

**Work in progress: please do not quote without authors' permission**

**‘Nothing about us without us?’**

**Reflections on the utilisation of best worst scaling**

**discrete choice experiment methods to obtain adolescent specific values**

**for the Child Health Utility 9D**

**Julie Ratcliffe<sup>1</sup>, Terry Flynn<sup>2</sup>, Frances Terlich, Katherine Stevens<sup>3</sup>, John Brazier<sup>3</sup>, Michael Sawyer<sup>4</sup>**

1. Flinders Clinical Effectiveness, Flinders University, South Australia
2. Centre for the Study of Choice, University of Technology, Sydney
3. Health Economics and Decision Science, SchARR, University of Sheffield
4. Discipline of Paediatrics, University of Adelaide, South Australia

**Paper presented to the Health Economists Study Group Meeting**

**Bangor University**

**29<sup>th</sup> June-1<sup>st</sup> July 2011**

**Corresponding author:**

Julie Ratcliffe

Flinders Clinical Effectiveness

Flinders University

Bedford Park

South Australia

Tel +61 8 8275 2858

E-mail: [julie.ratcliffe@flinders.edu.au](mailto:julie.ratcliffe@flinders.edu.au)

## **Background**

Adolescence is a transitional stage of physical and mental human development generally occurring between the ages of 11 and 17 years, commencing at the onset of puberty and terminating at legal adulthood (Williams et al., 2002). It is a time when the prevalence of several health risk behaviours, e.g. alcohol use, cigarette smoking and illicit drug use, rises markedly and individuals become increasingly responsible for their own health and health care. As such, this period of development is a key point when educational and preventative efforts may have a significant impact upon both short and long term health outcomes. Since the publication of the United Nations Convention on the Rights of the Child (United Nations, 1989), the importance of involving adolescents in decision-making related to their own health and health care and of recognising adolescents as consumers of health care in their own right has been recognised (Kleinert, 2007; Tylee et al., 2007). The recent influential report by the National Health and Hospitals Reform Commission for Australia noted the need for more information in relation to adolescents' attitudes about their own health status and their views about adolescent orientated treatment programs (National Health and Hospitals Reform Commission, 2009). Such information is an essential prerequisite for the planning and development of preventive strategies and clinical treatment programs designed to improve adolescent health. Researchers in health economics and other disciplines are also recognising the importance of adolescent health and the need to measure and value the health related quality of life of both children and adolescents (Kleinert, 2007; Griebisch et al, 2005) for incorporation into the economic evaluation of adolescent health care treatment and preventive programs.

The Child Health Utility 9D [CHU9D], is a recently developed instrument that has been specifically designed for the estimation of quality adjusted life years [QALYs] for the economic evaluation of health care treatment and preventive programs targeted at young people (Stevens 2009; Stevens 2010, Stevens 2011). The dimensions of HRQoL for inclusion in the CHU9D instrument were identified from in-depth qualitative interviews with young people with a variety of chronic and acute health problems [n=74] which aimed to explore how their health affects their lives. The CHU9D has 9 attributes: worried, sad, pain, tired, annoyed, schoolwork, sleep, daily routine, ability to join in activities, with 5 different levels representing increasing levels of severity within each attribute. Whilst it was originally developed for use with younger children aged 7 to 11 years, there is increasing interest internationally in the application of the instrument with adolescents in the 11-17 year age group. Several recent studies have demonstrated the practicality, face and construct validity of the CHU9D in the Australian adolescent general population (Ratcliffe et al., 2011; Ratcliffe et al., 2011A; Ratcliffe et al., 2011B). The CHU9D has undergone psychometric testing in general UK primary school samples, and in clinical populations and has demonstrated good

practicality and validity (Stevens 2011). The instrument is currently being applied in a number of research programmes internationally focused upon the adolescent age group including the economic evaluation new innovative adolescent treatment programs for type 1 diabetes, attention deficit hyperactivity disorder, mental health, obesity prevention and liver transplantation. Presently there exists a single health state valuation algorithm for the CHU9D generated from health state valuation interviews with 300 members of the UK adult general population (Stevens, 2010A). The algorithm was developed by employing a statistical model to infer health state values for all possible health states defined by the CHU9D. This model was based upon the direct valuation of 64 CHU9D health states, reflecting a range of mild, moderate and severe health states.

Although there is no accepted gold standard scaling method for eliciting health state values for the estimation of QALYs, historically health economists have tended to favour the choice based valuation methods of Standard Gamble and Time Trade Off (Brazier et al., 2007). Both methods have traditionally been applied in adult populations. The existing health state valuation algorithm for the CHU9D utilises the Standard Gamble valuation method applied in an adult general population sample (Stevens, 2010A). Standard Gamble questions typically involve consideration of a probability of immediate death which may be considered ethically inappropriate and upsetting for adolescents to contemplate. The Time Trade Off was developed specifically for use in health care as a less complex alternative to Standard Gamble that overcomes the problems of explaining probabilities to respondents (Torrance, 1976). However, Time Trade Off questions typically also involve consideration of death and there is evidence to suggest that duration effects and time preference effects can have an impact on the elicitation of health state values derived from Time Trade Off questions (Arnesen and Trommald, 2005). In addition, both methods place a considerable cognitive burden on respondents who are required to evaluate a series of separate health states successively until the point of indifference is found.

