Am J Psychol . 1904;15(1):72-101.
19. 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. Health Qual Life Outcomes. 2015;13(1):106.
20. Wailoo AJ, Hernandez-Alava M, Manca A, Mejia A, Ray J, Crawford B, et al. Mapping to estimate health-state utility from non–preference-based outcome measures: an ISPOR good practices for outcomes research task force report. Value in Health. 2017;20(1):18-27.
21. 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.
22. Mortimer D, Segal L. Comparing the incomparable? A systematic review of competing techniques for converting descriptive measures of health status into QALY-weights. Med Decis Making. 2008;28(1):66-89.
23. 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.

24. 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.
25. Tobin J. Estimation of relationships for limited dependent variables. Econometrica. 1958:24-36.
26. Powell JL. Least absolute deviations estimation for the censored regression model. J Econom. 1984;25(3):303-25.
27. Manning WG, Mullahy J. Estimating log models: to transform or not to transform? J Health Econ. 2001;20(4):461-94.
28. Chen G, Khan MA, Iezzi A, Ratcliffe J, Richardson J. Mapping between 6 multiattribute utility instruments. Med Decis Making. 2016;36(2):160-75.
29. Ospina R, Ferrari SLP. A general class of zero-or-one inflated beta regression models. Comput Stat Data Anal. 2012;56(6):1609-23.
30. 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.
31. Mpundu-Kaambwa C, Chen G, Russo R, Stevens K, Petersen KD, Ratcliffe J. Mapping CHU9D Utility Scores from the PedsQLTM 4.0 SF-15. PharmacoEconomics. 2017;35(4):453-67.
32. Hyndman RJ, Koehler AB. Another look at measures of forecast accuracy. Int J Forecasting. 2006;22(4):679-88.
33. 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.
34. Wong CKH, Lam CLK, Rowen D, McGhee SM, Ma K-P, Law W-L, 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.
35. Cole, T.J., Freeman, J.V. and Preece, M.A. Establishing a standard definition for child overweight and obesity worldwide: international survey. BMJ. 2000;320:1-6.
36. Pearson ES, Please NW. Relation between the shape of population distribution and the robustness of four simple test statistics. Biometrika. 1975;62(2):223-41.
37. Pregibon D. Goodness of link tests for generalized linear models. Applied Statistics. 1980:15-4.

38. Hosmer Jr DW, Lemeshow S, Sturdivant RX. Applied logistic regression: John Wiley & Sons; 2013.
39. 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.
40. Rivero-Arias O, Ouellet M, Gray A, Wolstenholme J, et al. Mapping the modified Rankin scale (mRS) measurement into the generic EuroQol (EQ-5D) health outcome. Med Decis Making . 2010;30(3):341-54.
41. Sullivan PW, Ghushchyan V. Mapping the EQ-5D index from the SF-12: US general population preferences in a nationally representative sample. Med Decis Making. 2006;26(4):401-9.
42. 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(7):801-13.
43. Chuang L-H, Whitehead SJ. Mapping for economic evaluation. Br Med Bull. 2012;101(1).
44. Office for National Statistics (ONS). Labour Force Survey. Available at: https://www.ons.gov.uk/surveys/informationforhouseholdsandindividuals/householdandindividualsurveys/labourforcesurveylfs. [Accessed November 3, 2017].

