Measuring utility values in children aged 6-7 years: CHU-9D vs. EQ-5D-Y

Canaway A. and Frew E.

Health Economics Unit, University of Birmingham

Abstract

Purpose

To examine the performance of the Child Health Utility 9D (CHU-9D) and EuroQol 5D-Youth (EQ-5D-Y) in children aged 6-7 years old.

Method

The CHU-9D and EQ-5D-Y were interviewer administered to 160 children aged 6-7 years at six schools across the West Midlands. Missing values, time taken to complete instruments and interviewer ratings were recorded to assess feasibility/acceptability. Construct validity was assessed by testing convergent validity hypotheses. Reliability was examined via a test-retest of a sub-sample. Psychometric properties were further examined by exploring distributions of utility scores, qualitative notes, and design of the questionnaires.

Results

No missing responses were recorded with over 80% of children's understanding being rated as good/excellent for both questionnaires. The average completion time for both instruments was less than 3 minutes demonstrating excellent feasibility/acceptability. Evidence of construct validity was recorded with 12 of the 13 hypotheses being supported. Test-retest reliability was relatively poor for both instruments with weighted kappa coefficients ranging from fair to moderate.

Conclusion

Children aged 6-7 years can feasibly complete utility instruments when interviewer-administered. The reliability of the instruments is of concern and requires further study. In respect to content validity and other psychometric properties, the CHU-9D is favoured to the EQ-5D-Y.

Introduction

The past decade has seen an overwhelming shift towards the use of generic preference-based measures (GPBM) in economic evaluations. In England and Wales, the National Institute of Health and Clinical Excellence's (NICE) guide to the methods of technology appraisal recommends the use of preference based health related quality of life (HRQL) measures [1, 2] and is symbolic of the shift towards the use of preference-based measures in economic evaluations.

Adult GPBM have been extensively developed and used in the health economics literature to assess the cost-effectiveness of interventions across disease groups [3]. The past decade has seen increased interest in the development of generic HRQL measures for use in children for example the Paediatric Quality of Life Initiative (PedsQL) [4]; the majority of these however tend to be non-preference based measures. The inclusion of children's HRQL is becoming increasingly important with children playing a larger part in health research [5]. There are clear differences between children and adults, and it would therefore be wrong to assume the dimensions of adult GPBM's are relevant to children, whose physical functioning and psycho-social skills are still developing [6]. Adult measures are therefore inappropriate for use in young children.

Proxy respondents have frequently been used to measure HRQL on behalf of young children e.g. McAuley *et al.* [7]. Evidence has repeatedly found parents to be poor proxies [8-11]. It has also been shown that parent's views of their child's health are affected by their own health status [8], children sometimes conceal their true feeling from potential proxies [9] and parents introduce bias into the analysis [10]. This is demonstrated by Achenbach [11] who investigated using a meta-analysis the relationship between child and parent responses. The average correlation was just 0.22. Despite their convenience, proxy measures are clearly flawed. A growing body of research [5, 12] has found increasingly strong arguments for eliciting data from children themselves. Riley [5] found that children aged as young as 6 years old can independently self-report on their own health status giving a perspective that proxies cannot fulfil. It is therefore important that if feasible, children's views are incorporated into health economic evaluations.

In 2010, two new GPBM were developed that are designed specifically for children, these are the EQ-5D-Y [13] and the CHU-9D [14]. The EQ-5D-Y has been tested for feasibility, validity and reliability in children aged 8+ years with promising results [15], whilst the CHU-9D has been tested for 7-11 year olds [14]. Given the issues surrounding proxy responses and the increasing evidence that children aged 6 years old can self report on their HRQL, it is important to assess the capabilities of these instruments in children aged 6-7 years. This study will assess the performance of the EQ-5D-Y and the CHU-9D alongside the PedsQL instrument which is a validated non-preference based paediatric HRQL measure.

Methods

Study population

Through contacts within the local Primary Care Trusts, six mainstream schools from the West Midlands were recruited into the study. Informed consent from parents and assent from children was a prerequisite for children to participate in the study.