Discrete Choice Experiments [DCEs] have their theoretical basis in random utility theory (Thurstone, 1927; Lancaster, 1966; McFadden, 1973). Random utility theory is based upon a psychological model that relates the probability of choosing a particular item in preference to other items to the unobserved cardinal utility, or total benefit, associated with each health state. DCEs are usually operationalised in conjunction with a valuation function that relates the mean utility for a given item (in this case health state) to a set of explanatory variables (McFadden, 1973). This operationalisation of DCEs is usually conducted within a conditional logistic regression model and its generalisations (McFadden, 1973). DCEs have become a very popular tool for eliciting preferences for health care process attributes in addition to, or in isolation of, health outcomes (Ratcliffe and Buxton, 1999). DCEs have also been applied to estimate values for different health or

quality of life states (Coast et al., 2008; Hakim et al., 1999; Ryan et al., 2005; Ratcliffe et al., 2009). Profile case best worst scaling [BWS] offers an attractive option for health state valuation exercises with vulnerable population groups e.g. adolescents and older people since it involves a potentially easier choice task to traditional DCE (Flynn, 2010). Traditional DCE involves presenting the respondent with a number of choices in which they are required to indicate their preferences *between* two or more health states whereas profile case BWS presents the respondent with a number of choices represented by *one health state only* and the respondent is asked to indicate the best and worst attribute *within* the health state under consideration (Flynn et al., 2007; Flynn et al., 2008). Profile case BWS also allows the impact of attributes to be compared meaningfully on a common scale which is not possible using traditional DCE methodology (Flynn et al., 2007; Marley et al., 2008). Profile case BWS has recently been successfully applied in a number of studies including the estimation of a cancer patient specific scoring algorithm for the EQ-5D and to develop scoring algorithms for the ICECAP-O capability index for older people and the Carer Experience Scale (Szeinbach et al., 1999; Coast et al., 2008; Al Janabi et al., 2010).

The main objective of this pilot study was to apply profile case BWS to obtain adolescent specific values for the CHU9D. A secondary aim was to assess the feasibility of a web based method of data collection for the valuation of health states defined by the CHU9D. A major purported advantage of a web based method of data collection is its ability to engage specific respondent groups, including community based adolescents, who may otherwise be hard to reach in sufficient numbers (Gwaltney et al., 2008). A web based method of data collection also has potential cost and time saving advantages in comparison with large-scale mail out surveys or face to face interviews.

## **Methods**

A web based survey was developed for administration to a community based sample of adolescents aged 11-17 years recruited from an on-line panel company following parent and adolescent consent for participation. The survey included three main sections. In Section A, respondents were asked to complete the CHU9D instrument. In addition to providing an indicator of their overall health, completion of the CHU9D helped to familiarise respondents with the wording, formatting and range of each of the 9 attributes of the CHU9D prior to the profile case BWS task. In Section B, respondents were presented with a series of CHU9D health states and asked to indicate the best and worst attribute for each health state. As it is not feasible to present every possible health state to participants for valuation, the full factorial generates  $5^9 = 1,953,125$  health states, a fractional factorial design was generated to reduce the number of health states to a manageable number for the purposes of a web based survey whilst retaining maximum statistical efficiency for the estimation of model parameters. Whilst it would have been ideal to estimate main effects and all two way

interactions, a fractional factorial design that would allow the estimation of these effects would still have too many health states to be practical. A fractional factorial that permitted the estimation of main effects, whilst maintaining the design properties of level balance and near orthogonality was generated in 50 health states. Complete orthogonality in the design was not possible due to the need to eliminate a small number of implausible health states (Louviere et al., 2000; Burgess and Street 2006). We blocked the design into 5 versions so that each participant was presented with a maximum of 10 health states for the CHU9D. Blocking the design has been demonstrated previously to promote participant completion rates and minimise error due to fatigue (Hensher et al., 2005). The 10 CHU9D health states in each block were chosen to include a range of mild, moderate and severe health states. Each health state description consisted of the 9 common attributes of the CHU9D with different levels for each of the 10 health states presented. Participating adolescents were asked to indicate the best and worst attribute levels of each health state. An example question from the survey is included in Appendix 1.

Basic socio-demographic information including age and gender and additional questions relating to general health status and whether or not the respondent had a disability or long standing health condition, were collected in the final section of the survey. Socio-economic status was measured by applying the Family Affluence Scale (Anderson et al., 2008), a measure of socioeconomic position designed for self report by adolescents age 11-17 years. The instrument includes four items relating to family affluence: (1) “Does your family own a car, van or truck?” [no/yes one/yes two or more] (2) “Do you have your own bedroom for yourself?” [no/yes] (3) “During the past 12 months how many times did you travel away on holiday/vacation with your family?” [not at all/once/twice/more than twice], (4) “How many computers does your family own?” [none/one/two/more than two]. The Family Affluence Scale is constructed as a 0-7 point scale with 1 point each for having one car, one computer and one room and one extra point each for having more than one car, holiday or computer. Hence lower scores represent lower levels of affluence and vice versa. The Family Affluence Scale was collected in eight categories ranging from 0 to 7, which were recoded into 3 groups for the analysis, low:0-3, intermediate:4-5 and high (Ravens-Sieberer et al., 2010). Participants were also asked to indicate how difficult they found the task was to complete on a scale from 1 to 4 where 1 indicates ‘not difficult’ and 4 indicates ‘very difficult’.

### ***Data analysis***

The individual responses to the CHU9D from Section A of the web based survey were converted to utilities using the existing UK adult general population algorithm developed by Stevens (Stevens, 2010A) based upon the standard gamble method. The difference in mean utilities between groups

according to respondent characteristics was assessed by employing the Kruskal-Wallis one way ANOVA and pairwise comparison Mann-Whitney U tests.

The profile case BWS data from section B was employed to estimate a sequential marginal model using conditional logit regression in the statistical package STATA for the prediction of CHU9D health state values (Flynn et al., 2008A; Marley et al., 2008). Marginal methods of analysis model the possible attribute levels that can be chosen. The profile case BWS data was aggregated over best–worst pairs (which ‘sums to the margins’ of tables of all possible best-worst pairs) to estimate the attribute level utilities. Analysing choices in a random utility framework implies that  $U_{iq}$ , the utility respondent  $q$  derives from choosing item  $i$ , is split into an explainable component ( $V_{iq}$ ) and a random component ( $\varepsilon_{iq}$ ).