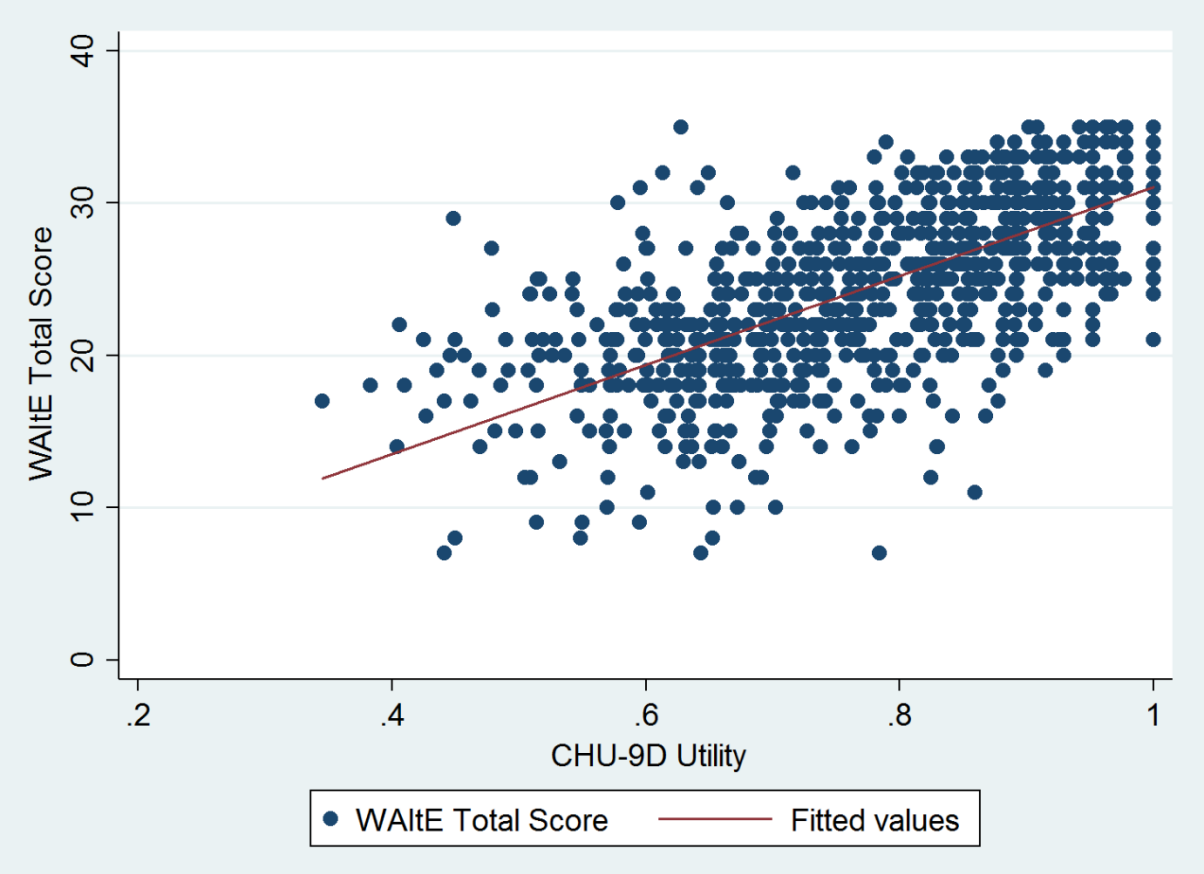


Figure 1- Correlation between the WAItE Total Score and the CHU-9D Utility Score

Table 1- Participant Characteristics (*N* = 975)			
Column	**1**	**2**	**3**
Characteristic	***n* (%)**	**WAItE Total Score**	**CHU-9D Utility Score**
Full Sample (%)	975 (100%)	-	-
Mean (SD)	-	25.39 (5.99)	0.81 (0.14)
Median (IQR)	-	26.00 (21.00-30.00)	0.84 (0.70-0.92)
Gender			
Males	482 (%)	26.25 (5.72)	0.82 (0.14)
Females	493 (%)	24.55 (6.14)	0.79 (0.15)
Age			
11-15	361 (%)	28.60 (5.43)	0.88 (0.12)
16-18	614 (%)	23.50 (5.48)	0.76 (0.14)
Ethnicity			
White British/Irish	741 (76.00%)	25.70 (6.14)	0.82 (0.14)
Asian or Asian British	80 (8.21%)	24.59 (5.23)	0.77 (0.14)
Black Caribbean/African	46 (4.72%)	27.32 (4.87)	0.83 (0.13)
Chinese/East Asian	56 (5.74%)	20.16 (3.46)	0.66 (0.09)
Mixed Ethnicity	43 (4.41%)	26.67 (4.87)	0.82 (0.12)
Other	9 (0.92%)	23.55 (6.00)	0.77 (0.16)
Weight Status			
Normal	353 (36.21%)	27.10 (5.64)	0.84 (0.14)
Overweight	221 (22.67%)	26.10 (5.67)	0.81 (0.14)
Obese	401 (41.13%)	23.49 (5.93)	0.78 (0.14)
Self-Assessed Weight			
Very Overweight	97 (9.95%)	19.66 (4.81)	0.68 (0.11)
Moderately Overweight	141 (14.46%)	22.15 (5.74)	0.75 (0.14)
Slightly Overweight	251 (25.74%)	24.23 (5.08)	0.78 (0.13)
About the right Weight	418 (42.87%)	28.26 (5.16)	0.87 (0.13)
Slightly Underweight	59 (6.05%)	27.33 (5.40)	0.84 (0.14)
Moderately Underweight	7 (0.72%)	25.86 (4.67)	0.74 (0.15)
Very Underweight	2 (0.21%)	22.5 (6.36)	0.69 (0.44)
Self- Assessed Health			
Excellent	168 (17.23%)	29.95 (4.72)	0.92 (0.10)
Very Good	275 (28.21%)	27.58 (5.03)	0.86 (0.11)
Good	313 (32.10%)	23.71 (5.29)	0.77 (0.13)
Fair	157 (16.10%)	21.96 (5.36)	0.72 (0.14)
Poor	62 (6.36%)	20.48 (5.81)	0.68 (0.16)
Illness or Disability			
Yes	560 (57.44%)	23.06 (5.68)	0.75 (0.14)
No	415 (42.56%)	28.54 (4.85)	0.88 (0.11)

