FLINDERS UNIVERSITY
ADELAIDE AUSTRALIA

FLINDERS
CLINICAL CHANGE

Filling critical gaps in our knowledge on chronic disease management, and later and end-of-life health care

Working Paper

No. 2009/10

Valuing Child Health Utility 9D health states in a young adolescent sample

By Julie Ratcliffe, Leah Couzner, Terry Flynn, Michael Sawyer, Katherine Stevens, John Brazier

Flinders UNIVERSITY
ADELAIDE • AUSTRALIA

Valuing Child Health Utility 9D health states in a young adolescent sample: A pilot study to assess the feasibility of best worst discrete choice, standard gamble and time trade off methods

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

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

Corresponding author:

Julie Ratcliffe

Flinders Centre for Clinical Change & Health Care Research

Flinders University

Bedford Park

South Australia

Tel +61 8 8275 1297

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

Acknowledgements

We are particularly grateful to Darren McLachlan, Westminster School for his support and help with study administration. We would also like to thank the staff, parents and children from Westminster School, Adelaide who consented to and participated in this study. Brita Pekarsky provided helpful comments on a previous version of this paper. The usual disclaimed applies.

1. Background

Adolescents and health state preferences

Adolescence is a period of life when treatment adherence/compliance is often poor and professional treatment programs are underutilised [Shaw 2001; Kovcs et al 1992; Dolgin et al 1986]. This is a significant problem as non-compliance and poor utilisation of health interventions can lead to increased morbidity and mortality which is associated with higher health care costs [Berg et al 1993]. Numerous studies have stressed the importance of involving and engaging adolescents in decisions concerning their health status and health care if we are to improve treatment adherence/compliance and service utilisation in this age group [Milne 2000; Tylee et al 2007; Kleinet 2007] Furthermore, since the publication of the United Nations Convention on the Rights of the Child [United Nations, 1989] policy makers have emphasised the importance of involving children and adolescents in health care decision-making and of recognising adolescents as consumers of health care in their own right [Department of Health 2003; Cavet and Sloper 2004]. The measurement and valuation of child and adolescent health has recently been identified as a key area for future research by the Organisation for Economic Co-operation and Development [OECD, 2006].

Medical advances have meant that there are now higher rates of survival from historically fatal childhood conditions often with the persistence of chronic health conditions. When measuring and valuing health outcomes in childhood and adolescence it is therefore important to focus upon the health related quality of life associated with survival and not just the length of survival per se. Quality adjusted life years (QALY's) are increasingly been utilised as a health outcome measure to calculate the benefits of new treatments and interventions within cost utility analyses (CUA), a technique of economic evaluation [Drummond et al 2005; Brazier et al 2007]. Despite this overall trend, published CUA studies of adolescent specific treatment programs are scant in comparison with those reported upon for adults [Ungar and Santos, 2004]. A recent critical review highlighted the methodological shortcomings of the majority of studies undertaken in this area particularly in the methods used to measure and value health outcomes [Griebsch et al 2005].

Where QALY's have been estimated within the framework of CUA for adolescent specific treatment programs, studies have tended to incorporate the views of clinicians or adults as the source of preferences rather than adolescents. Hence the values of adults have been used to assess the perceived benefits of new treatments and interventions for adolescents. However it is not clear that the views of adults are in accordance with those of adolescents on this issue. There is a large

body of evidence to suggest that the views of proxy adult respondents and adolescents in relation to the 'measurement' of health related quality of life differ significantly [Josefiak et al 2008; Varni et al 2007; Le Coq et al 2000; Theunissen et al 1998]. Although evidence relating to the 'valuation' of health related quality of life is more limited in comparison, it does suggests that the views of adults and adolescents about the value of identical health states may differ significantly [Norquist et al 2008, Wasserman et al 2005, Saigal et al 1999].

Saigal et al [1999] assessed differences in preferences for neonatal outcomes from the perspective of health care professionals, parents and adolescents. The standard gamble (SG) technique was used to elicit preferences for five health states based upon the responses of extremely low birth weight children at 8 years of age to the Health Utilities Mark 2 (HUI2) instrument. A total of 742 participants were recruited and interviewed over a three year time period including 100 neonatologists, 103 neonatal nurses, 264 adolescents (aged 12-16 years) and 275 parents of the recruited adolescents. The reliability of the standard gamble was not formally assessed within this study. The findings indicated that the utilities attached to identical neonatal health impairment states varied by as much as 0.10 for parents and adolescents (mean health state value 0.82 for adults and 0.72 for adolescents) and these differences were statistically significant. This implies that the incremental QALY gain from restoring low birth weight babies to full health would be more than 1.5 times as large using the adolescent value compared with the adult value (incremental QALY gain 0.18 for adults and 0.28 for adolescents). Similarly in a study to measure health state preferences for hemophilia, Wasserman et al [2005] applied the SG technique to compare the values attached to identical hemophilia disease specific health states by adults and adolescents aged 14 to 18 years. Statistically significant differences were found between the adolescent and adult participants for identical health states with the SG instrument ($P = 0.045$) with older participants being prepared to take more risk overall. In a recent study of antidepressants used to treat adolescent depression the authors expressed concern that only health state values from adults were available to estimate QALYs for the depressed adolescents and these may represent an unreliable guide to the true benefits from treatment since adults tend to underestimate the impact of this condition (and the consequent treatment impact) upon the adolescent's health and quality of life [Norquist, 2008]. These findings highlight the possibility that CUA studies of adolescent treatments and interventions which utilise health state values obtained from adults may reach incorrect conclusions about the adolescents' perceptions of the relative health benefits of alternative new treatment programs or health services for adolescents.