The equation to be estimated was of the following form:

$$U_{iq} = V_{iq} + \varepsilon_{iq}$$

where  $V_{iq} = \beta_{11}$  worried\_not<sub>i</sub> +  $\beta_{12}$  worried\_little<sub>i</sub> +  $\beta_{13}$  worried\_bit<sub>i</sub> +  $\beta_{14}$  worried\_quite<sub>i</sub> +  $\beta_{15}$  worried\_very<sub>i</sub> +  $\beta_{21}$  sad\_not<sub>i</sub> +  $\beta_{22}$  sad\_little<sub>i</sub> +  $\beta_{23}$  sad\_bit<sub>i</sub> +  $\beta_{24}$  sad\_quite<sub>i</sub> +  $\beta_{25}$  sad\_very<sub>i</sub> +  $\beta_{31}$  pain\_not<sub>i</sub> +  $\beta_{32}$  pain\_little<sub>i</sub> +  $\beta_{33}$  pain\_bit<sub>i</sub> +  $\beta_{34}$  pain\_quite<sub>i</sub> +  $\beta_{35}$  pain\_very<sub>i</sub> +  $\beta_{41}$  tired\_not<sub>i</sub> +  $\beta_{42}$  tired\_little<sub>i</sub> +  $\beta_{43}$  tired\_bit<sub>i</sub> +  $\beta_{44}$  tired\_quite<sub>i</sub> +  $\beta_{45}$  tired\_very<sub>i</sub> +  $\beta_{51}$  annoyed\_not<sub>i</sub> +  $\beta_{52}$  annoyed\_little<sub>i</sub> +  $\beta_{53}$  annoyed\_bit<sub>i</sub> +  $\beta_{54}$  annoyed\_quite<sub>i</sub> +  $\beta_{55}$  annoyed\_very<sub>i</sub> +  $\beta_{61}$  schoolwork\_not<sub>i</sub> +  $\beta_{62}$  schoolwork\_little<sub>i</sub> +  $\beta_{63}$  schoolwork\_bit<sub>i</sub> +  $\beta_{64}$  schoolwork\_quite<sub>i</sub> +  $\beta_{65}$  schoolwork\_very<sub>i</sub> +  $\beta_{71}$  sleep\_not<sub>i</sub> +  $\beta_{72}$  sleep\_little<sub>i</sub> +  $\beta_{73}$  sleep\_bit<sub>i</sub> +  $\beta_{74}$  sleep\_quite<sub>i</sub> +  $\beta_{75}$  sleep\_very<sub>i</sub> +  $\beta_{81}$  routine\_not<sub>i</sub> +  $\beta_{82}$  routine\_little<sub>i</sub> +  $\beta_{83}$  routine\_bit<sub>i</sub> +  $\beta_{84}$  routine\_quite<sub>i</sub> +  $\beta_{85}$  routine\_very<sub>i</sub> +  $\beta_{91}$  activities\_not<sub>i</sub> +  $\beta_{92}$  activities\_little<sub>i</sub> +  $\beta_{93}$  activities\_bit<sub>i</sub> +  $\beta_{94}$  activities\_quite<sub>i</sub> +  $\beta_{95}$  activities\_very<sub>i</sub>

Where  $\beta_{11}$  refers to the coefficient on the variable for attribute 1 level 1 (worried\_not<sub>i</sub>),  $\beta_{12}$  the coefficient on the variable for attribute 1 level 2 (worried\_little<sub>i</sub>) etc.

Assuming that the random components are distributed extreme value type 1 (EV1) enables choice data to be analysed using the conditional (multinomial) logit model:

$$P_{iq} = \frac{e^{\lambda v_{iq}}}{\sum_{j \in C} e^{\lambda v_{jq}}}$$

Where  $P_{iq}$  is the probability that respondent  $q$  chooses alternative  $i$ ,  $j$  represents all the relevant alternatives in choice set  $C$ , and  $\lambda$  represents the EV1 scale parameter which is inversely proportional to the standard deviation of the random component  $\sigma_{iq} = \pi\lambda^{-1}6^{-1/2}$

The estimates from the BWS DCE task are initially anchored to the least valued attribute level. Since these estimates are on an interval scale, a linear transformation can be applied in order that the full health state takes the value 1 and the 'PITS' health state [the health state comprising the lowest level on each of the nine attributes of the CHU9D descriptive system] takes the value 0 (Coast et al., 2008). However, in order for the estimates to have QALY properties 0 must represent the death state, not the PITS state. This can be achieved (mathematically and conceptually) by using the most severe or PITS health state value from a separate TTO or SG exercise to rescale the original estimates to ensure that the 0 represents death to enable the estimation of QALYs (Flynn, 2010A). In this case, the existing adult general population value for the most severe or PITS health state, the health state comprising the lowest level on each of the nine attributes of the CHU9D descriptive system, from the UK adult general population scoring algorithm was used to re-scale the estimates from the profile case BWS to ensure that the zero represented death, rather than the utility of the most severe health state (Ratcliffe et al., 2009; Flynn et al., 2008A; Flynn 2010A).

Finally, a selection of health states defined by the CHU9D instrument were scored using the new adolescent specific scoring algorithm. The health state values generated were then compared with the values for identical health states generated from application of the existing adult general population scoring algorithm.

## **Results**

The completion rate for the survey was 70%, with 590 of the total sample of consenting respondents [n=843] fully completing the survey. The characteristics of the respondents are shown in Table 1. The respondents were reasonably balanced in gender with 56% of respondents being male and there were a similar proportion of respondents in each age group [11 to 17 years]. A relatively small proportion of respondents [11%] indicated that they were living with a long standing illness or disability. The vast majority of respondents [70%] reported themselves as in excellent or very good general health. The CHU9D values corresponding to respondent characteristics, generated by application of the existing adult general population algorithm, are also presented in Table 1. The CHU9D was able to discriminate between respondents according to their self-reported general health, with the mean CHU9D utilities ranging from 0.926 for those who reported themselves in excellent health to 0.655 for those who reported themselves in poor health and these differences

were statistically significant [ $P < 0.001$ , Kruskal-Wallis Test]. The mean CHU9D utility for the minority of respondents who reported themselves as living with a long standing illness or disability [ $n=67$ ] was also found to be lower than the majority [ $n=440$ ] who reported themselves as living without long standing illness or disability, and these differences were also found to be statistically significant [ $P=0.002$  Mann-Whitney  $U$  test].