* As per Cole et al [34], individuals in the 85th percentile of the weight distribution for their age and gender were classed as overweight, and those in the 95th percentile were classed as being obese.

Table 2- Goodness-of-fit results from the full estimation sample (*N* =975)							
Column	1	2	3	4	5	6	7
Econometric Model	Mean (SD)	Min	Max	MAE	MSE	Abs diff <0.03	Abs diff <0.05
Observed	0.8056 (0.1440)	0.3454	1	-	-	-	-
Model Specification 1- Mapping from WAItE Total Score							
1. OLS	**0.8056 (0.1025)**	0.4953	0.9902	0.0775	**0.0102**	26.5%	42.8%
2. Tobit	0.8125 (0.1113)	0.4769	1.0152	0.0767	0.0103	27.7%	44.8%
3. CLAD	0.8200 (0.1089)	0.4891	1.0129	0.0768	0.0105	**29.8%**	**44.9%**
4. Two-Part	**0.8056 (0.1034)**	0.5086	0.9864	0.0772	**0.0102**	26.5%	43.4%
5. GLM	**0.8056 (0.1025)**	0.5392	1.0057	0.0771	**0.0102**	26.7%	43.1%
6. MM	0.8192 (0.1046)	0.5004	1.0021	**0.0765**	0.0104	29.1%	43.9%
7. BB	0.8136 (0.0816)	0.4962	0.9273	0.0824	0.0112	21.5%	37.3%
8. FMM	0.8077 (0.0985)	0.5102	0.9864	0.0777	**0.0102**	27.2%	41.6%
9. OLOGIT	0.9033 (0.1193)	**0.3779**	**1**	0.1179	0.0244	19.6%	33.6%
Model Specification 2- Mapping from Individual WAItE Items							
1. OLS	**0.8056 (0.1037)**	0.4944	0.9931	0.0760	**0.0100**	28.0%	43.7%
2. Tobit	0.8125 (0.1126)	0.4770	1.0238	0.0757	0.0101	29.4%	45.2%
3. CLAD	0.8160 (0.1092)	0.4908	1.0071	**0.0751**	0.0102	29.1%	47.6%
4. Two-Part	**0.8056 (0.1040)**	0.5067	0.9909	0.0759	**0.0100**	28.0%	44.6%
5. GLM	**0.8056 (0.1039)**	0.5381	1.008	0.0758	**0.0100**	28.3%	44.4%
6. MM	0.8181 (0.1083)	0.4859	0.9933	0.0752	0.0103	**30.4%**	**48.1%**
7. BB	0.8138 (0.0831)	0.4968	0.9273	0.0813	0.0109	21.6%	38.6%
8. FMM	0.8087 (0.0979)	0.5160	0.9925	0.0766	**0.0100**	26.4%	42.7%
9. OLOGIT	0.8957 (0.1193)	**0.4012**	**1**	0.1117	0.0227	20.9%	36.2%