Data Collection

Five researchers were trained to administer the questionnaires and to complete the interviewer forms. The room was organised with desks positioned in each corner to form 'work stations' where the question pack could be administered. To ensure that ordering bias [16] was minimised, the ordering of the CHU-9D, the EQ-5D-Y and the PedsQL in the question pack were randomised and study ID's were implemented to ensure anonymity. The question packs were interviewer administered with the questions being read verbatim to the child. The interviewers were given an "interviewer's form" to complete during/after interviewing each child. The form recorded the time taken to complete each instrument, questions struggled on and a section to rate on a scale of 1-5 (1 being 'very poor', 5 being 'excellent') how they felt the child understood each instrument. The form allowed for any qualitative notes that the interviewers found pertinent to the study. A test re-test of the questionnaires was carried out on a sub-sample (n=24) to examine reliability of the questionnaires. This was done by administering the questionnaires in the morning and then again in the afternoon.

Measures

EQ-5D-Y

The EQ-5D-Y is a GPBM adapted directly from the EQ-5D to calculate utility values for those aged 8 years and older (see [13] for full details). The rationale of the EQ-5D-Y is to have a measure that corresponds to an adult utility measure enabling the calculation of utility values using the same framework from childhood to adulthood [15]. The EQ-5D-Y features a 5 dimension descriptive system with the wording of the questions in each dimension modified to make it pertinent to a younger age range. For example, mobility is referred to as 'walking about' [15]. As with the adult version there are three severity levels for each dimension (no problems, some problems, lots of problems). The adult EQ-5D utility algorithm [17] can then be directly applied to generate utility values between -0.59 and 1 (perfect health) for the 243 possible health states.

CHU-9D

The CHU-9D was developed in response to the dearth of paediatric GPBM's [14]. It has been developed for use in 7-11 year old children and was created through extensive thematic content analysis of qualitative interviews with children to identify dimensions that are important to children [18]. The CHU-9D features some dimensions common to other measures e.g. 'pain' and 'able to join in usual activities' along with dimensions specific to the CHU-9D e.g. 'school work/homework', 'tired' and 'sleep'. Each dimension has 5 severity levels (e.g. I don't feel sad today, I feel a little bit sad today, I feel a bit sad today, I feel quite sad today, I feel very sad today). An algorithm has been formulated allowing the calculation of utility values for 1,953,125 unique health states with scores ranging from 0.33 to 1 (perfect health).

PedsQL 4.0 – Young Child Version

The PedsQL is a validated non-preference based HRQL measure [4] and features 23 items containing four core domains: physical functioning (8 items), emotional functioning (5 items), social functioning (5 items) and school functioning (5 items). The PedsQL addresses issues that are frequently problems for children, e.g. 'is it hard for you to walk'. The young child version (ages 5-7 years) of the PedsQL 4.0 has three possible response options (not at all, sometimes, a lot). Responses to the PedsQL were reverse linearly transformed (0 = 100, 2 = 50, 4 = 0), so that higher scores equate to better quality of life [4]. Total PedsQL score was calculated by summing the score for each item and then dividing by the number of items answered. Physical functioning summary scores and psycho-social summary scores were also calculated using the same method but restricted just to either the physical functioning domain or the emotional, social and school domains. This validated measure was included to help assess validity of the 2 instruments of interest in this paper.

Analysis

The acceptability of the two economic instruments was examined by analysing the percentage of missing values [19] whilst feasibility was assessed by examining the time taken to complete the instruments [20]. Acceptability/feasibility was further assessed by examining the interviewer ratings for each instrument. Further psychometric properties were determined by analysing the distribution of responses for each measure and floor/ceiling effects [21].

Reliability was assessed by examining the percentage agreement and kappa coefficients (weighted and un-weighted) to assess agreement between the test and re-test responses [22]. In line with Ravens-Sieberer et al [15], Landis and Koch's guidelines [23] were used to assess degree of agreement using kappa coefficients. A kappa coefficient < 0.2 indicates poor agreement, 0.21 to 0.40 indicates fair agreement, 0.41 to 0.6 indicates moderate agreement, 0.61 to 0.80 indicates substantial agreement and > 0.81 indicates almost perfect agreement.

Construct validity was assessed using convergent validity which is tested by examining whether dimensions that theoretically should be related from different instruments actually are related. This would usually be assessed by correlating dimensions from the instrument to theoretically similar dimensions in a widely validated instrument. As there is no widely validated preference-based measure for use in children aged 6-7 years to use as a comparator, core-scales from the PedsQL and theoretically similar dimensions from the other GPBM were used to formulate hypotheses to test. Coefficient guidelines from Cohen [24] were used to assess the magnitude of correlation between the dimensions. Correlations less than 0.29 are classified as low, 0.3 to 0.49 are deemed moderate, and those above 0.5 are classified as high. In total, 13 hypotheses testing theoretically similar dimensions/domains were formulated to test convergent validity. For example, the PedsQL psycho-social summary score contains an item relating to sadness i.e. 'do you feel sad', the CHU-9D contains a dimension addressing

sadness, given this, it was hypothesised that the 'sad' dimension of the CHU-9D should be negatively correlated with the PedsQL psycho-social summary score. Qualitative notes were taken to inform any other arising issues.