Differentiation according to socioeconomic position as measured by the Family Affluence Scale resulted in differences in utilities for the CHU9D with higher levels of reported affluence being associated with higher values on average and the differences in mean values between groups according to socioeconomic position were statistically significant [ $P=0.001$ , Kruskal-Wallis Test]. The frequencies of responses to the CHU9D are presented in Table 2. Participants generally reported themselves in good health according to the CHU9D classification, with 68 respondents [11%] reported themselves at the highest level for all 9 CHU9D attributes. The CHU9D health state values corresponding to these response patterns ranged from a minimum of 0.33 to a maximum of 1.00 with a mean of 0.85.

The results from the re-scaled conditional logit model to estimate part-worth utilities for all attribute levels on the QALY scale are presented in Table 3. Re-anchoring the original conditional logit model ensured that the PITS or most severe health state [55555555] adolescent specific value was identical to that generated by application of the existing UK adult scoring algorithm [0.33]. It can be seen that the highest level for the activities attribute, being able to join in with all activities today, exhibited the greatest impact upon utility. The lowest level for the annoyed attribute, I feel very annoyed today, exhibited the lowest impact upon utility. Figure 1 illustrates the results from the re-scaled conditional logit model presented in a graphical format. Relatively large differences in value were found between the highest and lowest levels [indicating “no problems” and “severe problems”] for all nine attributes relating to the CHU9D. In general, there was little differentiation between the three middle levels for all attributes indicating only limited additional value for adolescents of moving between these levels. In common with the findings from our previous pilot study (Ratcliffe et al., 2011) four of the five levels relating to the attribute activities were valued relatively highly, with a large difference in value between the middle levels and the lowest level for this particular attribute, a unique pattern which was not apparent for any of the other CHU9D attributes.

Summing the utilities for each attribute level enables estimation of a total utility for every possible health state defined by the CHU9D. Table 4 presents predicted health state values grouped according to the lower levels, levels 3-5, of the mental health, daily activities and physical health



attributes. It can be seen that the largest differences in values occur in relation to the mental health attributes of the CHU9D: worried, sad and annoyed, with application of the adolescent algorithm producing consistently lower mean values than the adult algorithm for identical CHU9D mental health impairment states with differences ranging between 0.083 and 0.121 on the 0-1 QALY scale. These results suggest that adolescents may place more weight upon the attributes relating to mental health than would be implied by application of the existing algorithm based upon adult values.

The mean time taken to complete the survey was 13 minutes (SD: 37 minutes), although it is important to note that it was not possible to identify respondents who did not complete the survey in a single session. Table 5 indicates that the majority of respondents who completed the survey (n=306, 52%) indicated that they found the survey not difficult to complete. Comments received about the survey, where indicated, were also positive in general with very few of the completing respondents indicating that they found significant problems in understanding it.

## **Discussion**

The profile case BWS approach exhibited good completion rates and our pilot study findings indicate that it represents a practical and feasible methodology for the valuation of health states with this age group. The pilot study findings indicate that there may be important differences in the values attached to CHU9D attributes between adolescents and adults with adolescents, in general, appearing to place more weight upon the attributes relating to mental health. However, it should be noted that the existing adult based scoring algorithm for the CHU9D was produced using an alternative method, the Standard Gamble, an approach that explicitly varied length of life within the preference based task. Use of a discrete choice based task that omits length of life (as was the case here) may produce different estimates for identical health states. Thus the difference in findings may be due to a method effect rather than a true difference in underlying values between adolescents and adults. It is important that further research is conducted to substantiate these preliminary findings in larger, more representative community based samples of adolescents and to make direct comparisons with adult populations utilising an identical profile case BWS approach to indicate whether true differences in CHU9D health state values may or may not exist for these two distinct population groups. Further research should properly adjust for any variance heterogeneity if the BWS population estimates are to demonstrate unbiasedness (Flynn et al., 2010B). It is also important to assess the implications of any differences between adolescents and adults in the values attached to identical CHU9D health state values for economic evaluation: in terms of their potential impact upon the results of cost effectiveness analyses of preventative strategies and adolescent treatment and service programmes.

There are difficulties associated with valuation of the state dead within ordinal tasks such as ranking and discrete choice experiments and these were first described in detail by Flynn and colleagues (Flynn et al., 2008A). Specifically, use of a statistical model such as conditional (multinomial) regression to anchor quality of life values from ordinal data to death is inappropriate in the presence of respondents who do not conform to the assumptions of conventional random utility theory. This is clearest when estimating values for that group of respondents observed in valuation samples who refuse to consider any living state to be worse than death as in such circumstances the model cannot be estimated. Craig et al., have since described the state dead as “the inextricable anchor of the quality-adjusted life years scale” (Craig et al., 2009).

Since the health state values from the profile case BWS exercise are anchored to the least valued attribute level, they must be re-scaled to the full health - death state required for the estimation of QALYs. Flynn and colleagues indicate that this can be achieved (mathematically and conceptually) by using the most severe or PITS health state value from a separate TTO or SG exercise to rescale the original estimates to ensure that the 0 represents death to enable the estimation of QALYs (Flynn, 2010A). In this study, the existing adult general population value for the most severe or PITS health state, from the UK adult general population scoring algorithm was used to re-scale the estimates from the profile case BWS to ensure that the zero represented death, rather than the utility of the most severe health state. However, this may be viewed as a limitation and further research should assess the possibility of obtaining adolescent specific value/s for the purposes of re-scaling the profile case BWS estimates. The findings from our previously published pilot study indicate that TTO and SG tasks are unreliable in younger adolescents, aged 11-13 years (Ratcliffe et al., 2011). However, it is possible that older adolescents may exhibit a stronger understanding and consequent ability to complete such tasks. Preliminary evidence from a study to obtain adolescent specific values for the AQOL-6D has indicated that a modified version of a TTO task conducted with a sample of older adolescents (aged 15 years plus) may be practical and feasible (Moodie et al., 2010). Further research should be conducted to assess this possibility for health states defined by the CHU9D.