CHU-9D – Child Health Utility 9D; MAE – mean absolute error; MSE – mean squared error. **Numbers in bold** are the best value in each column. For column 1, this is the closest value to the observed mean CHU-9D utility. For columns 2 & 3 this is the closest value to the observed minimum and maximum CHU-9D utility. For columns 4 & 5 this is the lowest MAE/MSE value. For columns 6 & 7 this is the highest number of absolute differences lower than 0.03/0.05.

Table 3- Goodness-of-fit results from the validation analyses (Model Specification 1)

	k-fold validation (*N* =975)					Random sample 1 (*n*=500)					Random sample 2 (*n*=300)					Random sample 3(*n*=100)				
	Observed = 0.8056					Observed = 0.8090					Observed = 0.8146					Observed = 0.8136				
Column	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
	Mean Utility	*MAE*	*MSE*	*% Abs diff< 0.03*	*% Abs diff< 0.05*	*Mean Utility*	*MAE*	*MSE*	*% Abs diff< 0.03*	*% Abs diff< 0.05*	*Mean Utility*	*MAE*	*MSE*	*% Abs diff< 0.03*	*% Abs diff< 0.05*	*Mean Utility*	*MAE*	*MSE*	*% Abs diff< 0.03*	*% Abs diff< 0.05*
OLS	0.8069	0.0776	0.0103	26.5%	42.7%	**0.8101**	0.0805	**0.0108**	23.2%	41.2%	0.8089	0.0767	0.0098	25.0%	42.3%	0.8029	0.0743	0.0095	33.0%	48.0%
Tobit	0.8124	0.0769	0.0104	27.8%	44.6%	0.8173	0.0799	0.0110	24.4%	**43.6%**	0.8161	0.0753	0.0098	26.7%	43.0%	0.8097	0.0744	0.0099	33.0%	49.0%
CLAD	0.8167	0.0769	0.0104	28.3%	44.7%	0.8277	**0.0780**	0.0112	**28.0%**	43.0%	0.8226	0.0749	0.0099	26.3%	42.7%	**0.8154**	0.0724	0.0095	33.0%	**51.0%**
Two-Part	0.8128	**0.0767**	0.0103	28.2%	**45.1%**	0.8102	0.0804	**0.0108**	23.2%	42.0%	0.8091	0.0761	0.0098	25.3%	42.3%	0.8030	0.0744	0.0096	31.0%	48.0%
GLM	**0.8056**	0.0772	**0.0102**	27.1%	42.9%	**0.8101**	0.0803	**0.0108**	23.4%	41.0%	0.8091	0.0759	**0.0097**	25.0%	42.0%	0.8030	0.0751	0.0097	31.0%	47.0%
MM	0.8191	0.0767	0.0104	**28.6%**	44.1%	0.8240	0.0794	0.0111	27.0%	43.0%	0.8224	**0.0743**	0.0098	**30.0%**	**44.7%**	0.8160	**0.0721**	0.0095	**34.0%**	50.0%
BB	0.8137	0.0825	0.0112	21.4%	37.5%	0.8172	0.0842	0.0117	22.2%	38.0%	**0.8157**	0.0823	0.0108	19.3%	26.0%	0.8109	0.0743	**0.0090**	27.0%	44.0%
FMM	0.8076	0.0779	0.0103	26.2%	41.8%	0.8120	0.0805	**0.0108**	24.2%	40.0%	0.8109	0.0770	0.0099	25.0%	42.0%	0.8052	0.0737	0.0094	33.0%	47.0%
OLOGIT	0.9028	0.1182	0.0244	19.5%	33.2%	0.9098	0.1201	0.0251	19.2%	32.2%	0.9053	0.1108	0.0220	21.7%	36.7%	0.8989	0.1124	0.0200	14.0%	32.0%

CHU-9D – Child Health Utility 9D; MAE – mean absolute error; MSE – mean squared error. **Numbers in bold** are the best value in each column. For columns 1, 6, 11 & 16 this is the closest value to the observed mean CHU-9D utility. For columns 2, 7, 12 & 17 this is the lowest MAE value. For columns 3, 8, 13 & 18 this is the lowest MSE value. For columns 4, 9, 14 & 19 this is the highest number of absolute differences lower than 0.03. For columns 5, 10, 15 & 20 this is the highest number of absolute differences lower than 0.05.