Results

The Sample

A total of 160 children were recruited into the sample from 6 different schools across the West Midlands. A test re-test was conducted on a subgroup of 24 pupils. All children in the study were aged 6-7 years old (101 children aged six years, 59 children aged 7 years).

Descriptive Statistics of Utility Instruments

Table 1 shows the descriptive statistics for the two utility-based instruments. The mean utility score for the CHU-9D was higher than the EQ-5D-Y (0.86 vs. 0.72). Given these two instruments are supposed to be measuring the same thing i.e. health related quality of life on a scale of 0 to 1, it is an important difference. Given the non-normal distributions of data within this study non-parametric tests were preferred to parametric tests. A Wilcoxon signed ranks test was used to test whether this difference was statistically significant or due to chance. It was found that this difference was statistically significant ($p < 0.05$). The EQ-5D-Y has a substantially larger range of utility values which is unsurprising given the larger range of possible values for the EQ-5D-Y. Given that the sample comprises seemingly healthy school children it is surprising to see such a low utility value observed using the EQ-5D-Y (-0.594) i.e. the worst possible health state.

Feasibility and Acceptability

Table 2 shows the time taken to complete the two instruments. The results show that both utility-based instruments on average were completed in less than 3 minutes and were therefore very little burden to the interviewer/child. There was no (0%) missing responses for both instruments. Table 3 shows the children's understanding of the questionnaires as rated by the interviewers. The interviewers rated children's understanding as good or excellent for both instruments in over 80% of the children. Only 7.1% of children interviewed were rated as having a poor or very poor understanding. Both instruments are clearly feasible when interviewer administered to children aged 6-7 years old.

Reliability

Table 4 shows the results for the test re-test reliability element of the study.. For the EQ-5D-Y percentage agreement ranges from 70.8% to 89.6% whilst weighted kappa coefficients ranged from fair to moderate. The range of percentage agreement for the CHU-9D was smaller than that of the EQ-5D-Y with 72.9% ('tired' dimension) being the lowest and 86.5% ('work' and 'routine' dimensions) being the greatest. The agreement as assessed by the weighted kappa coefficients for the CHU-9D dimensions was found to be fair to moderate, on average it was slightly higher for the CHU-9D dimensions than for the EQ-5D-Y dimensions. The weighted kappa coefficients for both measures indicate relatively fair to moderate reliability.

Psychometric Properties of Utility Instruments

Table 5 shows the percentage of respondents reporting each possible answer for all EQ-5D-Y dimensions and CHU-9D dimensions. Ceiling effects are prevalent in the dimensions of both instruments. Four of the EQ-5D-Y dimensions have over 70% of respondents reporting the top level. 'Pain' is the exception with 60% of respondents reporting the highest level. The CHU-9D has greater variation in ceiling effects for each dimension. Given that the children are healthy enough to attend mainstream school it is unsurprising to see ceiling effects within these dimensions.

Figure 1 shows the distribution of utility values across the sample using both utility-based instruments to examine ceiling effects of the instruments and the distribution of responses. The ceiling effect associated with the EQ-5D-Y is considerably larger than that of the CHU-9D with over 30% reporting full health. There is very little evidence of a floor effect taking place which is consistent with what would be expected in the general population. In terms of the distribution of utility values as a whole for each instrument; the distribution of the EQ-5D-Y values is worrying. As can be seen in figure 1 and is frequently seen in the adult version of the EQ-5D [24], there is clustering of indices resulting in the response distribution being bi-modal. The CHU-9D distribution on the other hand is a shape we would expect for the general population i.e. many healthy children with decreasing numbers of children reporting worse health states.

Validity

Table 6 shows the correlation coefficients investigating the 13 hypotheses created to investigate construct validity via convergent validity. Twelve of the hypotheses are statistically significant at the 5% level. The hypothesis that was not statistically significant hypothesised that the EQ-5D-Y mobility dimension would be positively correlated with the PedsQL physical functioning summary score, the lack of significance is surprising. Eight of the correlations using Cohen's [24] guidelines are deemed to be moderate whilst 5 of these are relatively low. There is some evidence of convergent validity for both instruments, with neither instrument standing out as having superior construct validity.

