Flinders Centre for Clinical Change and Health Care Research

Flinders University
inspiring achievement

Work in progress: please do not quote without authors' permission

A pilot study to apply best worst scaling discrete choice experiment methods to obtain adolescent specific values for the Child Health Utility 9D

Julie Ratcliffe[1], Terry Flynn[2], Frances Terlich[2], Katherine Stevens [3], John Brazier [3], Michael Sawyer[4]

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

Corresponding author:

Julie Ratcliffe

Flinders Clinical Effectiveness

Flinders University

Bedford Park

South Australia

Tel +61 8 8275 1297

E-mail: julie.ratcliffe@flinders.edu.au

Acknowledgements

This study was supported by a Flinders University seeding grant. The usual disclaimer applies.

Abstract

Aims: The main objective of this pilot study was to apply profile case best worst scaling discrete choice experiment methods to obtain adolescent specific values for the Child Health Utility 9D (CHU9D), a new generic preference based measure of health related quality of life developed specifically for application in cost effectiveness analyses of treatments and interventions targeted at young people. 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.

Methods: A web based survey was developed including three main sections. Section A comprised the CHU9D instrument. In Section B the respondent was asked to indicate the best and worst attribute levels from a series of 10 health states defined by the CHU9D presented one at a time. Section C comprised a series of socio-demographic questions. The survey was administered to a community based sample of consenting adolescents (n=590) aged 11-17 years. A conditional logistic regression model was applied to estimate values (part worth utilities) for each level of the 9 attributes relating to the CHU9D. The values were re-anchored by the existing adult general population value for the most severe or PITS health state to ensure that zero represented death, rather than the utility of the most severe health state. A marginal utility matrix was then estimated to generate an adolescent specific scoring algorithm on the full health=1 and dead=0 scale required for the calculation of QALYs.

Results: The results indicate that participants were able to readily choose “best” and “worst” attribute levels for the CHU9D health states. Large differences in value were found between the first and fifth levels (indicating “no problems” and “severe problems”) for all nine attributes relating to the CHU9D. In general, there was little differentiation between the middle levels of all attributes indicating only limited additional value for adolescents of moving between these levels. Comparison of the adolescent specific algorithm and the existing adult algorithm revealed some significant differences in values for identical health states which may have important implications for the application for the CHU9D to value adolescent treatment and service programmes particularly for mental health. In general, adolescents appeared to place more weight upon the CHU9D attributes relating to mental health (worried, sad and annoyed) than would be implied by application of the existing algorithm based upon adult values.

Conclusion: This study provides preliminary indications that there may be potentially important and systematic differences in the valuations attached to identical health states by adolescents in comparison to adult population groups. The study findings lend support to the potential future application of profile case best worst scaling discrete choice experiment methods to undertake large scale health state valuation studies directly with young adolescent population samples and provide support for the feasibility and acceptability of a web based mode of administration for this purpose.

Background

Adolescence is a significant period of transition, associated with major biological, psychological and social role changes [Williams 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. The attitudes that adolescents develop about their own health and health care have the potential to have a long term influence on their health outcomes. 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.

In many countries, adolescent health is marginalised and neglected [NHHRC, 2009]. Few countries routinely monitor the health status of adolescents in the general population to identify health trends and inequalities in this age group [Raven-Sieber et al, 2008]. There is also a great need for more information in relation to adolescents' attitudes about their own health status and their views about adolescent orientated treatment and service programs. Such information is an essential prerequisite for the planning and development of preventive strategies and clinical treatment programs designed to improve adolescent population health. A significant impediment to obtaining this information is the paucity of reliable and valid instruments available to measure self-reported health status in adolescents [Theunissen et al, 1998]. This contrasts with the situation for adults, where a range of instruments are available to assess self-reported health status [Bowling, 1997]. It appears to reflect a situation where young people are often not taken seriously enough to seek and respect their views and opinions in relation to their own health and parent proxy reports are often used in preference [Josefiak et al, 2008].