Specific evidence relating to the reliability of the application of SG and TTO techniques in adolescent samples is virtually non existent. The reasons for this are unclear but may partly reflect a predominant view amongst researchers that these conventional approaches are unsuitable for administration in child and adolescent samples due to the complexity of the task and/or ethical difficulties in their administration. In addition, many decision-making bodies including the National Institute for Clinical Excellence in the UK recommend that health state values be obtained from a 'representative' sample of the general population [NICE, 2008]. In practice this has been widely interpreted to imply 'adult only' general population preferences although in principle a truly representative sample of the population would include adolescents and children. In a study which was undertaken some time ago but which remains widely quoted, Juniper et al [1997] applied SG to estimate the value that 52 children (aged 7-11 years) with symptomatic asthma placed on their own health state. The study authors concluded that the standard gamble technique was unreliable in this population and that children needed at least grade 6 level reading skills (11 -12 years) to complete it.

Although there are a number of instruments available for measuring the health related quality of life of children and young adolescents including the PEDSQol [Varni et al 2001] and the KIDSCREEN [Ravens-Sieberer et al 2008], they are not suitable for use in the estimation of QALYs within economic evaluation. These instruments provide summary scores across several dimensions attributes but they are not preference based. In other words, the instruments have simple summative scoring algorithms which assume equal intervals between response choices and dimensions are of equal importance. A one unit improvement in emotional heath for example is assumed equivalent to a one unit improvement in pain. The PEDSQol and the KIDSCREEN are also not amenable to valuation for the estimation of QALYs because the size of the descriptive systems and the content, formatting and wording of the included dimensions precludes valuation [Stevens 2008]. For example, the PEDSQoL descriptive system has 23 items in total and 5 levels for each item. In addition to dimensions relating to health including problems with health and activities, this instrument contains dimensions which are not normally associated with the measurement and valuation of health for economic evaluation i.e. how the respondent feels that they get along with other children/teenagers. The KIDSCREEN is available in 3 versions of differing lengths (52 items, 27 items and 10 items). Whilst the 10 item version is short enough to potentially allow it to be amenable to valuation, this instrument also contains dimensions which are not normally associated with the measurement and valuation of health for economic evaluation including the

child/adolescent's satisfaction with their relationship with parents, the atmosphere at home and the nature of their relationships with other children/adolescents.

Measuring and valuing adolescent health

Presently, the most popular mechanism for obtaining health state values to calculate QALYs, particularly within the context of a clinical trial, is to employ generic preference based measures of adult health such as the AQoL (Assessment of Quality of Life), the EQ-5D (EuroQol) or the SF-6D [Brazier et al, 2007]. Generic preference based instruments comprise two main common elements: a descriptive system for completion by patients (or members of the general population) comprising a set of items with multiple response categories covering the different dimensions of health related quality of life and an off the shelf scoring algorithm reflecting general population values for the health states described by the instrument. To obtain the scoring, the health state values or utility weights for a sample of health states described by the instrument are typically derived from an adult general population sample using conventional valuation methods such as SG or TTO. Statistical modelling methods or multi attribute utility theory models are then typically employed to generate a scoring algorithm which allows the estimation of utility weights for all possible health states defined by the instrument [Brazier et al, 2007]. The advantages of the application of generic preference based instruments include their perceived ability to be used across all conditions, little additional respondent burden allowing their relative ease of incorporation into most clinical trials and routine data collection systems, and ease of valuation of responses using off the shelf scoring algorithms provided by the instrument developers. However, although there are a range of different instruments available, they all have different properties in terms of their coverage, sensitivity, responsiveness and the methods used to value the health states and there is no 'gold standard' or general consensus as to the appropriate instrument to use. In addition, the AQoL, EQ-5D and SF-6D have all been scored using adults as the source of preferences, have been designed for use in adults and have been widely applied in this capacity.[1]*

To be applicable for adolescent populations an instrument needs to be sensitive to changes in development and changes in health, and be sufficiently flexible to accommodate differing cognitive abilities of adolescents at various ages with regard to reporting and valuing health states [Petrou 2003; Stevens 2008]. Two generic preference based instruments have been developed which are

[1]* Work is currently underway to adapt the AQoL for application in adolescents and to develop a child friendly version of the EQ-5D: the EQ-5DY. However both of these projects are at a developmental stage and information has not yet been made available in the public domain.