Discussion

Feasibility, Validity and Reliability

Given the increased interest in the use of GPBM's in children and the ability of 6 year olds to self report their health; it is important to investigate the relative merits of available instruments to inform instrument choice. The results show that both instruments are clearly feasible and acceptable when interviewer administered. Both of the utility instruments were completed in less than 3 minutes, no missing values were recorded and interviewer ratings of understanding were generally very good for both instruments.

In terms of validity, fair criterion and construct validity was found in both instruments with pre-analysis hypotheses generally being proved correct. Test re-test reliability of both instruments was found to be relatively poor with moderate agreement between test and re-test scores. Given that the test re-test was conducted over a short time frame this is somewhat surprising as it would be expected that health would not have changed in such a short period of time. This raises concerns as to whether children aged 6 to 7 years can reliably answer these questionnaires. The low reliability combined with high response rate raises the question whether the children's responses are meaningful or whether they just see it as a 'box ticking' exercise. Some of the change in test-retest responses particularly in the CHU-9D however may be due to the transient nature of some of the dimensions e.g. the low test-retest reliability of the 'tired' dimension.

The EQ-5D-Y has been tested by Ravens-Sieberer *et al*. [15] in a multinational study for feasibility, validity and reliability in children aged to 8-11 years. The CHU-9D has to the authors knowledge not had such an extensive analysis. Ravens-Sieberer *et al*. [15] found evidence of good feasibility, fair to moderate test-retest reliability and construct validity. Despite the younger age range within this study, the findings from this study are broadly similar to those of Ravens-Sieberer *et al.* with good feasibility being found even at age 6-7 years old when interviewer administered.

Other Psychometric Properties

Dimensional ceiling effects were observed in both utility instruments as would be expected in a healthy population which is consistent with Ravens-Sieberer *et al's* (2010) findings for the EQ-5D-Y. The ceiling effects associated with the utility index scores and the distributions of utility values for each instrument raise some concern. Given that the sample comprises pupils healthy enough to attend mainstream schools; we expected to see a negative skew within the distribution of utility values with more pupils classified as being in good health, and fewer pupils within the worsening health states giving us a smooth upwards sloping curve. As can be seen in figure 1, this is the case with the CHU-9D. The EQ-5D-Y on the other hand is bimodal and features a significant ceiling effect with over 30% being in 'perfect health'. The implication of this is that the EQ-5D-Y is less capable at distinguishing between health states approaching full health. The CHU-9D has a more expansive descriptive system with 5 severity levels and as a result is capable of detecting smaller changes in HRQL. The Euroqol

group are however reported to be in the process of developing a 5 level version of the EQ-5D-Y [15] which may lead to improved performance of this instrument in the future.

In terms of the distribution of utility values using the EQ-5D-Y, the distribution is less stable than that of the CHU-9D. This may be due to the inclusion of the 'N3 term' associated with the EQ-5D-Y. When an individual has one dimension of the EQ-5D-Y in the worst health state the N3 term is deducted from their utility score. The consequence as seen in figure 1 is a bi-modal distribution of responses rather than a smooth upward sloping curve as would be expected. A further concern with the EQ-5D-Y relates to the number of children reporting to be in health states worse than death (n=7) given the children are healthy enough to attend school. This raises doubts regarding the appropriateness of the adult EQ-5D tariffs for use in children. Normative arguments suggest that societal preferences should be used [15] however whether the public when valuing the health state should be informed that it is children who will be experiencing the health state is less clear. The severity of some of the health states as derived by the EQ-5D-Y using the adult tariff in this paper provides a basis for arguing for the inclusion of this information.

Severity and Relativity

Given the children's age in this study, there is some concern about their concept of relativity in terms of severity. This was picked up by a qualitative note in the study whereby a child reported that he was 'in a lot of pain' and elaborated by saying it was because they 'fell over at break time'. By reporting as being in 'a lot' of pain for the utility instruments, there is a significant decrease in their utility score. This is especially the case for the EQ-5D-Y where pain is heavily weighted and the N3 term is also triggered. As a result, a minor short term event that triggers the N3 term can have a substantial effect on the child's utility score as measured by the EQ-5D-Y. The N3 term has also been found to be problematic in adults and the root cause of bimodal distributions of utility values [25]. It is hypothesised that given children's sensitivity to minor events, this adverse effect may be amplified in children.