Recently, researchers in Australia and Canada have utilised the traditional DCE approach to generate health state values for other descriptive systems without the need for re-scaling, most notably the EQ-5D, by the direct inclusion of a separate length of life attribute within the DCE choice task (Viney et al., 2011; Bansback et al., 2009). Both of these studies were conducted in adult samples. The inclusion of a separate length of life attribute involves interacting a significant number of length of life levels with, at a minimum, the fractional factorial design that permits the estimation of main effects. This approach would be very difficult to conduct to obtain adolescent

specific values for the CHU9D for two main reasons. Firstly, the CHU9D descriptive system itself is a relatively large descriptive system and it would necessitate the direct presentation for valuation of many health states. Secondly, the inclusion of an additional length of life attribute within the traditional DCE approach would introduce a further element of complication to the valuation task. Respondents would be required to choose the most preferred health state [where the health states for presentation would include a length of life attribute in addition to the CHU9D attributes], from a choice set of two or more competing states, repeated over a necessarily large number of choice sets to observe trade offs.

The findings from this pilot study provide support for the practicality and face validity of the CHU9D for application with adolescents in the general population aged 11-17 years. A relatively high proportion of the total sample of participating respondents successfully completed the CHU9D. As expected, in this community based sample, the vast majority of respondents reported themselves in good health, with those reporting themselves as living with a long standing illness or disability having lower CHU9D values on average than those reporting themselves as living without any long standing illness or disability. The CHU9D was also found to be able to discriminate between groups with known differences based upon self-reported ratings of general health. These findings of construct validity are consistent with a number of separate studies we have recently conducted in other community based samples of adolescents (Ratcliffe et al., 2011A, Ratcliffe et al., 2011B).

The advantages of a web based mode of administration for a survey of this nature include its increasing familiarity, particularly for young people and its ability to engage large numbers of community based adolescents who would otherwise be more difficult to reach. Potential disadvantages include concerns about response rates, data quality, that participants may not provide accurate information and/or the intended recipient may not be the person who completes the survey (Gwaltney et al., 2008). It should be noted that these problems are not unique to a web based mode of administration as they may also arise with self-report pen and paper surveys. It has recently been recommended that careful explanation of the importance of a web based survey at the outset may lead to improvements in the thoughtfulness of responses (Norman et al., 2010). This survey included an introductory section which explained the purpose of the survey and stressed the absolute importance of its completion by adolescents age 11-17 years only, following parental and adolescent consent, for the validity of responses. The survey was completed by a majority of respondents and very few indicated significant problems in understanding it. However, a limitation of this pilot study is that the sample was not representative of the Australian adolescent general population. The majority of respondents were of high socio-economic status as defined by the

Family Affluence Scale [52%, n=306, classified in Family Affluence Scale sub-groups 6 and 7] and were computer literate and likely better educated. Hence the findings relating to the overall high reported levels of health status as measured by the CHU9D instrument, respondent understanding and relatively high completion rates for the profile case BWS approach may not be generalisable. Further research is required to assess the reliability of these findings in other, more diverse, community based samples of adolescents and clinical groups.

Overall, the findings from this pilot study lend support to the potential future application of profile case BWS methods to undertake large scale health state valuation studies directly with young adolescent population samples. The study findings also provide support for the feasibility and acceptability of a web based mode of administration for this purpose. Further research is required to substantiate these preliminary findings in larger, representative samples of adolescents and to assess the potential implications of differences between adolescents and adults in the values attached to identical CHU9D health states for the economic evaluation of adolescent treatment and service programmes.

### **Acknowledgements**

This study was approved by the Flinders University Social and Behavioural Research Ethics Committee-project number 4701. Financial support from a Flinders University seeding grant is gratefully acknowledged. The usual disclaimer applies.