Table 4- Goodness-of-fit results from the validation analyses (Model Specification 2)

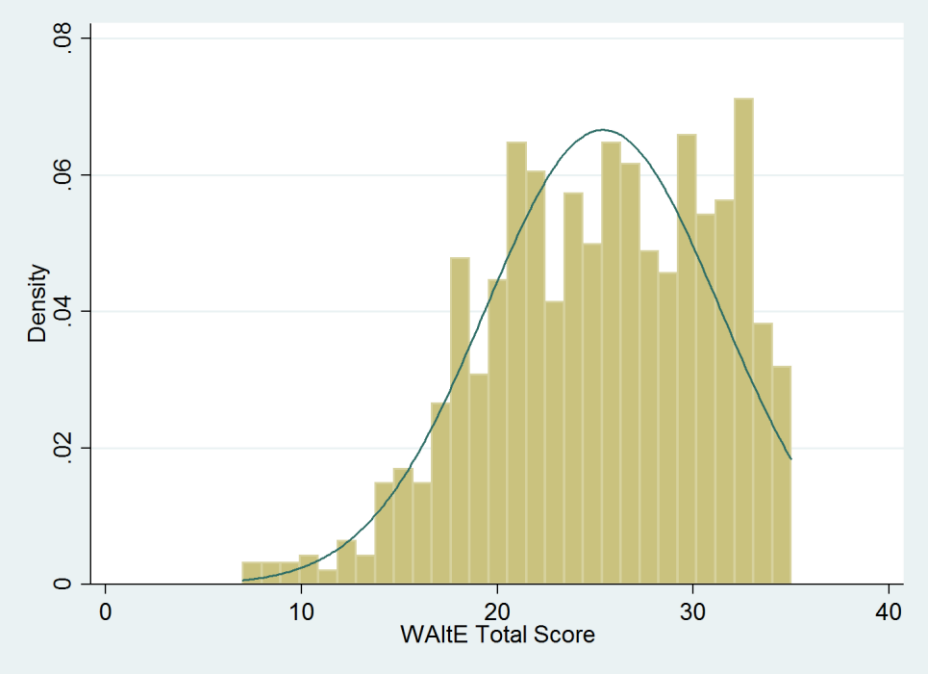
	k-fold validation (*N* = 975)					Random sample 1 (*n*=500)					Random sample 2 (*n*=300)					Random sample 3 (*n*=100)				
	Observed = 0.8056					Observed = 0.8090					Observed = 0.8146					Observed = 0.8136				
	Mean Utility	*MAE*	*MSE*	*Abs diff< 0.03*	*Abs diff< 0.05*	*Mean Utility*	*MAE*	*MSE*	*Abs diff< 0.03*	*Abs diff< 0.05*	*Mean Utility*	*MAE*	*MSE*	*Abs diff< 0.03*	*Abs diff< 0.05*	*Mean Utility*	*MAE*	*MSE*	*Abs diff< 0.03*	*Abs diff< 0.05*
OLS	0.8061	0.0770	0.0102	28.0%	42.8%	**0.8120**	0.0785	**0.0105**	25.8%	41.8%	0.8109	0.0747	**0.0093**	25.0%	41.7%	0.8035	0.0725	0.0090	32.0%	49.0%
Tobit	0.8130	0.0768	0.0104	28.7%	44.7%	0.8195	0.0785	0.0107	27.4%	43.2%	0.8184	0.0737	0.0094	27.0%	45.0%	0.8100	0.0742	0.0096	29.0%	47.0%
CLAD	0.8151	**0.0758**	0.0104	**30.1%**	**47.3%**	0.8247	**0.0777**	0.0109	28.6%	45.8%	0.8250	0.0722	**0.0093**	29.0%	**47.7%**	0.8168	0.0701	0.0089	34.0%	51.0%
Two-Part	**0.8057**	0.0764	**0.0101**	28.5%	44.4%	**0.8120**	0.0783	**0.0105**	25.4%	43.2%	0.8112	0.0743	**0.0093**	25.0%	45.3%	0.8040	0.0723	0.0091	32.0%	49.0%
GLM	0.8060	0.0769	0.0102	27.4%	42.7%	**0.8120**	0.0784	**0.0105**	25.0%	42.6%	0.8110	0.0740	**0.0093**	26.7%	42.7%	0.8038	0.0737	0.0092	31.0%	46.0%
MM	0.8181	0.0760	0.0104	29.3%	46.3%	0.8247	**0.0777**	0.0109	**28.8%**	**47.2%**	0.8232	**0.0720**	**0.0093**	**31.7%**	47.0%	0.8177	**0.0680**	0.0087	**37.0%**	**55.0%**
BB	0.8140	0.0818	0.0110	21.3%	38.2%	0.8187	0.0832	0.0115	21.8%	39.0%	0.8177	0.0805	0.0104	20.0%	37.0%	**0.8118**	0.0723	**0.0085**	26.0%	46.0%
FMM	0.8078	0.0768	0.0101	26.4%	42.5%	0.8148	0.0789	0.0106	24.4%	41.4%	**0.8137**	0.0754	0.0095	23.3%	42.7%	0.8066	0.0723	0.0090	29.0%	47.0%
OLOGIT	0.8953	0.1117	0.0226	21.3%	36.3%	0.8957	0.1144	0.0239	20.6%	35.2%	0.8957	0.1052	0.0212	23.0%	39.0%	0.8957	0.1062	0.0188	16.0%	34.0%

