The Development of a Preference Based Paediatric Health Related Quality of Life Measure for use in Economic Evaluation

Katherine Josie Stevens
BSc (Hons) (York), MSc (York)

A thesis prepared for the degree of Doctor of Philosophy

School of Health and Related Research
The University of Sheffield

08 December 2008

Testing the Draft Descriptive System in a General Paediatric Population

7.1 Introduction

Having developed a draft descriptive system (chapters 5, 7a and 7b), it was important to test its performance, to see if it is reliable and valid in a paediatric population. Matza (2004) notes that when developing a questionnaire, it is important to carry out pilot testing and cognitive debriefing with children so that you can determine whether children understand the questions and provide reliable and valid responses and also test whether this is the same across all ages. Cognitive debriefing invites children to comment on the measure in terms of the content, layout, terminology and response options.

This chapter reports on the testing of the instrument in a general paediatric population and chapter 8a reports on testing with a clinical paediatric population.

To assess the psychometric performance of an instrument, the most commonly used criteria are practicality, reliability and validity (Brazier 1999). These criteria should really only be used on the descriptive system prior to any scoring (Brazier 1999).

Practicality looks at issues such as the length, difficulty and acceptability of an instrument to its intended population, usually done by measuring response and completion rates, distributions across response levels and time taken to complete (Brazier 1999). It can also look at the reading age for the instrument.

Reliability looks at whether a measure can produce a series of results over repeated measurements on an unchanged population with a minimum amount of random error (Hays 2005), that is, whether it produces the same results on a population where there is no evidence of change. There are different types of

reliability, including test - retest (stability over time), inter-rater (agreement between raters), intra-rater (agreement of the same rater's scores) and internal consistency (whether items and domains in a scale are related) (Brazier 1999).

Validity examines whether an instrument measures what it is intended to measure in its descriptive system (Hays 2005). There are different types of validity, including content validity (whether all aspects of the attribute to be measured are covered), face validity (do the items appear to be measuring what they claim to measure) and construct validity (whether the measure can discriminate between groups that it should be able to, as defined by a measure of a similar construct). In addition, sensitivity to change or responsiveness (whether the instrument is responsive to change) can also be looked at (Hays 2005).

Psychometric criteria are useful firstly for assessing the performance of an instrument, and are also useful for refining an instrument. Here they are used for both purposes, in addition to testing whether children can self complete the measure and selecting which is the best form of wording for each dimension. The aim was therefore to test the draft descriptive system in a paediatric sample of the general population in terms of its psychometric properties, to determine whether children are able to self complete and to test the alternative wordings for dimensions.

7.2 Methods

7.2.1 Piloting

Prior to the main study, the descriptive system was first piloted on the same 10 children who had piloted the ranking work in chapter 7b. After applying the scales to the 17 questions, a few other additions were made to form the draft descriptive system. Where the dimension was quite broad; daily routine and joining in activities, an explanation of what this included was given in brackets afterwards. This explanation was based on the qualitative research. An

additional question was added at the end, which was question 17 from the Health Utilities Index 2 (HUI2) (Health Utilities Index). This question is not actually part of the HUI2 descriptive system but is just a general question asking respondents to rate their own health and can be useful when testing the validity of a measure. It is the same scale that was used for the sampling of children in the qualitative interviews in chapter 5. The recall period was chosen to be today as this made sense for the type of question children were answering (severity) and also as discussed in chapter 6a shorter recall period is better for children.

All 10 children were given the questionnaire to complete by themselves and told to ask for help if they needed it.

7.2.2 Results

All children managed to self complete, although one child asked for the instructions to be read to them. Children enjoyed completing the questionnaire and made helpful suggestions, such as what font size was appropriate and suggested the use of more colour. There were no problems with reading and understanding the questions although some children commented that some of the questions were similar.

7.2.3 Conclusions

The piloting was successful and the final draft descriptive system took into account the comments of the children in terms of font size and the use of colour. Colour pictures were added to make it more friendly and accessible for children. These pictures were not added next to questions as they could bias the respondent answers as they may interpret the picture differently from the wording of the question. The final draft descriptive system taken forward for the main study is shown in Appendix 7A.

7.2.4 Main study

All children from both schools involved in the research whose parents had consented them into the study originally were eligible to be included in order to

achieve as large a sample as possible. Children were approached in school by the researcher (KS) and asked if they would like to take part in this stage of the research. If the child consented then the draft descriptive system was given to them as a questionnaire to complete. (Appendix 7A) The child completed the questionnaire in a quiet place on their own, either in the school library or the dining room. The researcher sat next to them for the duration to be available to help with any questions. The researcher explained that all answers would remain confidential, there were no right or wrong answers and that they could ask for help from the researcher at any time, whether that be reading or explanation of meaning or what to do. The researcher then allowed the child to self complete the questionnaire by themselves whilst remaining sat next to them in case they had any problems. The start and end time were recorded.

Questions evaluating the questionnaire were then asked to obtain data on practicality and validity. These questions are shown in Appendix 7B. The first seven questions (C1 to C7) were administered by the researcher to the child after they had completed the questionnaire and the final five questions (K1 to K5) were completed by the researcher once the child had gone back to lessons. The aim of the questions administered by the researcher to the child was to find out the following:

- Whether children had been off school ill in the previous week in order to see if the descriptive system could discriminate between those who had and those who had not. (question C1 (construct validity))
- Whether children thought there was anything missing from the content of the descriptive system (question C2 (content validity))
- Whether children thought any questions were the same and if they were, which wording they preferred (question C3 and C4 (testing the wording))
- Whether there were any questions the children did not understand (question C5 (face validity))

- Whether there were any questions the children found difficult (question C6 (face validity))
- Whether there were any other comments they had (question C7 (overall comments))

The aim of the questions completed by the researcher was to document whether children read and understood the instructions (practicality) and questions (face validity) themselves , whether they needed help on particular questions, (face validity) the nature of that help and whether they were able to self complete.