suitable for application in young adolescent samples. The first instrument is the Health Utilities Index 2 (HUI2), which was originally designed to be used specifically in oncology studies and since has been adapted to become a generic instrument [Torrance et al 1996]. This measure takes a 'within skin' approach to the concept of health in that the dimensions are very impairment based e.g. levels of hearing and vision rather than estimating the impact of the impairment upon health related quality of life. The scoring algorithm for the HUI2 was developed with adults using conventional scaling methods. The second instrument is the Child Health Utility 9D (CHU9D) which was developed at the University of Sheffield, UK by Stevens [Stevens, 2008; Stevens 2008a, Stevens 2008b]. The CHU9D is a generic preference based measure of health related quality of life developed specifically for application in young people. The instrument has 9 dimensions (worried, sad, pain, tired, annoyed, schoolwork, sleep, daily routine, ability to join in activities) with 5 levels within each dimension. A unique feature of this instrument is that it was developed exclusively with young people. The response scales, wording and formatting are based upon a qualitative study of interviews with over 70 young people with a wide range of health problems. The instrument has undergone psychometric testing in both general primary school and in clinical paediatric populations and has demonstrated good practicality and validity [Stevens 2008]. Although originally developed for use with 7-11 year olds, the item content [see Appendix 1] is also very appropriate for young adolescents. For this reason and because the CHU9D dimensions focus upon health related quality of life rather than symptoms we have selected it for use in this study in preference to the HUI2. Presently there exists a single scoring algorithm for the CHU9D. This original scoring algorithm is based upon UK adult general population values (n=300) using the SG technique. However, Stevens [2008] has noted that it cannot be assumed that the health state values obtained from adults will accurately reflect those of adolescents and further research is needed to obtain CHU9D health state valuations directly from young people.

Conventional valuation methods

Although there is no accepted gold standard scaling method for eliciting health state values for the estimation of QALYs, historically the majority of health economists have tended to favour the choice based valuation methods of SG and TTO [Brazier et al, 2007]. Both methods have traditionally been applied in adult populations. SG has the most rigorous foundation in theory in the form of expected utility theory of decision making under uncertainty. However, theoretical arguments have been raised against the use of SG in health state valuation [Richardson 1994] and it has been argued that SG valuations may be influenced by factors other than a person's attitude to the health state, including probability weighting and loss aversion [Bleichrodt, 2002]. In addition

SG questions typically involve consideration of a probability of immediate death which may be considered ethically inappropriate and upsetting, particularly for young adolescents to contemplate. The TTO was developed specifically for use in health care as a less complex alternative to SG that overcomes the problems of explaining probabilities to respondents. However, TTO questions typically also involve consideration of death. There are concerns about the theoretical and empirical basis of the TTO. There is evidence to suggest that duration effects and time preference effects can have an impact on the elicitation of TTO values [Sutherland et al 1982; Dolan and Gudex 1995]. Therefore in reality SG and TTO valuations can both be seen as containing biases as measures of the value of health. 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 successively until the point of indifference is found.

Recently, there has been an increasing interest in using discrete choice experiment (DCE) methodology to estimate cardinal values for health states to calculate QALYs on the basis that these biases and problems may be avoided by using this technique [Brazier et al 2007]. DCE methods may be more easily understood and interpretable by vulnerable populations e.g. elderly and young people thereby allowing health state valuation tasks to be carried out for the first time directly using young adolescent population samples.

Discrete choice experiments (DCE)

DCEs have their theoretical basis in random utility theory [McFadden 1974, Thurstone 1927]. Random utility theory is based upon a psychological model that describes the probability of ranking a particular health state higher than another given the (unobserved) cardinal utility associated with each health state. It is usually operationalised in conjunction with a valuation function that relates the mean utility for a given health state to a set of explanatory variables [McFadden 1973; Lancaster 1966]. This operationalisation of DCEs is usually conducted within a conditional (multinomial) logistic regression model and its generalisations [McFadden 1973]. Although DCEs have become a very popular tool for eliciting preferences in health care, the vast majority of published studies using DCE methodology have tended to focus upon the possibility that individuals derive benefit from non-health outcomes and process attributes in addition to health outcomes [Ryan and Gerard 2003]. A limited number of studies have used DCEs to estimate values for different health or quality of life states [Ratcliffe et al 2009; Coast et al 2008; Ryan et al 2005; Hakim and Pathak 1999]. With the exception of the study by Coast et al, all of these studies have also employed a traditional DCE approach which involves choosing the most preferred scenario from a choice set of

two or more competing scenarios, repeated over a number of choice sets so as to observe trade-offs [Louviere et al 2000].

The profile version of a Best-Worst Scaling DCE, devised by Finn and Louviere [1992] and first introduced to European health care research by McIntosh and Louviere [2002] allows the impact of attributes to be compared meaningfully on a common scale. This approach also involves a potentially easier choice task than traditional DCE's because respondents are asked to indicate the best and worst attribute for each of a number of scenarios or health states, presented one at a time [Marley et .al 2008; Flynn et al 2008; Flynn et al 2007]. The promotion of respondent understanding of the choice task is an issue of crucial importance for the elicitation of health state values from vulnerable population groups e.g. young adolescents and the elderly. A Best-Worst Scaling DCE has recently been successfully applied by Flynn and colleagues in the valuation of the ICECAP capability index, a new instrument for measuring and valuing quality of life designed for use in the economic evaluation of health and social care interventions for the elderly [Coast et al 2008].

This pilot study was designed as the precursor to a large scale health state valuation study to develop a scoring algorithm for the CHU9D based upon the preferences of young adolescents. The pilot study sought to test the following:
[1] The feasibility of applying Best-Worst Scaling DCE questions to a series of health states defined by the CHU9D in a young adolescent sample.
[2] To compare Best-Worst Scaling DCE questions to conventional TTO and SG questions in terms of ease of understanding and completeness of the valuation task.