Utility Values

Both the CHU-9D and the EQ-5D-Y are GPBM's and as such are attempting to measure the same thing i.e. utility. It is therefore surprising to find that the mean utility values provided by the two instruments are so different. The mean (median) utility value associated with the EQ-5D-Y is 0.72 (0.80) compared to the CHU-9D's mean (median) of 0.86 (0.90). This difference is statistically significant ($P < 0.05$) and demonstrates that the instruments are not measuring 'utility' consistently. This is likely due to the differing descriptive systems used in each measure and the different valuation processes used. By having 9 dimensions the CHU-9D captures a host of dimensions that the EQ-5D-Y discounts. The different methods of eliciting utility values used to create the utility algorithms will also lead to differences between the utility instruments. Green *et al.* [26] found that health state valuations using time-trade off (TTO) and standard gamble (SG) lead to different utility scores. Given the EQ-5D-Y utility algorithm is based on TTO, whilst the CHU-9D utilises the SG, such differences are unsurprising. In adult measures differences of 0.05 are common between mean utility

values derived from different measures e.g. the SF-6D and EQ-5D [25]. The magnitude of difference between these two instruments however is over double this and represents an important difference that could influence the decision in an economic evaluation.

Time Horizon

The time frame of the questionnaires is an important aspect to utility instruments. As discussed by Petrou [6], children have a significantly diminished recall over time; therefore the specified time period is an important feature. Both questionnaires negate this by making the instruments specific to 'today'. There is however a difference in how this is applied within the instruments. The EQ-5D-Y only mentions the time frame once at the start of the questionnaire conversely the CHU-9D reinforces the time frame by referring to 'today' at the end of every response option. This helps to remind the child that we are interested in their health today specifically. The EQ-5D-Y on the other hand runs the risk of regressing to a longer/inappropriate time horizon.

Thematic Content Analysis vs. Transitional Instrument

The two measures were created via very different approaches. The EQ-5D-Y was developed in response to the dearth of transitional instruments measuring HRQL from childhood into adulthood [13]. The EQ-5D-Y therefore adapted the adult EQ-5D in order to create a transitional instrument with wording suitable to the younger age groups. The CHU-9D however was formulated using qualitative research with children to create a utility based instrument with dimensions pertinent and relevant to children [15]. The different methods both have their relative advantages. The EQ-5D-Y has the advantage of being applicable through childhood and into adulthood; the CHU-9D however through its unique descriptive system has superior content validity by addressing issues that are important to children.

Capability and Functioning

Given the age of the children, there is a worry that some of the utility scores were affected by a divergence between capabilities and functioning's. This was detected by two qualitative notes by the interviewers. The first comment came with regards to the dimension of the EQ-5D-Y that addresses 'looking after myself'; the options refer to the degree of problem children have 'washing or dressing myself'. One child said they had 'some problems' as he "couldn't tie his own tie and his parents had to help". In a similar case another child said their parents helped dress them as it was 'quicker' that way. These children may well be perfectly healthy enough to dress themselves however out of choice they choose not to do so. Consequently there is a disproportionate decrease in utility as measured by the CHU-9D and EQ-5D-Y which both address functioning's and not capabilities. An exploration of the

capabilities approach as developed by Al-Janabi *et al.* [27] in children could be an avenue for future studies using young children to address this problem.

Conclusion

The applicability of the EQ-5D-Y and CHU-9D for use in children as young as 6-7 years was explored in this paper. Both instruments are very similar in terms of feasibility, reliability and construct validity with reliability being a particular concern warranting further investigation. In respect to content validity, the process by which the CHU-9D was formulated is superior to the EQ-5D-Y with thematic content analysis of qualitative interviews with children identifying areas of importance to children. In addition to this, the reinforcement of the time frame throughout the CHU-9D is a further benefit. The EQ-5D-Y however has the benefit of being a transitional instrument feeding into the adult EQ-5D measure. In terms of psychometric properties the CHU-9D is superior for use in school children aged 6-7 years.

These findings suggest that utility scores using the CHU-9D and EQ-5D-Y can feasibly be elicited from children aged 6-7 years when interviewer administered; they however should be used with caution. Both measures need further testing to examine reliability and validity in children within this age group. The psychometric properties need to be further explored in clinical settings to explore the capabilities of these instruments in a wide range of settings.