7.2.5 Analysis

Two of the key psychometric criteria – practicality and validity were tested in this study. It was not possible to test the third key criteria, reliability, due to time and resource constraints. It was also not possible to test responsiveness or sensitivity to change as only one observation per respondent was obtained.

The practicality of the draft descriptive system was tested by looking at the mean time taken to complete, the response rate, the completion rate and distribution across levels by question and whether the child read and understood the instructions. This was also tested by age to see whether there were any implications for reading age.

Content validity was assessed by whether children identified things they felt were missing from the content of the descriptive system. Face validity was tested by looking to see if children were able to read and understand the questions and looking at what they found difficult or needed help on and why. Construct validity was tested by splitting the data according to whether children had been absent from school due to illness in the last week and looking at the distribution of the responses across the levels compared to those who had not

been absent. If there was a difference with those absent from school at lower levels, this provides some evidence of construct validity. Differences in the distributions were tested using a non parametric Mann Whitney test, which is the non parametric test for comparing independent samples where a parametric test is not possible (as the assumptions of parametric tests are not valid here). Construct validity was also examined by scoring the descriptive system and seeing if this correlated with the child's rating of their own health. Assuming that the questions can be scored so that the highest level is 1 and the next level down is 2, a score can be generated for each child by summing the levels across the questions. This assumes that there are equal intervals between levels. Only one form of wording (question) for each dimension was used so as not to double count (as there was more than one type of wording for each question (chapter 6a)). The questions included were worrying, sad, pain, schoolwork, tired, annoyed, sleep, embarrassed, jealous, daily routine and joining in activities. When compared to the child's rating of their own health, these should correlate, in that a higher score should be associated with a lower level of health. This scoring is very crude and does not take into account any preferences, however it does give an indication of whether the descriptive system is able to detect differences between children with different levels of health. The score was also calculated for those who had been off school in the last week and those who had not to test if there was a significant difference. It would be expected those who had been off school would have a higher mean score. It should be noted that some of these tests of construct validity compare 2 indicators of health over 2 different time frames – the child is rating their

health today and this is being compared to whether they have been off school ill in the last week. This may partially limit the conclusions that may be drawn from these tests as the child may be fully recovered by the time they are back at school and therefore whether they have been off school in the last week ill may not be the most reliable indicator. However, it is likely that children who have been off school ill may still not be fully recovered and so some effect may be picked up. Stronger conclusions can perhaps therefore be drawn from the other test of construct validity which use the same time frame (the score and the child's rating of their own health).

Whether children were able to self complete was assessed by looking at the proportion who were able to self complete (question K5).

Finally, questions identified as the same by children were examined to see if there was a substantial preference for one form of wording.

7.3 Results

7.3.1 Sample

34 children from School A (Firs Hill Community Primary School) and 119 from School B (Hunter's Bar Junior School) were approached, giving a total sample of 153. Three children did not consent to take part, giving a total of 150 children who participated. The characteristics of the sample who participated are shown in Table 7.1. For comparison, the characteristics of the same age population in Great Britain are also shown, according to the 2001 census (National Statistics).

Table 7.1: Characteristics of the participating sample

	N	%	Census (%)
Male	79	47.9	51.2
Y3	35	22.9	19.7*
Y4	37	24.2	20.3*
Y5	45	29.4	20.7*
Y7	37	23.5	20.1*
Excellent health	71	47.7	-
Very good health	44	29.9	-
Good health	25	17.0	-
Fair health	4	2.7	-
Poor health	2	1.3	-
Health missing	4	-	-
White	93	72	88.7
Mixed/dual heritage	14	9.3	2.8
Asian or Asian British	30	20.4	5.5
Black or Black British	4	2.7	2.5
Chinese	0	0	0.4
Other	7	2.7	0.4
Ethnicity missing	3	-	-

* The census data are for 8,9,10 and 11 year olds. As this work was carried out in the summer term, it is assumed most Y3 children would be 8.

A good split across year group and a good male – female balance was achieved. The level of health was less balanced, with more children in the excellent and very good health categories, however this is unsurprising given this is a school population. Asian/Asian British children were slightly over represented and white children were slightly under represented compared to the population in Great Britain.

7.3.2 Practicality

The time taken to complete the questionnaire was low with a mean of 3.83 minutes and ranged from 3.07 minutes to 4.89 minutes across year group, decreasing with age as might be expected. The mean, median, minimum and maximum time broken down by year group is shown in Table 7.2 in Appendix 7.C.

The response rate was excellent at 98% across all year groups and ranged from 92% to 100% across year group. The response rate by year is shown in Table 7.3 in Appendix 7.C.

Table 7.4 below shows the completion rate for each question for all children in the sample. There was very little missing data; all missing data came from the Y3 year group and from different questions. Table 7.5 in Appendix 7.C shows the completion rate for each question for Y3.