2. Methods

Participants

Permission was sought and ethical approval was granted to conduct the pilot study in Westminster School, an independent school in the southern suburbs of metropolitan Adelaide. Parents and pupils within the young adolescent target age range of 11–13 years (n=45) were initially invited to give informed consent for the pupil to participate in the pilot study.

Measures

Participants were asked to self-complete the CHU9D instrument as the first phase of the interview prior to undertaking the DCE task. This helped familiarise them with the wording and formatting of

the CHU9D health states for valuation. Following completion of the CHU9D the remainder of the interview was tape recorded to enable a detailed investigation of participants' responses and explanations within each task. The second phase of the interview comprised the Best-Worst Scaling DCE task whereby participants were presented with a series of 5 CHU9D health states, asked to indicate the best and worst features for each health state and indicate the reasons for their responses. The 5 CHU9D health states were chosen to include a mix of a broad range of dimension levels (mild, moderate and severe) with the exception of health state card 4 which reflected an intermediate level for all dimensions making it potentially more difficult for participants to choose the best and worst features (Table 4). At the final phase of the interview participants were randomised to receive additional SG questions (administered by interviewer 1: LC) or TTO questions (administered by interviewer 2: JR). Each participant was asked to value two health states, a health state labelled 'practice health state' and a second health state which was chosen sequentially from the five health state cards presented from the previous Best-Worst Scaling DCE task.

For the SG task the 'ping-pong' method was used with a visual aid in the form of a probability wheel [Torrance, 1986]. The ping-pong method helps respondents to determine their point of indifference between the certain outcome and the gamble by iterating between values for the probability of success P. The probability wheel is an adjustable disk with two sectors, each a different colour. It is constructed so that the relative size of the two sectors can be readily changed. The two alternatives are presented to the individual on cards and the two outcomes of the gamble alternative are colour coded to match the two sectors of the probability wheel. The individual was informed that the probability of each outcome was proportional to the similarly coloured area of the disk.

Conventionally the SG task (for a health state considered better than being dead) has a full health versus immediate death comparison for the gamble alternative. However, following concerns raised by the ethics committee, this comparison was modified to reflect a full health versus 'PITS' (reflecting the lowest level on all of the 9 dimensions of the CHU9D) state comparison for the young adolescent sample to avoid any references to immediate death which may have been upsetting or difficult for young adolescents to contemplate.

The TTO task was based upon the widely used variant developed by the York EQ-5D team for interviewer administration, using props in the form of a sliding scale to represent life years [Dolan

et al, 1996]. Conventionally this task also contains references to death in the question wording and hence for ethical reasons (as above) the interview script was modified to remove this reference. Participants were asked to compare living in two health states for a maximum period of 10 years. However the reference to dying immediately after each specified period of time was removed and participants were informed that what happens to their health beyond the specified time periods was not known and should not be taken into account when thinking about and answering the questions. As the final phase of the interview, each participant was asked to indicate how difficult they found each task, which task they preferred and the reason/s for their preference.

3. Results

A total of 16 positive double consent responses from both a parent and pupil were achieved giving an overall response rate of 36%. The age and gender of the participants is summarised in Table 1. Although specific data on the education level of the participants in this sample were not available, year 7 (aged 13 years) exam results indicate that Westminster School performs above the South Australian state averages in all areas of learning including reading, writing, spelling, grammar and punctuation and numeracy. The interviews took between 15 and 30 minutes to complete and all participants completed all three phases of the interview.

The responses to the CHU9D are presented in Table 2 and the health state values obtained by applying the existing UK adult general population SG algorithm to the responses to the CHU9D instrument are presented in Table 3. Participants tended to report themselves in good health overall, although no participants reported themselves at the best level for all of the CHU9D dimension questions. With the exception of one participant who reported themselves as feeling very tired today, the worst levels for the CHU9D instrument (levels 4 and 5) didn't apply for this sample. The CHU9D health state values ranged from 0.77-0.92 with a mean of 0.85.

Table 4 presents the responses to the best worst scaling DCE task. For health state card 1 the best feature for 8 of the participants was 'I don't feel sad today' which represented the best level for this dimension. Somewhat surprisingly 5 respondents chose 'I can join in with a few activities today' as the best feature. This represents a relatively low level (level 4) in the activities dimension. When prompted to explain why they had chosen this feature as the best these participants appeared to focus upon the positive question framing associated with this dimension (in contrast to the negative question framing of the other dimensions) and hence interpreted being able to join in with a few activities as essentially a 'good thing'. For health state card 2, the best feature for 8 participants was