CHU-9D – Child Health Utility 9D; MAE – mean absolute error; MSE – mean squared error. **Numbers in bold** are the best value in each column. For columns 1, 6, 11 & 16 this is the closest value to the observed mean CHU-9D utility. For columns 2, 7, 12 & 17 this is the lowest MAE value. For columns 3, 8, 13 & 18 this is the lowest MSE value. For columns 4, 9, 14 & 19 this is the highest number of absolute differences lower than 0.03. For columns 5, 10, 15 & 20 this is the highest number of absolute differences lower than 0.05.

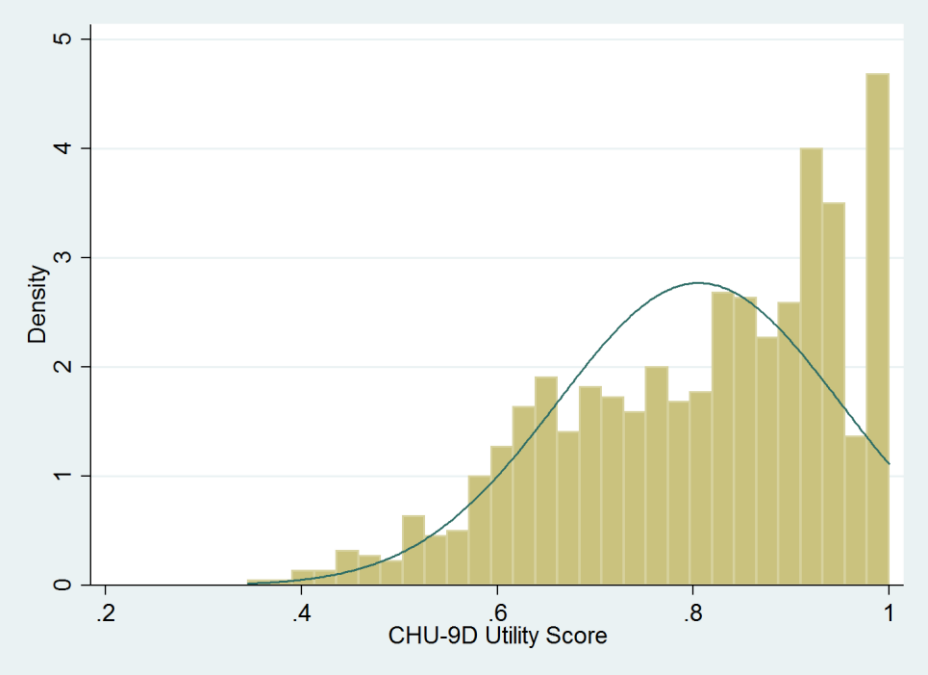
Table 5- Model performance for best-fitting models