Table 7.4: Completion rate by question, all cases

Question	Responses (n)	Missing (n)	%missing
q1 (Worrying)	150	0	0
q2 (Sad)	150	0	0
q3 (Weak)	150	0	0
q4 (Angry)	150	0	0
q5 (Pain)	149	1	0.7
q7 (Frustrated)	150	0	0
q7 (Hurting)	150	0	0
q8 (School Work)	150	0	0
q9 (Upset)	149	1	0.7
q10 (Tired)	149	1	0.7
q11 (Annoyed)	148	2	1.3
q12 (Scared)	149	1	0.7
q13 (Sleep)	149	1	0.7
q14 (Embarrassed)	149	1	0.7

q15 (Jealous)	149	1	0.7
q17 (Daily Routine)	149	1	0.7
q17 (Joining in activities)	149	1	0.7
q18 (Rating health)	149	1	0.7

The distribution of responses across the levels (or response options) for each question was good although the majority of children were responding with the top levels of each question, which is not surprising given the nature of the population being tested. For the last question (q18) which asked children to rate their health from excellent to poor, most children were either excellent, very good or good, with the majority of children very good. The tables 7.7 to 7.24 in Appendix 7.C show the distributions.

Table 7.25 below shows the number and percentage of children who read the instructions themselves. It also shows the breakdown by each year group. Overall, this was high (95.33%) and generally increased with age, although Y3 and Y4 were similar.

Table 7.25: Reading the instructions

Year	Read the instructions themselves	N	%
All cases	yes	143	95.33
3	yes	32	91.43
4	yes	31	91.18
5	yes	44	97.78
6	yes	37	100

Table 7.26 below shows the percentage of children who were able to understand the instructions without any help. It also shows the breakdown by

year group. This is also high overall (88%) and increases with age apart from Y4 which was 100%.

Table 7.26: Able to understand the instructions

Year	Able to understand the instructions	N	%
All cases	yes	132	88
3	yes	27	77.14
4	yes	34	100
5	yes	39	87.77
6	yes	32	88.89

7.3.3 Content Validity

The percentage of children saying something was missing from the content of the questionnaire was fairly low (around 10%, varying from 5.88 to 13.89% by age group). Table 7.27 below details what was stated as missing by the children.

Table 7.27: What children identified as missing from the content of the descriptive system (n=1 for all items)

breathing - hurts to breathe sometimes
doesn't eat enough, knows should try more stuff
eating - how much you can eat
energy/lively
going to the shop
itches - eczema
itching because of eczema
nose running and throat hurting
add itching – has impetigo
should ask about junk food - shouldn't eat so much of it
sometimes can't do the hoovering
sometimes gets bad headaches and then can't breathe properly

stressed - if feel upset, can't let go of it, which makes them more stressed
teeth - need some pulled out

7.3.4 Face Validity

Table 7.28 below shows the number and percentage of children who read the questions themselves (this was observed and recorded by the researcher). It also shows the breakdown by each year group. Overall this was high at 92% and ranged from 80% in the youngest age group to 100% in the oldest age group, increasing with age in between which is expected.

Table 7.28: Reading the questions

Year	**Read all the questions themselves**	N	%
All cases	yes	138	92
3	yes	28	80
4	yes	30	88.24
5	yes	44	97.78
6	yes	37	100

Table 7.29 below details which questions had to be read to children and also shows which year group they were. The number of children needing the particular question to be read is also shown.

Table 7.29: Which questions had to be read to children

Question Number	N	Year
11 (Annoyed)	2	Y3
8 (School work)	1	Y3
1 (Worrying)	1	Y3
All questions	4	Y3(2) Y4(2)
14 (Embarrassed)	2	Y4

17 (Daily routine)	1	Y4
17 (Able to join in activities)	1	Y5

Table 7.30 below shows whether help was needed on particular questions, broken down by year group and Table 7.31 details the nature of that help. The number of children asking for the type of help is also given. 35.57% of children needed help on questions across all groups, but for Y4 it was higher (50%), This help included anything that the children requested, including asking what a word said, explanations of terms or explanations of how it might affect your health. The nature of this help varied – many children asked for clarification about Q7 (hurting), as to what exactly hurt and a couple of children asked if it was the same as the pain question. One child asked if it meant emotional hurt. Another child asked if it was physical or emotional pain.

12 children asked what 'affected' meant in the context of school work and this also came up in the context of sleep – children wanted to know what affected sleep was. 7 children also asked what the word 'frustrated' meant.

Table 7.30: Help needed on questions

Year	Help needed on questions	N	%
All cases	yes	53	35.57
3	yes	13	38.24
4	yes	17	50
5	yes	12	27.77
6	yes	11	30.57

Table 7.31: Nature of help required

Question	Nature of help (N)
1 (Worrying)	Couldn't decide which level they were then filled levels 1 - 4 in and got muddled. Once is had been explained again, they understood it.
	why would you feel worried?
	asked if they could tick 2 boxes
	clarified how to answer it.
	what worrying meant
	worried about what?
2 (Sad)	asked for explanation of what you do to answer the question.
	clarified whether it was how they felt today, knew it was because of health.
3 (Weak)	what weak meant. After explanation about how to answer, went back and changed q1.
	clarified 'I don't feel weak' means I am not feeling weak.
	what weak means (3)
5 (Pain)	is it physical or emotional pain?
	asked if pain could be anywhere on their body.
	checked whether was about today (2)
	does it mean injuries?
7 (Frustrated)	what frustrated meant. (7)
7 (Hurting)	isn't this the same as q5? (2)
	what part of you that hurts? (4)
	asked what it meant - which bit of you hurts? Is it the same as pain?
	is it emotional or physical hurt - they asked ' is it because it hurts you to think about your health'
	what do you mean by hurting.
	what hurt meant.