"I don't have any pain today' and for a further 5 participants was 'I don't feel annoyed today' representing the best level for the pain and annoyed dimensions respectively. The activities feature 'I can join in with some activities today' (level 3) was chosen as the best feature by 3 participants and also as the worst feature by a further 4 participants. The explanations for these different responses indicated similarities with the responses to health state card 1 in that those participants who had chosen this feature as the best appeared to focus more upon the positive question framing associated with this dimension. Those participants who indicated this feature as the worst focused more upon the limitations of only being able to perform some activities when compared with the best level of being able to join in any activities. Health state card 3 represented the highest degree of concordance as to the best feature with 14 participants choosing 'I don't feel annoyed today' (level 1) as the best feature. There was less agreement relating to the worst feature although all of the worst features chosen represented low dimensions levels (either 4 or 5). Health state card 4 reflected an intermediate level for all attributes making it potentially more difficult for participants to choose and the majority of participants did express some difficulties in choosing the best and worst features for this particular health state. The activities dimension 'I can join in with some activities today' (level 3) was chosen as the best feature by 12 participants principally for the same reason as previously highlighted. There was less agreement in relation to the worst feature for this health state. For health state card 5, 13 participants chose 'I don't have any pain today' as the best dimensions and 3 participants indicated 'I can join in with a few activities today' (level 4) as the best feature again principally due to their interpretation of the positive nature of this statement.

The participant's responses to the SG and TTO tasks are summarised in Table 5. For the SG only one respondent (no.2) appeared to fully understand the task and was able to clearly identify a point of indifference between the gamble and the certain outcome. The remaining participants exhibited difficulties with the SG task in [1] identifying a point of indifference or [2] not appearing to fully consider the certain outcome in making their choices or [3] a combination of both of these factors. Those participants who did not appear to consider the certain outcome but who could identify a point of indifference tended to focus almost exclusively upon the probability of the treatment working in the gamble alternative and hence their values for the two health states considered were very similar or identical. For the TTO, again only one participant (no.11) appeared to understand the task and was able to identify a point of indifference for both of the health states presented. The remaining participants who received TTO questions all exhibited a very strong reluctance to trade healthy life years. These participants tended to view 10 years in the intermediate health state as a very long time (often appearing to view this time period as a life-time since they had been alive for

a similar time period) and expressed concerns at the consistency of being in an intermediate health state with problems on many dimensions for such a long period of time. Some participants expressed the absolute necessity to them of remaining healthy during the next ten years of their life because of their desire to achieve certain life goals and viewed the transition from adolescence into adulthood as being a crucial period in terms of the achievement of these.

When asked to indicate which valuation technique they preferred 5 of the 8 participants who received SG indicated a preference for that over the Best-Worst Scaling DCE questions principally because they liked using the probability wheel as a visual aid. In contrast only 1 of the 8 participants who received TTO indicated a preference for that technique.

4. Discussion

Overall, the results of this pilot study lend support to the view that Best- Worst Scaling DCE methods are more easily understood by young adolescent samples when compared with conventional approaches to health state valuation. We found that young adolescents were readily able to choose the best and worst features of a series of CHU9D health states and provide justifications for their answers. With the exception of health state 4 (which was deliberately chosen to reflect an intermediate level for all attributes making it potentially more difficult for participants to make choices), in general the dimension levels that were chosen as the best and worst for each health state corresponded with the highest and lowest levels for those dimensions. The exception to this rule was the activities dimension which presented problems of question framing and interpretation for some participants. The difficulties experienced in this pilot study with regard to the interpretation of this dimension could potentially be alleviated by careful explanation of this dimension to ensure that respondents remain fully aware throughout the valuation exercise of the range of alternative levels and are reminded of the limitations of e.g. being able to join in with only a few activities.

In contrast to the Best-Worst Scaling DCE questions, although the participants who were randomised to receive SG tended to view the technique positively, only 1 participant appeared to fully understand and engage with the choice task. Despite being prompted to tell the interviewer when the two choices were viewed as equal, participants expressed difficulties in identifying a point of indifference. Participants also tended to overlook the certain outcome when making choices and focused predominantly upon the probabilities within the gamble alternative. Participants who

received the TTO exhibited a very strong reluctance to trade healthy life years and were unable to identify a point of indifference.

The administration of SG and TTO in young adolescent samples presents additional problems due to the ethical difficulties associated with presentation of the concept of immediate death. This was most evident in applying the TTO technique. If participants had been informed (as is conventionally the case) that they would die immediately following the specified period of time in full health they may have been more likely to trade healthy life years. Further research is required to examine the tolerability and acceptance of the concept of immediate death in health state valuation exercises with adolescent samples. It may be the case that older adolescents will exhibit a stronger ability to deal with this concept, although this does not remove the possibility that ethics committees will continue to raise concerns about its utilization in this context with samples of young people.

For the estimation of QALYs, health state values are required on a scale where 1 is equivalent to full health and 0 is equivalent to death. In common with all limited dependent variable models (including traditional DCE's and ranking studies), the Best-Worst Scaling DCE model produces predicted valuations on an interval scale such that meaningful comparisons of differences are possible but the origins and units of the scale are defined arbitrarily by the identifying assumptions in the model [Salomon, 2003]. Since the health state values from the Best-Worst Scaling DCE exercise will be 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 [Ratcliffe et al 2009; Flynn et al, 2008]. A number of methods have been proposed in the literature for achieving this including re-anchoring by rescaling via the value for the most severe or PITS health state generated from TTO or SG (since at least conceptually, these methods estimate the trade off that respondents are willing to make between quantity and quality of life) to ensure that the zero represents death, rather than the utility of the most severe health state [Salomon 2003; McCabe et al 2006]. The findings of this pilot study suggests that TTO and SG tasks may be unreliable in young adolescent populations and therefore re-anchoring would need to be based upon TTO or SG values taken from an older adolescent or adult general population sample. Whilst the re-anchoring process could be viewed as a weakness of this approach in relation to its' ability to produce health state values which can be used for the estimation of QALY's, it is important to note that the avoidance of the concept of immediate death in young adolescents samples would also necessitate a re-anchoring or chaining process for SG and TTO methods. Hence even if the modified SG and TTO questions had been found to generate reliable health state values in this study, these would still have needed to be