Estimation Method	Observed CHU-9D Mean (SD)	Mean	Min	P.25	Median	P.75	Max	MAE	MSE	% Abs diff< 0.03	% Abs diff< 0.05
Two-Part											
Full Sample	0.8056	0.8056	0.5067	0.7419	0.8012	0.8911	0.9909	0.0759	0.0100	28.0%	44.6%
k-fold Validation	0.8056	0.8057	0.5069	0.7401	0.8012	0.8914	0.9901	0.0764	0.0101	28.5%	44.4%
Random Sample 1	0.8090	0.8083	0.5067	0.7315	0.8029	0.8918	0.9907	0.0783	0.0105	25.4%	43.2%
Random Sample 2	0.8146	0.8050	0.5067	0.7400	0.7983	0.8823	0.9877	0.0743	0.0093	25.0%	45.3%
Random Sample 3	0.8136	0.7984	0.5329	0.7425	0.7952	0.8626	0.9794	0.0723	0.0091	32.0%	49.0%
MM											
Full Sample	0.8056	0.8181	0.4859	0.7417	0.8221	0.9096	0.9933	0.0752	0.0103	30.4%	48.1%
k-fold Validation	0.8056	0.8181	0.4861	0.7419	0.8225	0.9099	0.9933	0.0760	0.0104	29.3%	46.3%
Random Sample 1	0.8090	0.8202	0.4858	0.7421	0.8278	0.9152	0.9933	0.0777	0.0109	28.8%	47.2%
Random Sample 2	0.8146	0.8177	0.4858	0.7468	0.8239	0.9043	0.9872	0.0720	0.0093	31.7%	47.0%
Random Sample 3	0.8136	0.8138	0.5090	0.7567	0.8273	0.8901	0.9764	0.0680	0.0087	37.0%	55.0%

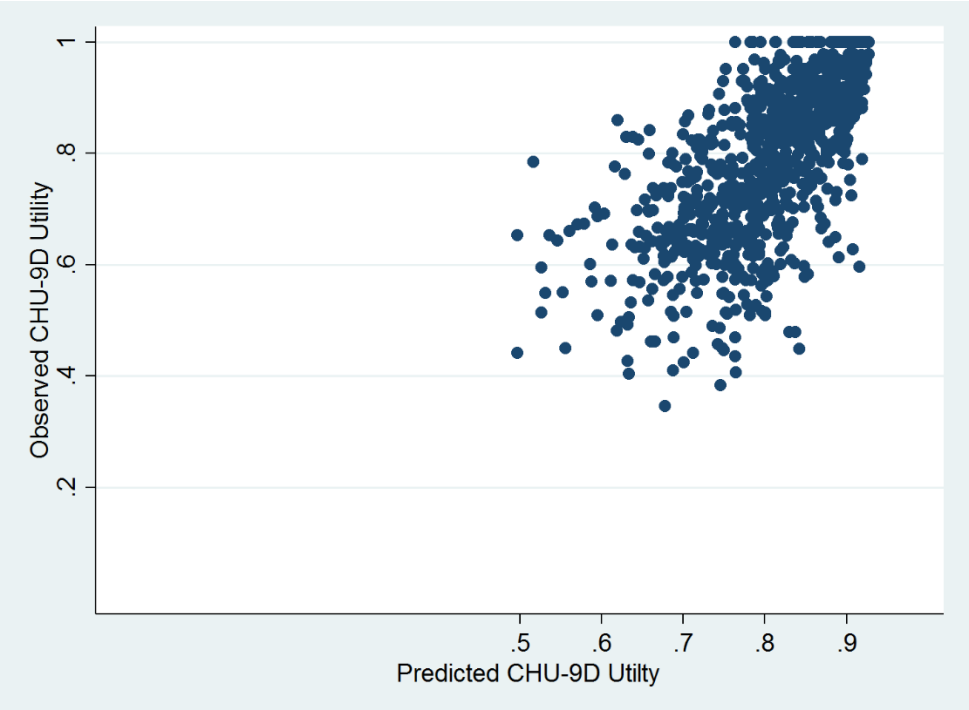
CHU-9D – Child Health Utility 9D; MAE – mean absolute error; MSE – mean squared error.