	what hurting meant and asked if it was any day or today
8 (School Work)	what affected meant (11)
	schoolwork - is it activities? Explained it was lessons - then understood fine
	knew what affected meant, but clarified that my schoolwork is not affected means having no problems
	what affected meant and in context of school work
10 (Tired)	can never get to sleep easily so is tired today - is this a health problem?
	was it about today
	what tired said
	checked that tired because went to bed late was invalid - so had understood correctly
12 (Scared)	How scared related to health
	why would you feel scared?
13 (Sleep)	is it about how well you slept last night?
	couldn't get to sleep, unsure if this was because of health
	had trouble with sleep but doesn't know why. Wasn't because of TV.
	how would your sleep be affected? What would cause it?
	not sure which level they were
	sleep - had problems because of scary film and asked if valid.
	What affected meant
	sleep affected - what this means in context of sleep
	What quite affected meant in relation to sleep
	sleep - sleep is not good generally but didn't know if this was because of health
14 (Embarrassed)	Why would health affect it? But realised for serious health problems it would
	said that get embarrassed because of eczema in P.E. because gets out of breath, but not happened today, so ticked level 1

	what embarrassed said and meant
	asked if was embarrassed because of health
	what embarrassed meant
	embarrassed - reading the word - but knew what it meant
15 (Jealous)	didn't feel jealous and didn't understand why it was there
	how would health affect you feeling jealous
	what jealous means
	why would you feel jealous
17 (Daily Routine)	thought washing was washing clothes and they don't do that
	what routine said, knew the meaning
	what does it mean. Understood fine on explanation
	daily routine - needed explanation
	routine – however once the researcher pointed to the explanation in brackets they got it fine
	in-between levels 1 and 2, unsure which to put, so decided was closest to 1
	what it means
	reading it
17 (Join in activities)	What activities were included
	Why related to health
	didn't understand, but when explained, changed answer
	is it because of you or because of e.g. your friends
18 (Rating Health)	where to rate - has asthma
	choosing their level
	is it healthy or health
	in between very good and good
General	checked it was about them
	explanations as to why things might be a problem
	kept checking how should answer - got it by q7 and then went back to change it. Forgot it was about today because of health but then got it.

	before had answered any questions, clarified what had to do
	what affected meant.

Approximately 23% of children had question(s) that they didn't understand. This decreased with age apart from a negligible increase from Y5 to Y6 (13.33% to 13.89%). This is shown in Table 7.32 below broken down by year group.

Table 7.32: Questions not understood

Year	Questions not understood	N	%
All cases	Yes	34	22.82
3	Yes	15	44.12
4	Yes	8	23.53
5	Yes	7	13.33
6	Yes	5	13.89

Table 7.33 below details whether any questions were judged to be difficult by children and is broken down by year group.

Table 7.33: Questions found difficult

Year	Questions found difficult	N	%
All cases	Yes	51	34.23
3	Yes	11	32.355
4	Yes	14	41.18
5	Yes	17	37.78
6	Yes	9	25

Approximately 34% of children judged one or more of the questions as difficult (ranging from 25% to 41%). The main problem was that they had to think about which level they were, this occurred in questions 7 (frustrated), 8 (school

work), 13 (sleep), 17 (daily routine), 17 (joining in activities). Several children found it difficult to judge what level they were when rating their health (question 18). Some asked if it was how healthy you were, others simply found it difficult to rate their own health. Three children asked how or why you would feel jealous because of your health (question15). Similarly 2 children asked why you would feel embarrassed because of your health (question 14). Some children had problems with question 13 (sleep) as they had poor sleep anyway and were not sure if it was because of their health. Table 7.34 details the questions which needed help and the reasons why.

Table 7.34: Which questions were found difficult and why

Question	What was difficult
2 (Sad)	relates to eating, feels sad can't eat more
5 (Pain)	depends on the time of day
	had to think carefully about them but not too difficult
7 (Frustrated)	what frustrated meant
	understood meaning but difficult to think what level they are
	felt frustrated, but not sure which level to choose
7 (Hurting)	had to think carefully about them but not too difficult
8 (School work)	schoolwork - took time to think about
	which level
	which level they were.
9 (Upset)	relates to eating, feels sad can't eat more
10 (Tired)	started thinking about another day and is usually tired so difficult to think about whether is because of health.
	lots to think about
11 (Annoyed)	reading the word, got it in the end

13 (Sleep)	because they are a bad sleeper
	knowing which level - but was thinking about other days and in general
	when get hot in night can't sleep. Get hot because of running around
	whether sleep was normal or was because of health - sometimes find it difficult to sleep but don't know why,
	which level to put (1st or 2nd)
	affected
	depends on how ill they are
	had to think carefully about them but not too difficult
14 (Embarrassed)	didn't know how would feel embarrassed
	not sure why, but did
	why would you feel embarrassed about your health?
	understood meaning but difficult to think what level they are
15 (Jealous)	but understood it as your friends may be more healthy
	not sure what it meant
	couldn't think how would feel jealous, but then thought about it and got it - jealous of others who can do rounders today but they can't because of pulled muscle. Q is quite good because gives you a chance to say how you feel
	how would you feel jealous?
	why would you feel jealous?
17 (Daily Routine)	knowing which level
17 (Join in activities)	because quite long
	didn't really understand it
	didn't understand, but when explained, changed answer
	due to not much available in their area
	had to be read to them
	it is not health stopping them do all activities

	not sure what it meant.
	sometimes felt jealous, not really a reason
	sometimes fine, sometimes cough a lot
	wasn't sure was level 1 or 2
	which level in between 2
	had to think carefully about them but not too difficult
18 (Rating health)	think they are in the middle, between excellent and v good but ticked v good
	can't rate own health
	deciding levels
	didn't know how good health is, just was guessing
	difficult to know which level, was thinking of healthy - e.g. diet and exercise
	hard to rate your own health
	knowing which level they are - wasn't sure what to answer
	not sure which level they were
	wasn't sure of their level
	which level they were
	difficult to know whether was excellent or v good, hard to rate
General	at beginning, difficult to think how feel, but then got it.
	choosing between levels
	where had to choose between top 2 levels - close
	which level they were