References

1. NICE 2004, *Guide to the methods of technology appraisal*. Available Online: http://www.nice.org.uk/niceMedia/pdf/TAP_Methods.pdf

2. NICE 2008, *Guide to the methods of technology appraisal*. Available Online: http://www.nice.org.uk/media/B52/A7/TAMethodsGuideUpdatedJune2008.pdf

3. Räsä, N., P., Roine, E., Sinotenen, H., Semberg-Konttinen, V., Ryynänen, P., & Roine, R. 2006, "Use of quality-adjusted life years for the estimation of effectiveness of health care: A systematic literature review.", *International Journal of Technology Assessment in Health Care*, vol. 22, pp. 235-241.

4. Varni, J., Burwinkle, T., & Seid, M. 2006, "The PedsQL^{TM} 4.0 as a school population health measure: Feasibility, reliability, and validity", *Quality of Life Research*, vol. 15, pp. 203-215.

5. Riley, A. 2004, "Evidence That School-Age Children Can Self-Report on Their Health", *Ambulatory Pediatrics*, vol. 4, no. 4, pp. 371-376.

6. Petrou, S. 2003, "Methodological issues raised by preference-based approaches to measuring the health status of children", *Health Economics*, vol. 12, pp. 697-702.

7. McAuley, K., Taylor, R., Farmer, V., Hansen, P., Williams, S., Booker, C., & Mann, J. 2010, "Economic evaluation of a community-based obesity prevention program in children: The APPLE project", *Obesity* no. 1, pp. 131-136.

8. Pal, D. 1996, "Quality of life assessment in children: a review of conceptual and methodological issues in multidimensional health status measures", *Journal of Epidemiology and Community Health*, vol. 50, pp. 391-396.

9. Binger, C., Ablin, A., Feverstein, R., Kushner, J., & Zoger, S. M. C. 1969, "Childhood leukaemia: emotional impact on patient and family", *New England Journal of Medicine*, vol. 280, pp. 414-418.

10. Ball A, Russell, E., Seymour, D., & et al 2001, "Problems in using health survey questionnaires in older patients with physical disabilities", *Gerontology*, vol. 47, pp. 334-340.

11. Achenbach, T., McConaughy, S., & Howell, C. 1987, "Child/Adolescent behavioural and emotional problems: implications of cross-informant correlations for situational specificity", *Psychological Bulletin*, vol. 101, pp. 213-232.

12. Eieser, C. & Morse, R. 2001, "Quality-of-life measures in chronic diseases of childhood", *Health Technology Assessment*, vol. 5, no. 4, pp. 1-156.

13. Wille, N. & Badia, X. e. a. 2010, "Development of the EQ-5D-Y: a child-friendly version of the EQ-5D", *Quality of Life Research*, vol. 19, pp. 875-886.

14. Stevens, K. Assessing the performance of a new generic measure of health related quality of life for children and refining it for use in health state valuation. Discussion Paper . 2010.
Ref Type: Unpublished Work

15. Ravnes-Sieberer, U., Willie, N., & et al 2011, "Feasibility, reliability, and validity of the EQ-5D-Y: results from a multinational study", *Quality of Life Research*, vol. 19, pp. 887-897.

16. Perreault, W. 1975, "Controlling order-effect bias", *Public Opinion Quarterly* pp. 544-551.

17. Szende, A., Oppe, M., & Devlin, N. 2007, *EQ-5D valuation sets: An inventory, comparative review and users' guide.* Rotterdam: EuroQol Foundation, Springer

18. Stevens, K. 2010, "Working with children to develop dimensions for a preference based generic paediatric health related quality of life measures", *Qualitative Health Research*, vol. 20, no. 3, pp. 340-351.

19. Essink-Bot, M., Krabbe, P., & Bonsel, G. A. N. 1997, "An empirical comparison of four generic health status measures: The Nottingham Health Profile, the Medical Outcomes Study 36-Item Short-Form health Survey, the COOP/WONCA Charts, and The EuroQol Instrument", *Medical Care*, vol. 35, pp. 522-537.

20. Brazier, J. & Deverill, M. 1999, "A Checklist for Judging Preference-Based Measures of Health Related Quality of Life: Learning From Psychometric Measures", *Health Economics*, vol. 8, pp. 41-51.