chained to death via the worst or PITS health state using health state values from an older adolescent or adult general population sample. The only other way to avoid rescaling DCE generated health state values via an external choice task is to conduct a very large and resource-intensive DCE study that includes a length of life dimension. This would necessitate the presentation for valuation of many health states since it would involve interacting a relatively large number of length of life levels with (at a minimum) the fractional factorial that permits the estimation of main effects, (also known as an orthogonal main effects plan which can be generated in 50 health states) using the CHU9D instrument [Burgess and Street, 2006; Flynn et al, 2008].

This pilot study was undertaken on a small sample of young adolescents and hence the findings must be interpreted with caution. The participants in our study were intelligent, well-educated and from a metropolitan area of Adelaide with relatively high socio-economic status. The difficulties that the participants in our study experienced with SG and TTO indicate that these techniques would be highly unreliable in young adolescents with lower education levels. Although health state values have been generated previously using adolescent age groups [Wassermann et al 2005; Saigal et al 1999], these studies have not reported specifically upon the reliability of the application of the SG technique in young adolescents (aged 11-13 years) and hence the extent to which our findings are reflective or not of previous work in this area is unclear. Whilst the participants in this study exhibited a good understanding of the Best-Worst Scaling questions further research is required to test the feasibility of this approach in young adolescents with lower education levels and in areas of lower socio-economic status. This pilot study was also undertaken using 'cold calling' techniques. We are therefore unable to rule out the possibility that allowing the participants more time to learn about and become more familiar with SG and TTO techniques could have resulted in them being able to more fully understand these approaches and thereby provide more reliable responses. Further research should be undertaken to assess the impact of training and learning in promoting understanding of conventional health state valuation methods in adolescent samples

This pilot study has demonstrated the feasibility of applying Best-Worst Scaling DCE methodology to value a series of health states defined by the CHU9D in a young adolescent sample. The potential difficulties associated with the administration of conventional TTO and SG questions in a sample of this age group, in terms of question framing, ease of understanding and completeness have also been highlighted. The findings of this pilot study lend support to the possibility of applying Best Worst Scaling DCE methodology in large scale health state valuation studies to be carried out

directly for the first time using young adolescent general population samples as the source of preferences.

References

Berg J, Dischler J, Wagner D, Raia J, Palmer-Shevlin N. Medication compliance: a health care problem. *Annals of Pharmacotherapy* 1993; 27 (suppl.): 2-21.

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.

Cavet J, Sloper P. The participation of children and young people in decisions about UK service development. *Child Care, Health and Development* 2004;30:613-21.

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.

Department of Health 2003 *Building on the best, choice responsiveness and equity in the NHS.* Department of Health, London.

Dolan P, Gudex C, Kind P, Williams A. Valuing health states a comparison of methods. *Journal of Health Economics* 1996;15:209-231.

Dolgin M, Katz E, Doctors S, Siegel S. Caregivers perception of medical compliance in adolescents with cancer. *Jnl of Adolescent Health Care* 1986; 7: 22-27.

Drummond M. Sculpher M, Drummond M, O'Brien B, Stoddart GL, Torrance GW. *Methods for the Economic Evaluation of Health Care Programmes*, 3rd edition, Oxford University Press, 2005 .

Finn A, Louviere J. Determining the appropriate response to evidence of public concern: the case of food safety. *Journal of Public Policy and Marketing* 1992;11:12-25.

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* 2008; 6:6.

Flynn TN, Louviere JJ, Peters TJ, Coast J. Best-Worst Scaling: What it can do for health care research and how to do it. *Journal of Health Economics* 2007; 26(1):171-37.

Flynn TN, Louviere JJ, Peters TJ, 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.

Hakim Z, Pathak D. Modelling the EuroQol data: a comparison of discrete choice conjoint and conditional preference modelling. *Health Economics* 1999; 82: 103-16.

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.

Juniper E, Guyatt G, Feeny D. Minimum skills required by children to complete health related quality of life instruments for asthma: comparison of measurement properties. *European Respiratory Journal* 1997; 10: 2285-94.

Kleinet S. Adolescent health: an opportunity not to be missed. *The Lancet* 2007; 369: 1057-1058.

Kovacs M, Goldston D, Obroosky D, Iyengar I. Prevalence and predictors of pervasive noncompliance with medical treatment amongst youths with insulin dependent diabetes mellitus. *Journal of the American Academy of Child and Adolescent Psychiatry* 1992; 31: 1112-1119.

Lancaster K. A new approach to consumer theory *Journal of Political Economy* 1966; 74:134-157.

LeCoq E, Boeke A, Bezemer P, Colland V, Van Eijk J. Which source should we use to measure quality of life in children with asthma: The children themselves or their parents? *Quality of Life Research* 2000; 9: 625-636.