7.3.5 Construct Validity

28 children (18%) had days off school due to illness in the week prior to completion of the questionnaire.

The tables 7.35 – 7.51 in Appendix 7.C show the distribution across the levels (or response options) for each question for those who were off school ill in the last week and those who were not. The distributions across the levels are different

between those children who were off school and those who were not. Those who were had lower percentages at the higher levels, apart from one question (question 17, daily routine). Question 13 (sleep) had quite a big difference and all other questions showed some difference, even if it was small. It was not possible to undertake a Mann Whitney test as the number of children in the off school group was too small (28). Instead, children were categorized into those who were at level 1 (optimal) and those who were not at level 1 (sub optimal) and a χ^2 test was undertaken to compare the groups. The results only showed a difference for question 2 (sad) and question 13 (sleep) (at $p<0.05$).

The scores generated from summing the levels on the descriptive system from the selected eleven questions correlated well with the child's rating of their health, in that a lower rating of health correlated with a higher score. (The score is the sum of the responses to questions). Although this is a crude scoring method and takes no account of preferences, this does demonstrate that the descriptive system can discriminate between different levels of health, giving some evidence of construct validity. Table 7.52 below shows the mean, median, minimum and maximum score for each level of health group. A one way ANOVA test showed a significant difference in mean score by level of health (F=2.076, 0.087) at $p<0.1$.

Table 7.52 Score by rating of health

Level of health	Mean	Median	Min	Max	n
1	23.3	21	18	44	51
2	29.3	25	19	94	74
3	30.2	28	21	57	30
4	54	54	49	59	2
5	-	-	-	-	0

Similarly, those children who were off school in the last week had a higher mean score than those who were not. Table 7.53 below shows the scores for those children who had been off school in the last week and those who hadn't.

Table 7.53: Score by whether off school or not in the last week

	Mean	Median	Min	Max	n
Off school	17.37	14	11	70	28
Not off school	19.87	17	11	38	119

The difference in mean score was significant at p=0.05 (independent samples t test) and again gives some evidence of construct validity as the questionnaire can discriminate between these groups.

7.3.6 Whether children were able to self complete

Table 7.54 shows the percentage of children able to self complete the questionnaire and is broken down by year group. It is very high overall at nearly 97%, years 4 to 7 are similar at 97% and year 3 is the lowest at 91% however it was expected that this may be lower as they are the youngest age group.

Table 7.54: Able to self complete

Year	Able to self complete	N	Valid %
All cases	yes	143	95.97
3	yes	31	91.18
4	yes	33	97.07
5	yes	44	97.78
6	yes	35	97.23

7.3.7 Wording

Approximately two thirds of children judged one or more questions to be the same as another question. This varied by year group, Y3 and Y4 was about 44% and the percentage was much higher in the older children (Y5 82% and Y7 75%). Table 7.55 below shows the break down by year group.

Table 7.55: Children judging any questions to be the same

Year	Questions the same	N	Valid %
All cases	Yes	94	73.09
3	Yes	15	44.12
4	Yes	15	44.12
5	Yes	37	82.22
6	Yes	27	75

One of the most frequently mentioned pairs of questions was 10 (tired) and 13 (sleep), however when children were asked which they preferred, on reflection the majority said to keep them both as they realised they were different. Question 5 (pain) and (7) hurting were the most frequently mentioned pair and there was a preference for pain of 25:12. Questions 2 (Sad) and 9 (upset) was another pair and there was an equal split of preferences for the different wordings at 20:20. Questions 4 (angry) and 11 (annoyed) were another frequently mentioned pair, with a preference for annoyed by 7:4. Questions 7 (frustrated) and 11 (annoyed) were stated as the same, with a preference for annoyed at 13:7. The rest of the questions had lower numbers of children and are listed in Table 7.57 in Appendix 7.C, which details all the combinations of questions that were judged to be the same, the number of children stating they were the same and the preferences of these children for the wording.

Other comments that were made by children after they had completed the questionnaire were favourable and many children commented that they had enjoyed completing the questionnaire and welcomed the opportunity to give their views. Suggestions to improve the questionnaire included increasing the size of the boxes, making the hurting question more specific, putting the writing in colour and making it clearer how to answer it, for example putting in an example question. Some children commented there were too many levels within each question and they are not all necessary. Some children noted that the explanation in brackets in questions like 16 (daily routine) and 17 (joining in

activities) was essential to understanding the question. The full list of comments are given in Table 7.56 in Appendix 7.C

7.3.8 Observations and other Comments by the Researcher

There were two main categories of observations by the researcher, firstly whether the children were thinking about their health when thinking about how to answer the questions (assessed by the question asking children whether they remembered the questionnaire was about their health at the end and also any think aloud comments the children made) and secondly how they went about answering the questions, as not every child did them in numerical order. The comments are listed below.