Researchers in health economics and other disciplines are increasingly recognising the importance of the measurement and valuation of health related quality of life (HRQoL) in both children and adolescents [Kleinet, 2007; Griebsch et al, 2005] for application in the economic evaluation of adolescent treatment and service programs. The overall aim of economic evaluation is to aid decision-makers to make efficient and equitable decisions by comparing the costs and benefits of health care treatment and service programmes [Brazier et al, 2007]. Economic evaluation requires the measurement and valuation of resources allocated across diverse diseases and treatment populations (including adolescents) with varying impacts upon health outcomes. Cost utility analysis (CUA) is the main approach used to measure and value the impacts of healthcare treatment and service programmes internationally and CUA is increasingly being applied in Australia and many other countries in the economic evaluation of new pharmaceuticals and medical services [Brazier et al, 2007; Harris et al, 2008]. Within CUA, survival and quality of life are combined into a single outcome measure, the quality adjusted life year (QALY). The QALY combines length of life with HRQoL on a single scale between zero and one, where zero is dead and one is full health [Brazier et al, 2007]. In order for an instrument to be usefully applied for the estimation of QALYs within CUA there is not only a need for it to be able to 'measure' health status but also a need for 'valuation' of the health states generated by the instrument i.e. it needs to be preference based. The PEDSQol [Varni, et al 2001] and the KIDSCREEN instruments [Ravens-Sieberer et al, 2008; Ravens-Sieberer et al, 2010] are amongst the most prominent instruments for the measurement of health related quality of life in adolescent populations. However, in their current form, neither instrument is suitable for the estimation of QALYs for application in CUA. The PEDSQoL and the KIDSCREEN provide summary scores across several attributes but they are not weighted by individuals' preferences for one dimension over another. The instruments have simple summative scoring algorithms which assume both equal intervals between response choices and that every dimension within the instrument is of equal importance. The instruments also reflect a broader interpretation of quality of life than that typically associated with the measurement and valuation of HRQoL for CUA including, for example, the child/adolescent's satisfaction with their relationship with parents, the atmosphere at home and the nature of their relationships with other children/adolescents.

There are now three preference-based generic instruments in the public domain that are suitable for the measurement and valuation of HRQoL in both children and adolescents. These are the AQoL-6D, CHU9D and HUI2. A further instrument is currently also in development – the EQ 5DY. The HUI2 is the most established instrument having been in existence for almost two decades [Torrance et al, 1996]. The HUI2 was developed in Canada by Torrance and colleagues, was originally designed for application in studies of childhood cancer, and is now widely used as a generic preference based measure for children and adolescents [Torrance et al, 1996]. The EQ-5DY (a child friendly version of the EQ-5D - a well known and widely applied generic preference based measure of adult health) has also been developed by a task force operating across seven countries on behalf of the EuroQol group [Wille et al, 2010]. Whilst this child friendly version of the instrument can presently be applied in the measurement of HRQoL, it cannot as yet be applied within the framework of economic evaluation as it does not have a preference based scoring algorithm attached to it. Both the EQ-5DY and the AQoL-6D for adolescents [Moodie et al, 2009] represent adaptations of existing instruments designed principally for application in adults. In contrast, the CHU9D utilised comprehensive qualitative interviewing with young people to identify the dimensions of HRQOL most relevant to this population [Stevens, 2009]. Where an instrument originates from an adaptation of an existing instrument, developed for use with adults, and not from an in-depth examination of the conceptualization, structure and content specific to children and adolescents' HRQoL, it may fail to include key aspects of HRQOL which are relevant to young people's views about their health and well-being [Stevens, 2010]. The EQ-5DY instrument developers have acknowledged the importance of this issue and suggest that further research is warranted to examine how the addition of further domains might improve the performance of the instrument [Wille et al 2010].