## References

- Al Janabi H, Flynn T, Coast J. (2010). Estimation of a Preference-Based Carer Experience Scale. *Medical Decision Making* 2011;31: 458-468
- Anderson A, Krolner R, Currie C et al. High agreement on family affluence between children and parents' self reports: international study of 11 year old children. *Journal of Epidemiology and Community Health* 2008; 62: 1092-1094.
- Arnesen T, Trommald M. Are QALYs based on time trade off comparable? – A systematic review of TTO methodologies. *Health Economics* 2005; 14: 39-53.
- Bansback N, Brazier J, Tsuchiya A, Anis A. A comparison of using Discrete Choice Experiments and the Time Trade Off to value health states for Quality Adjusted Life Years. Health Economists' Study Group Meeting, July 2009, University of Sheffield.
- Brazier J, Ratcliffe J, Salomon J, Tsuchiya A. *Measuring and valuing health benefits for economic evaluation*. Oxford University Press, 2007.
- Burgess L, Street D. The optimal size of choice sets in choice experiments *Statistics* 2006; 40(60): 507-515.
- Craig B, Busschbach J, Salomon J. Keep it simple: Ranking health states yields values similar to cardinal measurement approaches. *Journal of Clinical Epidemiology* 2009; 62: 296-305.
- Coast J, Flynn T, Natarajan L, Sproston K, Lewis J, Louviere J, Peters T. Valuing the ICECAP capability index for older people. *Social Science and Medicine* 2008; 67: 874-992.
- Flynn TN, Louviere JJ, Peters TJ, Coast J (2007). Best-Worst Scaling: What it can do for health care research and how to do it. *Journal of Health Economics* 26(1):171-37.
- Flynn T, Louviere J, Peters T, Coast J. Estimating preferences for a dermatology consultation using Best-Worst Scaling: Comparison of various methods of analysis. *BMC Medical Research Methodology* 2008; 8:76.
- Flynn T, Louviere J, Marley A, Peters T, Coast J. Rescaling quality of life tariffs from discrete choice experiments for use as QALYs: a cautionary tale. *Population Health Metrics* 2008A; 6:6.
- Flynn T. Valuing citizen and patient preference in health: recent developments in three types of best-worst scaling. *Expert Rev. Pharmacoeconomics Outcome Res.* 2010; 10(3): 259-267.
- Flynn T. Using Conjoint Analysis and Choice Experiments to Estimate QALY Values: Issues to Consider. *Pharmacoeconomics* 2010A; 28(9): 711-722.
- Flynn T, Louviere J, Peters T, Coast J. Using discrete choice experiments to understand preferences for quality of life. Variance scale heterogeneity matters. *Social Science and Medicine* 2010B; 70: 1957-1965.
- Griebsch I, Coast J, Brown J. Quality adjusted life years lack quality in paediatric care: A critical review of published studies. *Paediatrics* 2005; 115: 600-614.
- Gwaltney C, Sheilds A, Shiffman S. Equivalence of electronic and paper and pencil administration of patient-reported outcome measures: A meta-analytic review. *Value in Health* 2008;11(2): 322-333.
- Hakim Z, Pathak D. Modelling the EuroQol data: a comparison of discrete choice conjoint and conditional preference modelling. *Health Economics* 1999; 82: 103-16.
- Hensher D, Rose J and Greene W. *Applied Choice Analysis, A Primer*. 2005; Cambridge University Press.
- Klein S. Adolescent health: an opportunity not to be missed. *The Lancet* 2007; 369: 1057-1058.
- Lancaster K. A new approach to consumer theory *Journal of Political Economy* 1966; 74:134-157.
- Louviere J, Hensher D, Swait J. *Stated Choice Methods: Analysis and Application*. Cambridge University Press, 2000.
- Marley AAJ, Flynn TN, Louviere JJ. Probabilistic Models of Set-Dependent and Attribute-Level Best-Worst Choice. *Journal of Mathematical Psychology* 2008; 52:281-296.
- McFadden D. *Conditional logit analysis of qualitative choice behaviour*. 1973; University of California at Berkley, California.
- Moodie M, Richardson J, Rankin B, Sinha K, Lezzi A. Predicting time trade-off health state valuations of adolescents in four pacific countries using the AQoL-6D instrument. *Value in Health* 2010; 13: 1014-1027.

National Health and Hospitals Reform Commission (NHHRC): *A Healthier Future For All Australians: Final Report of the National Health and Hospitals Reform Commission*, June 2009.

Norman R, King M, Clarke D, Viney R, Cronin P, Street D. Does mode of administration matter? Comparison of online and face to face administration of a time trade off task. *Quality of Life Research* 2010; 19: 499-508.

Ratcliffe J, Brazier J, Tsuchiya A, Symmonds T, Brown M. A comparison of time trade off, discrete choice experiment and ranking data in estimating a preference based single index from the sexual quality of life questionnaire. *Health Economics* 2009;18: 1261-1276.

Ratcliffe J, Buxton M. Patient's preferences regarding the process and outcomes of life saving technology: an application of conjoint analysis to liver transplantation. *International Journal of Technology Assessment in Health Care* 1999; 15: 340-351.

Ratcliffe J, Couzner L, Flynn T, Sawyer M, Stevens K, Brazier J (2011) Assessing the feasibility of applying best worst scaling discrete choice methods to value Child Health Utility 9D health states in a young adolescent sample. *Applied Health Economics and Health Policy* 2011; 9(1):15-27.

Ratcliffe J, Stevens K, Brazier J, Flynn T, Sawyer M (2011A). An assessment of the construct validity of the CHU9D in the Australian adolescent general population. Submitted for publication.

Ratcliffe J, Stevens K (2011B). The measurement and valuation of health related quality of life in adolescents: an assessment of the practicality and validity of the Child Health Utility 9D in the Australian population. Paper submitted for presentation at the International Health Economics Association meeting, Toronto, Canada, July 2011.

Ravens-Sieber U, Erhart M, Rajmil L et al. Reliability, construct and criterion validity of the KIDSCREEN-10 score: a short measure for children and adolescents' well-being and health-related quality of life. *Quality of Life Research* 2010; 19: 1487-1500.

Ryan M, Netten A, Skatun D, Smith P. Using discrete choice experiments to estimate a preference based measure of outcome - an application to social care for older people. *Journal of Health Economics* 2005; 25: 927-44.40.

Stevens K.J, *Developing a preference-based measure of health in children*. PhD Thesis, ScHARR, University of Sheffield, 2008.

Stevens K.J. Developing a descriptive system for a new preference-based measure of health related quality of life for children. *Quality of Life Research* 2009; 18: 1105-1113.

Stevens K.J. Working with children to develop dimensions for a preference-based generic paediatric health related quality of life measure. *Qualitative Health Research* 2010; 20: 340-351.

Stevens K.J. Valuation of the Child Health Utility Index 9D (CHU9D). Health Economics and Decision Science Discussion Paper 10/07, ScHARR, University of Sheffield, 2010A.

Stevens K J. 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. *Applied Health Economics and Health Policy* 2011; 9 (3): 1-13

Szeinbach S, Barnes J, McGhan W. Using conjoint analysis to evaluate health state preferences. *Drug Inf J* 1999; 33: 849-858.

Thurstone LL. A law of comparative judgement. *Psychological Review* 1927; 34: 273-86.

Torrance GW. Social preferences for health states: an empirical evaluation of three measurement techniques. *Socio-Econ.Plan.Sci.* 1976; 10: 129-136.

Tylee A, Haller D, Graham T, Churchill R, Sanci L. Youth friendly primary care services: how are we doing and what more needs to be done. *The Lancet* 2007; 369: 1585-1573.

United Nations. *Convention on the rights of the child*. HMSO, London, 1989.