21. McHorney, C., Ware, J., Lu, J., & Sherbourne, C. 1994, "The MOS 36-item short-form health survey (SF-36): III Tests of data quality, scaling assumptions, and reliability across diverse patient groups", *Medical Care*, vol. 32, pp. 40-66.

22. Cohen, J. 1968, "Weighted kappa: Nominal scale agreement with provision for scaled disagreement or partial credit" *Psychological Bulletin*, vol. 70, pp. 213-220

23. Landis, J and Koch, G. 1977, "The measurement of observer agreement for categorical data", *Biometrics*, vol. 33, pp. 159-174

24. Cohen, J. 1988, *Statistical power analysis for the behavioural sciences*, 2nd edn, NJ: Lawrence Erlbaum Publishers.

25. Brazier, J., Roberts, J., Tsuchiya, A., & Busschbach, J. 2011, "A comparison of the EQ-5D and SF6D across seven patient groups", *Health Economics*, vol. 13, pp. 873-884.

26. Green, C., Brazier, J., & Deverill, M. 2000, "Valuing health-related quality of life. A review of health state valuation techniques", *Pharmacoeconomics*, vol. 17, no. 2, pp. 151-165.

27. Al-Janabi, H., Flynn, T., & Coast, J. (In Press), "Development of a self-report measure of capability wellbeing for adults: The ICECAP-A", *Quality of Life Research*

Appendix

Table 1 Descriptive Statistics of the EQ-5D-Y and CHU-9D

	Mean	Maximum	Minimum	Range	Percentile 25	Median	Percentile 75
CHU_9D Utility	.86	1.00	.33	.67	.81	.90	.96
EQ-5D-Y Overall Utility (Tariff)	.72	1.00	-.59	1.59	.62	.80	1.00

Table 2 Time taken to complete the EQ-5D-Y and CHU-9D

	N	Minimum	Maximum	Mean	Std. Deviation
Time Taken to complete CHU9D (seconds)	148	60.00	420.00	147.77	50.96
Time Taken to complete EQ-5D-Y (seconds)	144	60.00	240.00	109.79	31.96

Table 3 Interviewer ratings of child understanding for the EQ-5D-Y and CHU-9D

	Interviewer Rating	N	%
CHU-9D	Very Poor	2	1.3%
	Poor	9	5.8%
	Fair	16	10.4%
	Good	57	37.0%
	Excellent	70	45.5%
EQ-5D-Y	Very Poor	3	1.9%
	Poor	8	5.2%
	Fair	17	11.0%
	Good	62	40.3%
	Excellent	64	41.6%

Table 4 Test re-test correlations and percentage agreement

** Significant at P < 0.05 * Significant at P < 0.1

Measure and Dimension (n=24)	Percentage Agreement (Weighted Kappa)	Test Re-Test Correlation (Spearman's Rho)	Kappa Coefficient	Weighted Kappa Coefficient
EQ-5D-Y				
EQ-5D-Y Mobility	87.5**	0.47**	0.39**	0.53**
EQ-5D-Y Looking After Myself	70.8	0.24	0.17	0.20
EQ-5D-Y Usual Activities	89.6**	-0.61**	0.48**	0.58**
EQ-5D-Y Pain	77.1**	0.35	0.25*	0.27**
EQ-5D-Y Worried/Sad	72.9**	0.35	0.30**	0.29**
CHU-9D				
CHU-9D Worried	76.0**	0.49*	0.19**	0.31**
CHU-9D Sad	79.2**	0.55**	0.18*	0.32**
CHU-9D Pain	84.4**	0.49*	0.44**	0.46**
CHU-9D Tired	72.9**	0.52**	0.12	0.34**
CHU-9D Annoyed	81.3**	0.34	0.28**	0.34**
CHU-9D School Work	86.5**	0.63**	0.53**	0.52**
CHU-9D Sleep	84.4**	0.70**	0.45**	0.60**
CHU-9D Routine	86.5**	0.57**	0.31**	0.41**
CHU-9D Activities	83.3**	0.58**	0.40**	0.41**