The CHU9D is a recently developed instrument that has been specifically designed for application in cost effectiveness analyses of treatment and service programmes targeted at young people [Stevens 2009; Stevens 2010]. The dimensions of HRQoL for inclusion in the 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 (see Table 1) 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 7-11 years) our previous pilot testing provides strong support for the feasibility of the application of the CHU9D with older children within the adolescent age group [Ratcliffe et al, 2010]. 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 compared with full health (and also death in the case of Standard Gamble) 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 (multinomial) 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 J and Buxton M, 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) DCE 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 choice scenarios 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 choice scenarios represented by *one health state only* and the respondent is asked to indicate the best and worst attribute of 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 2008; Marley et al 2008]. Profile case BWS has recently been successfully applied in a number of studies to estimate a cancer patient specific scoring algorithm for the EQ-5D and to develop scoring algorithm 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 (at least 15,625 health states). A fractional factorial that permitted the estimation of main effects, whilst maintaining the design properties of level balance and 'near' orthogonality (complete orthogonality in the design was not possible due to the need to eliminate a small number of implausible health states) was generated in 50 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 features (attribute levels) of each health state. An introductory example question from the pilot study is included below to illustrate the specifics of the method:

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'.

Best	Health State X	Worst
	I feel quite worried today	
O	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	O
	I can join in with a few activities today	

Basic socio-demographic information including age and gender and additional questions relating to health (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 employing the Family Affluence Scale (FAS) [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). FAS 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 FAS 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:6-7 FAS level) [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 based upon the standard gamble method developed by Stevens [Stevens, 2010A]. 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 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 (ε_{iq}). The equation to be estimated was of the following form:

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

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

Where β_{11} refers to the coefficient on the variable for attribute 1 level 1 (worried_not$_i$), β_{12} the coefficient on the variable for attribute 1 level 2 (worried_little$_i$) 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:

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

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

Since the conditional logistic regression estimates from the profile case BWS are anchored to the least valued attribute level, they must be re-anchored on a 0 -1 (dead to full health) scale via an external choice task for 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-anchor 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 2. 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 generated from application of the existing adult general population algorithm are also presented in Table 2. 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 (n=145) to 0.655 for those who reported themselves in poor health (n=2) 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 FAS 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 3. 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 (Table 2).

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 4. Re-anchoring the original conditional logit model ensured that the PITS or most severe health state (555555555) 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, 2010] 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 5 presents health state values for a selection of mild moderate and severe CHU9D health states based upon employment of the adolescent specific algorithm generated from this pilot study and the existing UK adult algorithms respectively. Differences in values are evident with the adult values for identical health states being relatively higher for the vast majority of health states. Table 6 presents predicted health state values grouped according to 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 health (worried, sad and annoyed), with application of the adolescent algorithm producing consistently lower mean values than the adult algorithm for CHU9D health 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 6 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, 2010]. It is also essential 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.

The findings from this pilot study also provide support for the practicality and face validity of the CHU9D for application with community based adolescents aged 11-17 years. A relatively high proportion of the total sample of participating respondents successfully completed the CHU9D. As expected, for 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 2010A, Ratcliffe et al 2010B].

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 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 it's 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 (4% of respondents indicated that they found the questionnaire very difficult to complete). 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 FAS (52% (n=306) classified in FAS 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.

Since the health state values from the profile case BWS exercise are anchored to the least valued attribute level, they must be re-anchored to the death state via an external choice task for the estimation of QALYs . Flynn and colleagues were the first researchers to identify the difficulties involved with valuation of the state dead within ordinal tasks such as ranking and discrete choice experiments [Flynn etal, 2008A]. Craig et al have since described the state dead as “the inextricable anchor of the quality-adjusted life years scale” [Craig et al, 2009]. The findings from our previous pilot work indicate that TTO and SG tasks are unreliable in young adolescent populations [Ratcliffe et al, 2010]; therefore re-anchoring needs to be based upon TTO or SG values taken from an adult general population sample (as is the case here) or potentially an older adolescent sample could be utilised for this purpose. Recently studies have been conducted by researchers in Australia and Canada to utilise 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]. These studies have been 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 with an adolescent sample for two main reasons. Firstly, the CHU9D descriptive system itself is a relatively large descriptive system and it would necessitate the 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 as to observe trade offs.

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.

References

Al Janabi H, Flynn T, Coast J. (2010). Estimation of a Preference-Based Carer Experience Scale. *Medical Decision Making* (in press).

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.