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

McCabe C, Brazier J, Gilks P, Tsuchiya A, Roberts J et al. Using rank data to estimate health state utility models. *Journal of Health Economics* 2006; 25: 418-431.

McFadden D. *Conditional logit analysis of qualitative choice behaviour.* 1973; University of California at Berkley, California.

McIntosh W, Louviere J. Separating weight and scale value: an exploration of best-attribute scaling in health economics. Health Economists Study Group Meeting, Brunel University.

Milne A , Chesson R. Health services can be cool: partnerships with adolescents in primary care. *Family Practice* 2000; 17: 305-308.

Norquist G, McGuire T, Essock S. Cost effectiveness of depression treatment for adolescents *American Journal of Psychiatry* 2008; 165: 549-552.

Organisation for Economic Co-operation and Development (OECD). *Economic valuation of environmental health risks to children*. Paris, OECD Publishing, 2006.

Petrou S. Methodological issues raised by preference based approaches to measuring the health status of children. *Health Economics* 2003; 12: 697-702.

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* (forthcoming: 2009).

Ravens-Sieberer U, Gosch A, Rajmil L. The KIDSCREEN-52 quality of life measure for children and Adolescents: psychometric results from a cross cultural survey in 13 European countries. *Value in Health* 2008; 11: 645-658.

Ryan M, Gerard K. Using discrete choice experiments to value health care programmes: current practice and future research reflections. *Applied Health Economics and Health Policy* 2003; 2:1

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.

Saigal S, Stoskopf B, Feeny D, Furlong W, Burrows E, Rosenbaum P, Hoult L. Differences in preferences for neonatal outcomes among health care professionals, parents and adolescents. *JAMA* 1999; 281: 1991-1997.

Salomon J. reconsidering the use of rankings in the valuation of health states: a model for estimating cardinal values from ordinal data. *Population Health Metrics* 2003, 1.

Shaw R. treatment adherence in adolescents: development and psychopathology. *Clinical Child Psychology and Psychiatry* 2001; 6: 137-150.

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

Stevens K. Working with children to develop dimensions for a preference based generic paediatric health related quality of life measure. Health Economics and Decision Science Discussion Paper 08/04, ScHARR, University of Sheffield, 2008a

Stevens K. Developing a descriptive system for a new preference based measure of health related quality of life for children. Health Economics and Decision Science Discussion Paper 08/10, ScHARR, University of Sheffield, 2008b.

Theunissen NC, Vogels TG, Koopman HM, et al. The proxy problem: child report versus parent report in health-related quality of life research. *Qual Life Res*. 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 G. Measurement of health state utilities for economic appraisal. *Journal of Health Economics* 1986; 5:1-30.

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.

Ungar W, Santos M. Trends in paediatric health economic evaluation 1980 to 1999. *Arch Dis Child* 2004; 89: 26-29.

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

Varni J, Limbers C, Burwinkle T. Parent proxy report of their children's health related quality of life: an analysis of 13,878 parents' reliability and validity across subgroups using the PedsQL 4.0 Generic Score Scales. *Health and Quality of Life Outcomes* 2007, 5: 2.

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.

Wasserman J, Aday L, Begley C, Ahn C, Lairson D. Measuring health state preferences for hemophilia: development of a disease specific instrument. *Haemophilia* 2005; 11: 49-57

Appendix 1: Child Health Utility 9D (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. *Schoolwork/Homework (such as reading, writing, doing lessons)*

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 (things like eating, having a bath./shower, getting dressed)*

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 (things like playing out with your friends, doing sports, joining in things)*

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 1: Participants

Gender	**Male** **N (%)**	**Female** **N (%)**	
	10	6	
Age	**11 years** **N (%)**	**12 years** **N (%)**	**13 years** **N (%)**
	5	9	2

Table 2: Responses to CHU9D (n=16)

CHU9D Attributes and levels	***Frequency (%)***
1. Worried	
I don't feel worried today	7 (44%)
I feel a little bit worried today	5 (31%)
I feel a bit worried today	4 (25%)
I feel quite worried today	-
I feel very worried today	-
2. Sad	
I don't feel sad today	14 (88%)
I feel a little bit sad today	2 (12%)
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	4 (25%)
I have a little bit of pain today	4 (25%)
I have a bit of pain today	8 (50%)
I have quite a lot of pain today	-
I have a lot of pain today	-
4. Tired	
I don't feel tired today	8 (50%)
I feel a little bit tired today	5 (31%)
I feel a bit tired today	2 (13%)
I feel quite tired today	-
I feel very tired today	1 (6%)
5. Annoyed	
I don't feel annoyed today	11 (69%)
I feel a little bit annoyed today	4 (25%)
I feel a bit annoyed today	1 (6%)
I feel quite annoyed today	-
I feel very annoyed today	-
6. School Work/Homework	
I have no problems with my schoolwork/homework today	8 (50%)
I have a few problems with my schoolwork/homework today	5 (31%)
I have some problems with my schoolwork/homework today	3 (19%)
I have many problems with my schoolwork/homework today	-
I can't do my schoolwork/homework today	-

Table 2 (continued): Responses to CHU9D (n=16)