Viney R, Norman R, King M et al. Using a Discrete Choice Experiment to Value EQ-5D Health States. Paper submitted to the 8<sup>th</sup> World Congress of the International Health Economics Association, Toronto, Canada, July 2011.

Williams P, Holmbeck G, Greenley R. Adolescent Health Psychology. *Journal of Consulting and Clinical Psychology* 2002; 70: 828-842.

**Table 1: CHU9D utilities according to the characteristics of respondents**

<b>Characteristic</b>	<b>N=590</b>	<b>CHU9D utilities<sup>1</sup> Mean (SD)</b>
Number of females (%)	268 (45.3)	0.858 (0.117)
Number of males (%)	322 (54.7)	0.844 (0.114)
Mean age (SD)	14.5 (1.99)	
Age range (IQ range)	11-17 (13-16)	0.851 (0.115)
Long standing health condition or disability**:		
Yes (%)	67 (11.4)	0.803 (0.138)
No (%)	523 (88.6)	0.857 (0.138)
Socio-economic status**:		
Low Family Affluence Scale (%)	57 (9.7)	0.806 (0.134)
Medium Family Affluence Scale (%)	227 (38.5)	0.844 (0.118)
High Family Affluence Scale (%)	306 (51.9)	0.866 (0.107)
General health**:		
Excellent	145 (25)	0.926 (0.085)
Very Good	268 (45)	0.863 (0.097)
Good	129 (22)	0.792 (0.101)
Fair	39 (7)	0.731 (0.119)
Poor	9 (2)	0.665 (0.178)

<sup>1</sup> from application of the existing adult scoring algorithm.

\*\* indicates statistically significant difference between mean values at the 1% level.

**Table 2: Responses to the CHU9D**

<i>CHU9D Dimensions and levels</i>	<i>Frequency (%)</i>
<b>Worried</b>	
I don't feel worried today	363 (61%)
I feel a little bit worried today	126 (21%)
I feel a bit worried today	73 (13%)
I feel quite worried today	24 (4%)
I feel very worried today	6(1%)
<b>Sad</b>	
I don't feel sad today	429 (73%)
I feel a little bit sad today	91 (15%)
I feel a bit sad today	51 (9%)
I feel quite sad today	17 (3%)
I feel very sad today	4 (1%)
<b>Pain</b>	
I don't have any pain today	411 (69%)
I have a little bit of pain today	115 (19%)
I have a bit of pain today	43 (7%)
I have quite a lot of pain today	17 (3%)
I have a lot of pain today	6 (1%)
<b>Tired</b>	
I don't feel tired today	166 (28%)
I feel a little bit tired today	245 (41%)
I feel a bit tired today	102 (17%)
I feel quite tired today	60 (10%)
I feel very tired today	19 (3%)
<b>Annoyed</b>	
I don't feel annoyed today	342 (58%)
I feel a little bit annoyed today	147 (25%)
I feel a bit annoyed today	67 (11%)
I feel quite annoyed today	22 (4%)
I feel very annoyed today	14 (2%)
<b>Schoolwork/homework</b>	
I have no problems with my schoolwork/homework today	313 (53%)
I have a few problems with my schoolwork/homework today	160 (27%)
I have some problems with my schoolwork/homework today	74 (13%)
I have many problems with my schoolwork/homework today	25 (4%)
I can't do my schoolwork/homework today	20 (3%)
<b>Sleep</b>	
Last night, I had no problems sleeping	345 (58%)
Last night, I had a few problems sleeping	155 (26%)
Last night, I had some problems sleeping	61 (10%)
Last night, I had many problems sleeping	25 (4%)
Last night, I couldn't sleep at all	6 (1%)
<b>Daily routine</b>	
I have no problems with my daily routine today	418 (71%)
I have a few problems with my daily routine today	130 (22%)
I have some problems with my daily routine today	30 (5%)
I have many problems with my daily routine today	10 (2%)
I can't do my daily routine today	4 (1%)
<b>Activities</b>	
I can join in with any activities today	336 (57%)
I can join in with most activities today	138 (23%)
I can join in with some activities today	61 (10%)
I can join in with a few activities today	37 (6%)
I can join in with no activities today	20 (3%)