Online Supplementary Materials- Distribution of WAItE Total Score



Online Supplementary Materials- Distribution of CHU-9D Utility Score



Online Supplementary Materials- Observed vs Predicted CHU-9D Utilities

Online Supplementary Materials- Mapping equations from the WAItE to CHU-9D utility scores[a]			
Independent variables	Two-Part (Model Specification 2)		Robust MM-estimator (Model Specification 2)
	Logit Model [b]	OLS Model	
WAItE Tired	0.9280337*** (0.1451408)	0.0122953*** (0.0044152)	0.0107895** (0.0042061)
WAItE Walking	0.437054* (0.2437725)	0.0102579** (0.0045578)	0.015773** (0.0066986)
WAItE Concentration	0.4776855*** (0.1486862)	0.0262418*** (0.0036084)	0.0231428*** (0.0036076)
WAItE Embarrassed	0.1882361 (0.2081824)	0.0144265*** (0.0038541)	0.0131602*** (0.0046261)
WAItE Unhappy	0.3434785 (0.2650085)	0.0214737*** (0.0047145)	0.0288246*** (0.0061739)
WAItE Treated Different	0.082201 (0.2457673)	0.0180189*** (0.0045279)	0.0244563*** (0.0057471)
Age	-0.181171*** (0.0553389)	-0.0065962*** (0.0015402)	-0.006123*** (0.0014019)
Constant	-9.200232*** (1.601675)	0.5226641*** (0.0335752)	0.4798968*** (0.0359304)

Standard errors in parentheses. $p < 0.01$=***, $p < 0.05$=**, $p < 0.1$=*

[a] WAItE dimension scores reverse coded so that 1 indicates 'Always' and 5 indicates 'Never'.

[b] Dependant variable in Logit Model is a dummy variable with the value of 1 if CHU-9D=1 and 0 otherwise. OLS model only conducted on those who reported a CHU-9D utility below 1.

Online Supplementary Materials- Distribution of errors according to selected ranges in the CHU-9D utility score									
	n	Two–Part (Model Specification 1)		GLM (Model Specification 1)		CLAD (Model Specification 2)		GLM (Model Specification 2)	
		MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE
CHU-9D < 0.7	250	0.1158	0.0201	0.1149	0.0199	0.1230	0.0227	0.1125	0.0194
0.7 < CHU9D < 0.8	179	0.0582	0.0056	0.0568	0.0053	0.0675	0.0073	0.0548	0.0052
0.8 < CHU-9D < 0.9	247	0.0670	0.0070	0.0673	0.0069	0.0626	0.0065	0.0670	0.0068
0.9 < CHU-9D < 1	324	0.0662	0.0078	0.0671	0.0080	0.0549	0.0060	0.0664	0.0079
CHU-9D – Child Health Utility 9D; MAE – mean absolute error; MSE – mean squared error.									

Online Supplementary Materials- Variance Covariance Matrix from Mapping Algorithm

	WAItE_Tired	WAItE_Walking	WAItE_Concentration	WAItE_Embarassed	WAItE_Unhappy	WAItE_Treated_Different	Age	Constant
WAItE_Tired	.00001769							
WAItE_Walking	-.00001179	.00004487						
WAItE_Concentration	-3.675e-06	-2.145e-06	.00001301					
WAItE_Embarassed	-6.904e-07	-.00001378	9.133e-07	.0000214				
WAItE_Unhappy	-4.201e-06	.00001295	-2.231e-06	-.00001484	.00003812			
WAItE_Treated_Different	2.579e-06	-.00001727	-7.348e-08	6.717e-06	-.00002437	.00003303		
Age	5.949e-07	-6.250e-07	1.340e-06	8.062e-07	4.178e-07	6.238e-07	1.965e-06	
Constant	.0000149	-.00006815	-.00004512	-8.914e-06	-.00003883	-7.818e-06	-.00004078	.001291