Whether they were thinking about their health

- Questions 3(weak) and 10(tired) were because they went to bed late
- They were a bit tired because went to bed a bit late
- Other reasons too apart from health
- Remarked that was tired because they were up reading
- Said occasionally started to think about other things not just health
- Commented about their eczema, got a bit distracted
- Commented that kept forgetting was about health
- Mostly remembered was about today because of health. Drifted a little at the start.
- Started to think about other days a bit and some questions not because of health.
- The answer to Question 17 was because of friends
- Thinking mostly about today because of their health
- Though about their health until about q13/14. Question 17 was because of not being able to play with Y5, not health related
- Was not thinking about today, was putting all experience together

Style of answering

- Did across the page, i.e. 1, 7, 2, 8 etc

- Did in the order 1, 2, 3, 7, 8, 9 etc
- Did questions in a random order, missed q14 initially but saw it and completed it. Read out the answers as answered them for first 2 questions
- Forgot was about today - but was v quiet and may have just been saying this
- Didn't do them in numerical order. Not sure really understood. Stopped at q12 and thought this was the end.

7.4 Discussion

The response rate was very high (98% overall across all years) and completion rates were high with the only missing data appearing in the youngest children. The mean time to complete was under 4 minutes overall and decreased with age, which is what might be expected as the older children can complete it more quickly. This is low compared with evidence from the other generic paediatric measures reviewed in Chapter 3, which range from 10 to 40 minutes, with the exception of the HUI2/3 where one study found it was just 3.1 minutes, however this was just one study in cancer and it seems unlikely that this would be the case all the time as the HUI2/3 is 15 questions long and the response choices are often long. The percentage able to read the instructions was high (>91%) and mainly increased with age. The percentage of children able to understand the instructions was slightly lower (88% overall), as some children asked for an explanation after they had read the instructions. The distribution across the response levels was good although it was quite skewed as many children were choosing the upper levels, which reflects the type of population it was tested on; a fairly healthy sample from the general population. For some questions however, there was more of a spread across the levels, such as questions 5(pain), 7(hurting), 10(tired) and 13(sleep).

Content validity was good as the majority of things children felt were missing were symptoms, such as trouble breathing, itching, throat hurting, and headaches. Other things included hoovering, eating and stress. Every item

missing brought up by children had occurred in the interviews in the first part of this work. Children had said that itching can lead to pain/hurt and so this is covered by this dimension. Eating is part of the daily routine question, going to the shops is covered by the question about being able to join in the activities you want to. Headaches and throat hurting feed into the pain or hurting dimensions. Hoovering is part of the joining in activities dimension, dental/tooth problems led to pain and stress led to worry.

There was evidence of face validity too, with the percentage of children able to read the questions being high and increasing with age, up to 100% for Y7 children. Questions that had to be read were mainly to the younger children. The most common situation was having all questions read to them, if the children's reading was poor.

Approximately 30% of children needed some help with the questionnaire, sometimes this was minimal, for example asking what a word said and sometimes this was more extensive, for example asking for an explanation of the question. One of the most common problems was the word *affected* and what this meant in the context of both sleep and school work.

About one third of children judged one or more of the questions as difficult, with the main problems being deciding which level they were, why you would feel jealous or embarrassed because of your health and some children had problems distinguishing whether their poor sleep was down to their health or not as they had poor sleep anyway.

There was good evidence of construct validity, demonstrated by differences being detected between those who were and were not off school ill in the week prior to completing the questionnaire, both in terms of the distributions across the levels and the scores of the questionnaire. The exceptions was question 17 (daily routine) and the biggest difference was shown in question 13 (sleep). The results of the χ^2 tests when the off school and not off school groups were

compared by question only showed differences for 2 of the questions, even when the data were collapsed into 2 levels (optimal and sub optimal). This may be in part due to the healthy nature of this sample, the difference in recall period and the relatively low numbers tested in this study, however is something that requires further testing. A larger sample would have allowed more formal non parametric testing of the differences between distributions and this is a weakness of this study however is something that can be undertaken in future psychometric testing of the measure.

The measure of being off school for the week prior to completion was quite crude as the reason for absence was mostly acute health problems which had usually disappeared by the time children were back at school, however children were sometimes still on the tail end of a health problem and so this is probably why the differences do appear.

The percentage of children able to self complete was very high and increased with age. This question was recorded as 'yes' if the child was able to select the level that applied to them and tick the appropriate box themselves. It included children who had to have explanations for some words or words read to them, so in this sense they were able to self complete, but only with the assistance of another person on hand to read any words or explain meanings to them.

The strongest identifications of wordings being similar were *sad* and *upset*, *pain* and *hurt* and *frustrated* and *annoyed*. Overall the number of children identifying questions as being the same was quite low but this may be in part due to children not expecting questions on a questionnaire to be similar and thinking they may be being tested on differences in meanings between words.

Other comments noted by the researcher were mainly around the issue of ensuring that children remember they are supposed to be thinking about their health when answering the question. Some children answered the questions for different reasons, for example feeling sad because they had fallen out with a

friend or feeling tired because they had stayed up late watching television. Some children did not answer the questions in numerical order, however this did not make a difference to the completion rate.

Generally the questionnaire performed well in this population and demonstrated good practicality and evidence of content, face and construct validity. However, the sample was a relatively healthy one as all children were in main stream schooling which does limit the testing of the questionnaire. Further work is needed to test the questionnaire in a clinical population where children are sicker to test whether the questionnaire still demonstrates good psychometric properties here.