Table 5 Percentage of reported problems by severity for the EQ-5D-Y and CHU-9D

		N	%
EQ-5D-Y			
EQ-5D-Y Mobility Dimension	No problems walking around	139	86.9%
	Some problems walking around	14	8.8%
	A lot of problems walking around	7	4.4%
EQ-5D-Y Looking after myself	No problems washing or dressing myself	121	75.6%
	Some problems washing or dressing myself	27	16.9%
	A lot of problems washing or dressing myself	12	7.5%
EQ-5D-Y Usual Activities	I have no problems doing my usual activities	126	78.8%
	I have some problems doing my usual activities	23	14.4%
	I have a lot of problems doing my usual activities	11	6.9%
EQ-5D-Y Pain	I have no pain or discomfort	96	60.0%
	I have some pain or discomfort	50	31.3%
	I have a lot of pain or discomfort	14	8.8%
EQ-5D-Y Worried or Sad	I am not worried, sad or unhappy	113	70.6%
	I am a bit worried, sad or unhappy	34	21.3%
	I am very worried, sad or unhappy	13	8.1%
CHU-9D			
CHU-9D Worried	I don't feel worried today	109	68.1%
	I feel a little bit worried today	15	9.4%
	I feel a bit worried today	11	6.9%
	I feel quite worried today	14	8.8%
	I feel very worried today	11	6.9%
CHU-9D Sad	I don't feel sad today	123	76.9%
	I feel a little bit sad today	13	8.1%
	I feel a bit sad today	9	5.6%
	I feel quite sad today	5	3.1%
	I feel very sad today	10	6.3%
CHU-9D Pain	I don't have any pain today	106	66.3%
	I have a little bit of pain today	28	17.5%
	I have a bit of pain today	12	7.5%
	I have quite a lot of pain today	2	1.3%
	I have a lot of pain today	12	7.5%
CHU-9D Tired	I don't feel tired today	67	41.9%
	I feel a little bit tired today	42	26.3%
	I feel a bit tired today	13	8.1%
	I feel quite tired today	11	6.9%
	I feel very tired today	27	16.9%
CHU-9D Annoyed	I don't feel annoyed today	118	73.8%
	I feel a little bit annoyed today	9	5.6%
	I feel a bit annoyed today	14	8.8%
	I feel quite annoyed today	8	5.0%
	I feel very annoyed today	11	6.9%
CHU-9D School Work	I have no problems with my schoolwork today	112	70.4%
	I have a few problems with my schoolwork today	28	17.6%
	I have some problems with my schoolwork today	8	5.0%
	I have many problems with my schoolwork today	4	2.5%
	I can't do my schoolwork today	7	4.4%
CHU-9D Sleep	Last night I had no problems sleeping	93	58.1%
	Last night I had a few problems sleeping	27	16.9%
	Last night I had some problems sleeping	6	3.8%
	Last night I had many problems sleeping	3	1.9%
	Last night I couldn't sleep at all	31	19.4%
CHU-9D Routine	I have no problems with my daily routine today	135	84.4%
	I have a few problems with my daily routine today	12	7.5%
	I have some problems with my daily routine today	6	3.8%
	I have many problems with my daily routine today	1	.6%
	I have many problems with my daily routine today	6	3.8%
CHU-9D Activities	I can join in with any activities today	117	73.1%
	I can join in with most activities today	22	13.8%
	I can join in with some activities today	10	6.3%
	I can join in with a few activities today	6	3.8%
	I can join in with no activities today	5	3.1%

Table 6 Testing convergent validity hypotheses

Dimension and Measure	Dimension and Measure	Correlation Coefficient (Spearman's Rho)	Sig. Level (2 tailed)
EQ-5D-Y Mobility	CHU-9D Activities	.224**	0.00
EQ-5D-Y Looking After Myself	CHU-9D Routine	.338**	0.00
EQ-5D-Y Usual Activities	CHU-9D Activities	.347**	0.00
EQ-5D-Y Pain	CHU-9D Pain	.454**	0.00
EQ-5D-Y Pain	PedsQL Phys. Funct. Summary Score	-.345**	0.00
EQ-5D-Y Worried	CHU-9D Worried	.337**	0.00
EQ-5D-Y Worried	CHU-9D Sad	.376**	0.00
EQ-5D-Y Mobility	PedsQL Phys. Funct. Summary Score	-.145	0.07
EQ-5D-Y Worried	PedsQL Psycho-Social Summary Score	-.291**	0.00
CHU-9D School Work	PedsQL Psycho-Social Summary Score	-.324**	0.00
CHU-9D Sad	PedsQL Psycho-Social Summary Score	-.311**	0.00
CHU-9D Activities	PedsQL Phys. Funct. Summary Score	-.248**	0.00
CHU-9D Worried	PedsQL Psycho-Social Summary Score	-0.199*	0.01

Figure 1 Distribution of utility values for the EQ-5D-Y and CHU-9D