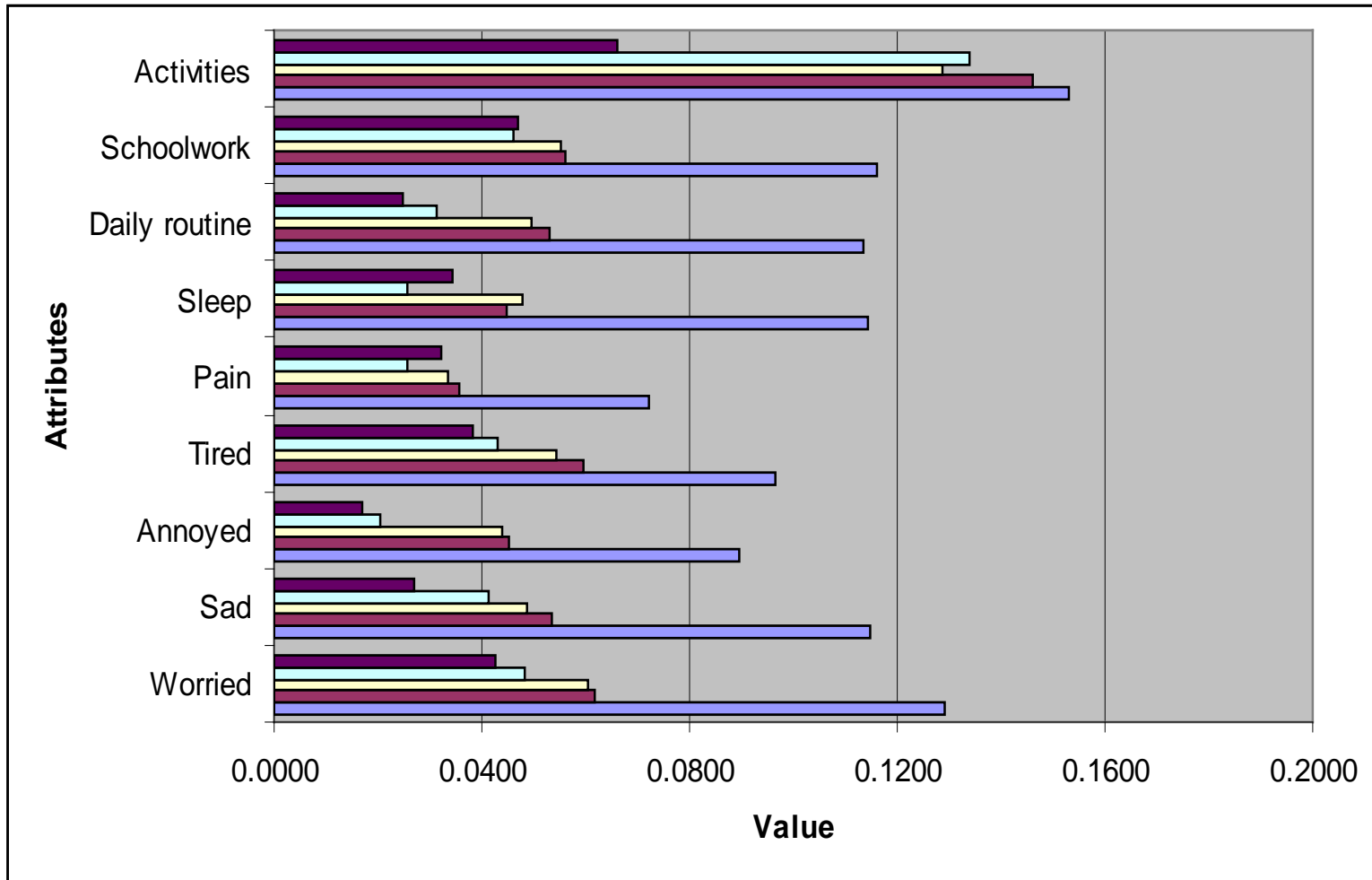


**Table 3: Re-scaled conditional logit estimates\***

<b>Worried</b>	Level 1	0.1292
	Level 2	0.0617
	Level 3	0.0602
	Level 4	0.0484
	Level 5	0.0425
<b>Sad</b>	Level 1	0.1148
	Level 2	0.0535
	Level 3	0.0485
	Level 4	0.0412
	Level 5	0.0269
<b>Annoyed</b>	Level 1	0.0895
	Level 2	0.0452
	Level 3	0.0438
	Level 4	0.0205
	Level 5	0.0169
<b>Tired</b>	Level 1	0.0967
	Level 2	0.0596
	Level 3	0.0542
	Level 4	0.0432
	Level 5	0.0381
<b>Pain</b>	Level 1	0.0722
	Level 2	0.0356
	Level 3	0.0335
	Level 4	0.0255
	Level 5	0.0323
<b>Sleep</b>	Level 1	0.1145
	Level 2	0.0449
	Level 3	0.0478
	Level 4	0.0255
	Level 5	0.0344
<b>Daily routine</b>	Level 1	0.1136
	Level 2	0.0529
	Level 3	0.0495
	Level 4	0.0313
	Level 5	0.0246
<b>Schoolwork</b>	Level 1	0.1163
	Level 2	0.0562
	Level 3	0.0551
	Level 4	0.0460
	Level 5	0.0469
<b>Activities</b>	Level 1	0.1531
	Level 2	0.1462
	Level 3	0.1287
	Level 4	0.1340
	Level 5	0.0661

\* Re-scaled such that the highest CHU9D health state, state 11111111, is equal to 1.00 and the lowest (PITS) CHU9D health state, state 55555555, is equal to 0.33, the corresponding PITS value from the adult SG algorithm

Figure 1: Graph of BWS CHU9D utilities by attribute level



**Table 4: Predicted health state values adult vs adolescent algorithm grouped by attributes**

<b>CHU9D Health state</b>	<b>Adult value</b>	<b>Adolescent value</b>	<b>Difference</b>	<b>Attributes at lower levels</b>
551151111	0.874	0.753	0.121	Worried, sad, annoyed
441141111	0.874	0.777	0.097	Mental health attributes
331141111	0.902	0.819	0.083	
111115155	0.734	0.755	-0.021	Schoolwork, daily routine,
111114144	0.828	0.828	-0.001	activities
111113133	0.853	0.850	0.003	Daily activities attributes
115511511	0.719	0.821	-0.103	Pain, tired, sleep
114411411	0.777	0.811	-0.034	Physical health attributes
113311311	0.898	0.852	0.046	

**Table 5: Level of difficulty indicated for survey**

<i>Level of difficulty</i>	<i>Frequency (%)</i>
Not difficult	306 (52%)
Slightly difficult	173 (29%)
Moderately difficult	86 (15%)
Very difficult	25 (4%)

## Appendix 1

We are now going to ask you a series of 10 questions in which we would like you to imagine you are living in the states of health described in each question.

For each **health state** we would like you to tell us which you think is the best feature and which is the worst feature by clicking on the button next to the feature you like best and the feature you think is worst. Although the health states appear similar the descriptions do differ in every health state so please read each one carefully before making your choices.

For Health State X if you think that ‘I don’t feel sad today’ is the best feature then you click on the button next to that feature in the column labelled ‘Best’. If you think that the worst feature is ‘I have many problems with my daily routine today’ then you click on the button next to that feature in the column labelled ‘Worst’.

<b>Best</b>	<b>Health State X</b>	<b>Worst</b>
	I feel quite worried today	
<b>O</b>	I don’t feel sad today	
	I have a bit of pain today	
	I feel quite tired today	
	I don’t feel annoyed today	
	I have a few problems with my schoolwork today	
	Last night I had some problems sleeping	
	I have many problems with my daily routine today	<b>O</b>
	I can join in with a few activities today	