Bowling A. *Measuring Disease. A review of disease specific quality of life measurement scales*. 2nd Edition, Open University Press, 2001.

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 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, 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, 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* 2010; 70: 1957-1965.

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.

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.

Harris A, Hill S, Chin G, Li J, Walkolm E. The role of value for money in public insurance coverage decisions for drugs in Australia: a retrospective analysis 1994-2004. *Medical Decision Making* 2008; 28: 713-722.

Hensher D, Rose J and Greene W. *Applied Choice Analysis, A Primer*. 2005; Cambridge University Press.

Josefiak T, Larsson B, Wichstrom L, Mattejat F, Ravens-Sieberer UR. Quality of life as reported by school children and their parents: a cross-sectional survey. *Health and Quality of Life Outcomes* 2008; 6:34.

Kleinet 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. *Centre for Health Economics, Monash University Research Paper* 2009 (43).

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, DCE 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 (2010) 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* (in press).

Ratcliffe J, Stevens K, Brazier J, Flynn T, Sawyer M (2010A). An assessment of the construct validity of the CHU9D in the Australian adolescent general population. Submitted for publication.

Ratcliffe J, Stevens K (2010B). 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, Wille N, Bullingher M and the BELLA study group. Health related quality of life in children and adolescents in Germany: results from the BELLA study. Eur Child Adolesc Psychiatry 2008; [Suppl 1] 17: 148-156.

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. 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. 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. 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.

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

Theunissen NC, Vogels TG, Koopman HM, et al. The proxy problem: child report versus parent report in health-related quality of life research. *Quality of Life Research*. 1998; 7: 387-397.

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

Torrance G, Feeny D, Furlong W, Barr R, Zhang Y, Wang Q. Multiattribute utility function for a comprehensive health status classification system. Health Utilities Index Mark 2. *Medical Care* 1996; 34: 702-22.

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

Varni J, Seid M, Kurtin P. PedsQL 4.0: reliability and validity of the Pediatric Quality of Life Inventory version 4.0 generic core scales in healthy and patient populations. *Medical Care* 2001; 39(80: 800-812.

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

Wille N, Badia X, Bonsel G et al. Development of the EQ-5DY: a child friendly version of the EQ-5D. *Quality of Life Research* 2010; 19: 875-886.

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

Table 1: CHU9D Dimensions and levels
1. *Worried*
I don't feel worried today
I feel a little bit worried today
I feel a bit worried today
I feel quite worried today
I feel very worried today
2. *Sad*
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
3. *Pain*
I don't have any pain today
I have a little bit of pain today
I have a bit of pain today
I have quite a lot of pain today
I have a lot of pain today
4. *Tired*
I don't feel tired today
I feel a little bit tired today
I feel a bit tired today
I feel quite tired today
I feel very tired today
5. *Annoyed*
I don't feel annoyed today
I feel a little bit annoyed today
I feel a bit annoyed today
I feel quite annoyed today
I feel very annoyed today
6. *School Work/Homework*
I have no problems with my schoolwork/homework today
I have a few problems with my schoolwork/homework today
I have some problems with my schoolwork/homework today
I have many problems with my schoolwork/homework today
I can't do my schoolwork/homework today
7. *Sleep*
Last night, I had no problems sleeping
Last night, I had a few problems sleeping
Last night, I had some problems sleeping
Last night, I had many problems sleeping
Last night, I couldn't sleep at all
8. *Daily routine*
I have no problems with my daily routine today
I have a few problems with my daily routine today
I have some problems with my daily routine today
I have many problems with my daily routine today
I can't do my daily routine today
9. *Able to join in activities*
I can join in with any activities today
I can join in with most activities today
I can join in with some activities today
I can join in with a few activities today
I can join in with no activities today

Table 2: CHU9D utilities according to the characteristics of respondents

Characteristic	N=590	CHU9D utilities[1] Mean (SD)
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 FAS (%)	57 (9.7)	0.806 (0.134)
Medium FAS (%)	227 (38.5)	0.844 (0.118)
High FAS (%)	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)

[1] from application of the existing adult scoring algorithm.
** indicates statistically significant difference between mean values at the 1% level.