In addition, it was not possible to test reliability in this study due to time constraints, which is another key criteria of a measure. This is something that will be important to test in the future.

7.5 Conclusions

Generally the descriptive system performed well in this population, with excellent response and completions rates. It also has a low mean time to complete and this decreases with age as you would expect. This mean time is also much lower than most other paediatric generic measures. The percentage of children able to self complete was high, and the majority of children were able to read and understand the instructions, however it should be noted there was a researcher present who was able to read any words or clarify meanings for them. The practicality is therefore very good but for some children it will be necessary to have a researcher present. The descriptive system demonstrated good content validity in that only 10% of children felt there was something missing from the content and the majority of the items identified were symptoms which had arisen in the previous interview work anyway and should feed into the existing dimensions. The distribution across the levels was also appropriate for this type of population. The descriptive system demonstrated reasonable face validity in that most children were able to read and understand

the instructions and the questions, however there are some areas of the questionnaire which could be improved in order to increase this, such as reducing the number of levels and making it clearer how to answer a question. Good construct validity was demonstrated as the descriptive system showed differences in the distributions between those who were off school in the week prior to completion, compared to those who were not and there was a difference in mean scores. The score also correlated well with the self report rating of health. The questions identified as the same were as expected and the strongest preference was to remove the hurting question and keep the pain question instead.

One of the main challenges is to keep children focused on the fact that they should be thinking about their health when they answer the questions. This may be improved through the use of an example question and when the final descriptive system is reduced, the instructions may appear at the top of the page to aid this.

Overall, this testing phase has demonstrated that in its current form, the descriptive system demonstrates very good practicality and validity and the vast majority of children are able to self complete provided there is an adult on hand to give clarification when required.

Appendix 7A Questionnaire used in the testing

The University Of Sheffield.

Health Questionnaire

These questions ask about how **your health and your health problems** (like a headache or asthma) are affecting you **today.**

For each question, read all the choices and decide which one is most like you **today** because of your health.

Then put a tick in the box next to it like this ☑

Only tick **one** box for each question. There are no right or wrong answers.

Thank you for filling this in!

1. Worrying

- ☐ I don't feel worried
- ☐ I feel a little bit worried
- ☐ I feel a bit worried
- ☐ I feel quite worried
- ☐ I feel very worried

2. Sad

- ☐ I don't feel sad
- ☐ I feel a little bit sad
- ☐ I feel a bit sad
- ☐ I feel quite sad
- ☐ I feel very sad

3. Weak

- ☐ I don't feel weak
- ☐ I feel a little bit weak
- ☐ I feel a bit weak
- ☐ I feel quite weak
- ☐ I feel very weak

4. Angry

- ☐ I don't feel angry
- ☐ I feel a little bit angry
- ☐ I feel a bit angry
- ☐ I feel quite angry
- ☐ I feel very angry

5. Pain

- ☐ I don't have any pain
- ☐ I have a little bit of pain
- ☐ I have a bit of pain
- ☐ I have quite a lot of pain
- ☐ I have a lot of pain
- ☐ I am really in pain

6. Frustrated

- ☐ I don't feel frustrated
- ☐ I feel a little bit frustrated
- ☐ I feel a bit frustrated
- ☐ I feel quite frustrated
- ☐ I feel very frustrated

7. Hurting

- ☐ It doesn't hurt
- ☐ It hurts a little bit
- ☐ It hurts a bit
- ☐ It hurts quite a bit
- ☐ It hurts quite a lot
- ☐ It hurts a lot
- ☐ It really hurts

8. School Work

- ☐ My school work is not affected
- ☐ My school work is a little bit affected
- ☐ My school work is a bit affected
- ☐ My school work is quite affected
- ☐ My school work is affected quite a lot
- ☐ My school work is really affected
- ☐ I can't do my school work

9. Upset

- ☐ I don't feel upset
- ☐ I feel a little bit upset
- ☐ I feel a bit upset
- ☐ I feel quite upset
- ☐ I feel very upset

10. Tired

- ☐ I don't feel tired
- ☐ I feel a little bit tired
- ☐ I feel a bit tired
- ☐ I feel quite tired
- ☐ I feel very tired

11. Annoyed

- ☐ I don't feel annoyed
- ☐ I feel a little bit annoyed
- ☐ I feel a bit annoyed
- ☐ I feel quite annoyed
- ☐ I feel very annoyed

12. Scared

- ☐ I don't feel scared
- ☐ I feel a little bit scared
- ☐ I feel a bit scared
- ☐ I feel quite scared
- ☐ I feel very scared

13. Sleep

- ☐ My sleep is not affected
- ☐ My sleep is a little bit affected
- ☐ My sleep is a bit affected
- ☐ My sleep is quite affected
- ☐ My sleep is affected quite a lot
- ☐ My sleep is affected a lot
- ☐ My sleep is really affected
- ☐ I can't sleep at all

14. Embarrassed

- ☐ I don't feel embarrassed
- ☐ I feel a little bit embarrassed
- ☐ I feel a bit embarrassed
- ☐ I feel quite embarrassed
- ☐ I feel very embarrassed

15. Jealous

- ☐ I don't feel jealous
- ☐ I feel a little bit jealous
- ☐ I feel a bit jealous
- ☐ I feel quite jealous
- ☐ I feel very jealous