7. Sleep	
Last night I had no problems sleeping	8 (50%)
Last night I had a few problems sleeping	4 (25%)
Last night I had some problems sleeping	4 (25%)
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	9 (56%)
I have a few problems with my daily routine today	7 (44%)
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	10 (63%)
I can join in with most activities today	4 (25%)
I can join in with some activities today	2 (13%)
I can join in with a few activities today	-
I can join in with no activities today	-

Table 3: CHU9D Utilities

Mean (SD)	***Median (IQ range)***	***Range***
0.85 (0.46)	0.85 (0.81-0.88)	0.77-0.92

Table 4: Responses to best worst scaling DCE question

Best N (%)	***Health State Card 1***		***Worst N (%)***
-	1.	I feel quite worried today	3 (19%)
8 (50%)	2.	I don't feel sad today	-
-	3.	I have a bit of pain today	-
-	4.	I feel quite tired today	-
2 (13%)	5.	I don't feel annoyed today	-
1 (6%)	6.	I have a few problems with my schoolwork today	2 (13%)
-	7.	Last night I had many problems sleeping	8 (50%)
-	8.	I have many problems with my daily routine today	3 (19%)
5 (31%)	9.	I can join in with a few activities today	-

Best N (%)	***Health State Card 2***		***Worst N (%)***
-	1.	I feel a little bit worried today	-
-	2.	I feel very sad today	3 (19%)
8 (50%)	3.	I don't have any pain today	-
-	4.	I feel very tired today	2 (13%)
5 (31%)	5.	I don't feel annoyed today	-
-	6.	I have some problems with my schoolwork today	-
-	7.	Last night I couldn't sleep at all	7 (44%)
-	8.	I have a few problems with my daily routine today	-
3 (19%)	9.	I can join in with some activities today	4 (25%)

Best N (%)	***Health State Card 3***		***Worst N (%)***
-	1.	I feel very worried today	2 (13%)
2 (13%)	2.	I feel a bit sad today	-
-	3.	I have quite a lot of pain today	3 (19%)
-	4.	I feel quite tired today	-
14 (88%)	5.	I don't feel annoyed today	-
-	6.	I have some problems with my schoolwork today	-
-	7.	Last night I had many problems sleeping	1 (6%)
-	8.	I can't do my daily routine today	5 (31%)
-	9.	I can join in with no activities today	5 (31%)

Best N (%)	***Health State Card 4***		***Worst N (%)***
-	1.	I feel a bit worried today	2 (13%)
-	2.	I feel a bit sad today	2 (13%)
-	3.	I have a bit of pain today	2 (13%)
1 (6%)	4.	I feel a bit tired today	-
1 (6%)	5.	I feel a bit annoyed today	-
1 (6%)	6.	I have some problems with my schoolwork today	4 (26%)
-	7.	Last night I had some problems sleeping	4 (26%)
1 (6%)	8.	I have some problems with my daily routine today	1 (6%)
12 (76%)	9.	I can join in with some activities today	1 (6%)

Table 4 (continued): Responses to best worst scaling DCE question

Best N (%)	***Health State Card 5***		***Worst N (%)***
-	1.	I feel quite worried today	1 (6%)
-	2.	I feel very sad today	5 (31%)
13 (81%)	3.	I don't have any pain today	-
-	4.	I feel quite tired today	1 (6%)
-	5.	I feel very annoyed today	2 (12%)
-	6.	I can't do my schoolwork today	6 (38%)
-	7.	Last night I had many problems sleeping	1 (6%)
-	8.	I have a few problems with my daily routine today	-
3 (19%)	9.	I can join in with a few activities today	-

Table 5: Responses to SG and TTO tasks

Participant no.	***Valuation technique***	***Health state number presented***	***Practice health state value [1]***	***Health state value [2]***	***Valuation technique preference***	***Interviewer judgement of task comprehension***	***Reason/s***
1	SG	1	0.50	0.60	SG	No	R1
2	SG	2	0.30	0.30	SG	Yes	R4/R5
3	SG	3	NIPI	0.30	SG	No	R1/R3
4	SG	4	NIPI	NIPI	SG	No	R1/R3
5	SG	5	0.30	0.30	SG	No	R1
6	SG	1	NIPI	NIPI	NP	No	R3
7	SG	2	NIPI	NIPI	NP	No	R1/R3
8	SG	3	NIPI	0.60	DCE	No	R1/R3
9	TTO	1	<0.01	<0.01	DCE	No	R2/R3
10	TTO	2	<0.01	<0.01	DCE	No	R2/R3
11	TTO	3	0.30	0.30	DCE	Yes	R4/R5
12	TTO	4	<0.01	<0.01	DCE	No	R2/R3
13	TTO	5	<0.01	<0.01	DCE	No	R2/R3
14	TTO	1	<0.01	<0.01	DCE	No	R2/R3
15	TTO	2	<0.01	<0.01	TTO	No	R2/R3
16	TTO	3	<0.01	<0.01	DCE	Ni	R2/R3

Abbreviations
NIPI: No indifference point identified
NP: No preference
R1: Did not consider certain outcome (SG) or intermediate health state (TTO) when making choices
R2: Reluctance to trade healthy life years
R3: Unable to identify a point of indifference
R4: Fully considered certain outcome (SG) or intermediate health state (TTO) when making choices
R5: Able to identify a point of indifference