Table 3: Responses to the CHU9D

CHU9D Dimensions and levels	*Frequency (%)*
Worried	
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%)
Sad	
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%)
Pain	
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%)
Tired	
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%)
Annoyed	
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%)
Schoolwork/homework	
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%)
Sleep	
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%)
Daily routine	
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%)
Activities	
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 4: Re-scaled conditional logit estimates*

Worried	Level 1	0.1292
	Level 2	0.0617
	Level 3	0.0602
	Level 4	0.0484
	Level 5	0.0425
Sad	Level 1	0.1148
	Level 2	0.0535
	Level 3	0.0485
	Level 4	0.0412
	Level 5	0.0269
Annoyed	Level 1	0.0895
	Level 2	0.0452
	Level 3	0.0438
	Level 4	0.0205
	Level 5	0.0169
Tired	Level 1	0.0967
	Level 2	0.0596
	Level 3	0.0542
	Level 4	0.0432
	Level 5	0.0381
Pain	Level 1	0.0722
	Level 2	0.0356
	Level 3	0.0335
	Level 4	0.0255
	Level 5	0.0323
Sleep	Level 1	0.1145
	Level 2	0.0449
	Level 3	0.0478
	Level 4	0.0255
	Level 5	0.0344
Daily routine	Level 1	0.1136
	Level 2	0.0529
	Level 3	0.0495
	Level 4	0.0313
	Level 5	0.0246
Schoolwork	Level 1	0.1163
	Level 2	0.0562

	Level 3	0.0551
	Level 4	0.0460
	Level 5	0.0469
Activities	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 111111111, is equal to 1.00 and the lowest (PITS) CHU9D health state, state 555555555, is equal to 0.33, the corresponding PITS value from the adult SG algorithm.

Figure 1: Graph of BWS CHU9D utilities by attribute level

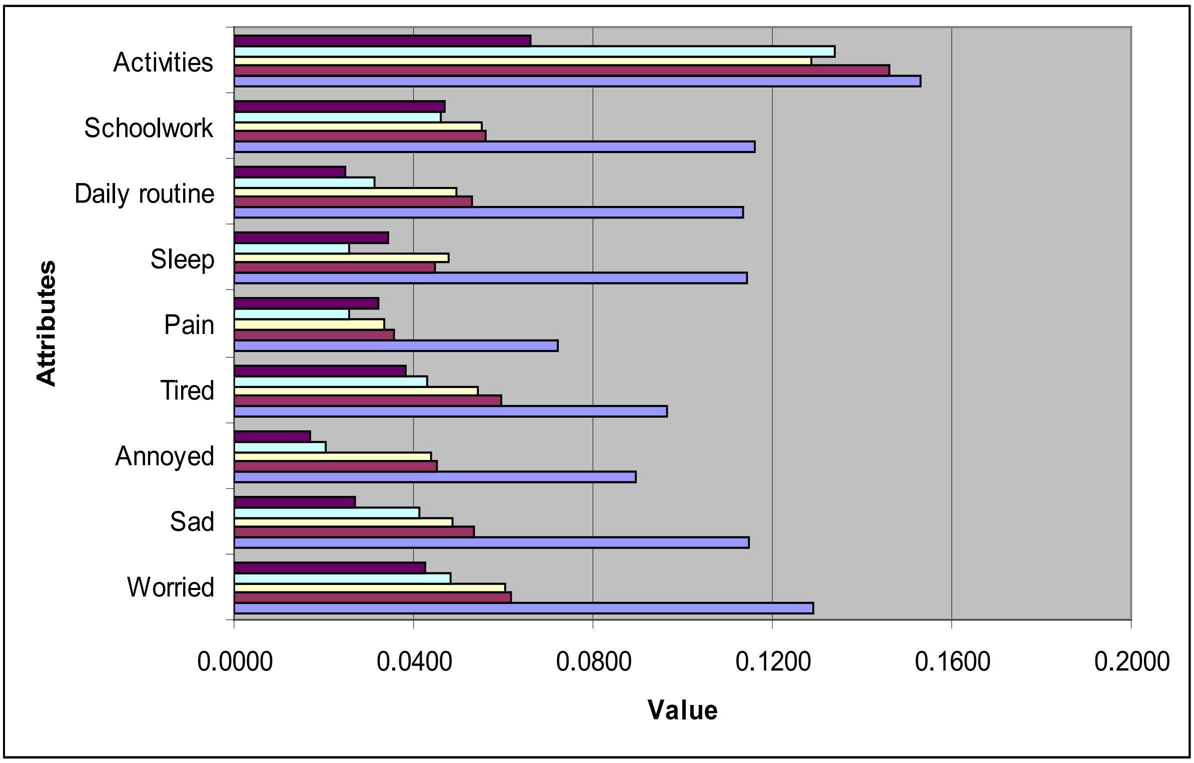

Table 5: Predicted health state values adult vs adolescent algorithm

CHU9D Health state	Adult value	Adolescent value	Difference
512433221	0.758	0.589	0.169
534112352	0.717	0.568	0.149
521313422	0.711	0.593	0.117
521234143	0.701	0.593	0.109
542321331	0.683	0.591	0.092
533221114	0.784	0.708	0.075
513142233	0.678	0.605	0.073
555555515	0.392	0.398	-0.006
555555555	0.326	0.329	-0.003
452123133	0.734	0.598	0.136
432511223	0.731	0.605	0.126
415332152	0.741	0.619	0.122
411223315	0.706	0.628	0.078
423215341	0.639	0.560	0.078
423131522	0.692	0.617	0.075
444444414	0.544	0.486	0.058
431352231	0.635	0.579	0.056
334522121	0.750	0.571	0.180
325134211	0.783	0.609	0.175
352132411	0.749	0.608	0.141
322413135	0.678	0.557	0.121
333333313	0.701	0.582	0.119
313514232	0.725	0.609	0.116
345213132	0.719	0.613	0.106
314321245	0.654	0.565	0.089
313212534	0.699	0.609	0.089
321421553	0.659	0.597	0.062
312351442	0.583	0.645	-0.062
341125322	0.634	0.610	0.025
351241253	0.593	0.605	-0.012
321152324	0.625	0.617	0.008
332245121	0.590	0.596	-0.007
253412341	0.678	0.541	0.137

214233521	0.702	0.572	0.130
251313224	0.713	0.595	0.118
222222212	0.728	0.616	0.112
231324531	0.659	0.548	0.112
235211423	0.698	0.590	0.109
232114355	0.616	0.512	0.104
CHU9D Health state	**Adult value**	**Adolescent value**	**Difference**
241532143	0.700	0.597	0.103
243131225	0.647	0.567	0.080
213125433	0.667	0.602	0.065
233451112	0.674	0.720	-0.046
224153133	0.655	0.610	0.045
225341331	0.622	0.584	0.038
212335154	0.672	0.638	0.034
211543312	0.686	0.689	-0.004
135123244	0.730	0.608	0.123
123523451	0.689	0.579	0.110
115422323	0.720	0.619	0.101
131232435	0.627	0.558	0.069
112254323	0.581	0.641	-0.060
124315213	0.714	0.662	0.052
153324122	0.679	0.631	0.048
122531334	0.699	0.653	0.046
142312513	0.698	0.657	0.041
154231332	0.669	0.635	0.033
111111112	0.963	0.993	-0.030
131435232	0.661	0.633	0.028
132143542	0.583	0.610	-0.027
143253251	0.582	0.607	-0.025
123342125	0.577	0.587	-0.011
111121111	0.967	0.963	0.003

Table 6: Predicted health state values adult vs adolescent algorithm grouped by attributes

CHU9D Health state	Adult value	Adolescent value	Difference	Attributes at lower levels
551151111	0.874	0.753	0.121	Worried, sad, annoyed
441141111	0.874	0.777	0.097	
331141111	0.902	0.819	0.083	Mental health attributes
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	
113311311	0.898	0.852	0.046	Physical health attributes

Table 7: Level of difficulty indicated for survey

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