16. Daily routine (things like eating, washing, getting dressed, getting ready)

- ☐ I have no problems with my daily routine
- ☐ I have a few problems with my daily routine
- ☐ I have some problems with my daily routine
- ☐ I have many problems with my daily routine
- ☐ I can't do my daily routine

17. Able to join in the activities that you want to (things like playing out with your friends, doing sports, joining in things)

- ☐ I can join in with any of the activities that I want to
- ☐ I can join in with most of the activities that I want to
- ☐ I can join in with some of the activities that I want to
- ☐ I can join in with a few of the activities that I want to
- ☐ I can join in with none of the activities that I want to

18. Overall, how would you rate your health today?

- ☐ Excellent
- ☐ Very good
- ☐ Good
- ☐ Fair
- ☐ Poor

Appendix 7.B: Questionnaire Evaluation

ID	Start time	End time	Total time

Questionnaire Evaluation

C1	Have you been off school poorly or ill in the last week?	Y☐ N☐	
C2	Was there anything else about how your health affects you that was missing?	Y☐ N☐	
C3	Were there any questions you thought were the same? Which ones?	Y☐ N☐	
C4	If so, which wording do you find easier?		
C5	Were there any questions you didn't understand? Which ones?	Y☐ N☐	
C6	Were there any questions you found difficult? Which ones? Why?	Y☐ N☐	
C7	Have you got any other comments about the questionnaire?	Y☐ N☐	

K1	Did they read the instructions themselves?	Y☐ N☐	
K2	Did they understand the instructions If no, what did they need help on?	Y☐ N☐	
K3	Did they read all the questions themselves If no, which had to be read to them	Y☐ N☐	
K4	Did they need help on particular questions? If yes, which ones What was the nature of the help	Y☐ N☐	
K5	Were they able to self complete?	Y☐ N☐	

Appendix 7.C

Table 7.2: Mean time taken to complete

	Mean (minutes)	Median (minutes)	Minimum (minutes)	Maximum (minutes)
All cases	3.83	3	1	12
Y3	4.89	5.00	2	10
Y4	4.09	3.00	2	12
Y5	3.44	3.00	1	11
Y6	3.07	3.00	2	7

Table 7.3: Response rate by year group

	n	%
Total	150	98
Y3	35	100
Y4	34	92
Y5	45	100
Y6	37	100

Table 7.5: Completion rate by question, Y3 (7/8 years)

Question	Responses (n)	Missing (n)	%missing
q1 (Worrying)	35	0	0
q2 (Sad)	35	0	0
q3 (Weak)	35	0	0
q4 (Angry)	35	0	0
q5 (Pain)	34	1	2.9
q7 (Frustrated)	35	0	0
q7 (Hurting)	35	0	0
q8 (School Work)	35	0	0
q9 (Upset)	34	1	2.9
q10 (Tired)	34	1	2.9

q11 (Annoyed)	33	2	5.7
q12 (Scared)	34	1	2.9
q13 (Sleep)	34	1	2.9
q14 (Embarrassed)	34	1	2.9
q15 (Jealous)	34	1	2.9
q17 (Daily Routine)	34	1	2.9
q17 (Joining in activities)	34	1	2.9
q18 (Rating health)	34	1	2.9

Distribution across the response options

Tables 7.7

Q1 Worrying

Level	n	%
1	93	72
2	40	27.7
3	10	7.7
4	5	3.3
5	2	1.3

Table 7.8

Q2 Sad

Level	n	%
1	111	74
2	21	14
3	10	7.7
4	7	4.7
5	1	0.7

Table 7.9

Q3 Weak

Level	n	%
1	103	78.7
2	31	20.7
3	9	7
4	5	3.3
5	2	1.3

Table 7.10

Q4 Angry

Level	n	%
1	119	79.3
2	17	10.7
3	9	7
4	3	2
5	3	2

Table 7.11

Q5 Pain

Level	n	%
1	83	55.7
2	42	28.2
3	18	12.1
4	4	2.7
5	0	0
7	2	1.3

Table 7.12

Q7 Frustrated

Level	n	%
1	100	77.7
2	33	22
3	9	7
4	5	3.3
5	3	2

Table 7.13

Q7 Hurting

Level	n	%
1	88	58.7
2	38	25.3
3	9	7
4	7	4
5	3	2
7	3	2
7	3	2

Table 7.14

Q8 School Work

Level	n	%
1	104	79.3
2	31	20.7
3	7	4.7
4	7	4
5	1	0.7
7	1	0.7
7	0	0

Table 7.15

Q9 Upset

Level	n	%
1	111	74.5
2	25	17.8
3	7	4
4	7	4
5	1	0.7
Missing	1	-

Table 7.16

Q10 Tired

Level	n	%
1	45	30.2
2	78	45.7
3	17	10.7
4	12	8.1
5	8	5.4
Missing	1	-

Table 7.17

Q11 Annoyed

Level	n	%
1	99	77.9
2	34	23
3	7	4.1
4	7	4.7
5	2	1.4
Missing	2	-

Table 7.18

Q12 Scared

Level	n	%
1	127	84.7
2	14	9.4
3	7	4.7
4	1	0.7
5	1	0.7
Missing	1	-

Table 7.19

Q13 Sleep

Level	n	%
1	81	54.4
2	37	24.8
3	9	7
4	8	5.4
5	7	4.7
7	1	0.7
7	3	2
8	3	2
Missing	1	-

Table 7.20

Q14 Embarrassed

Level	n	%
1	119	79.9
2	24	17.1
3	3	2
4	0	2
5	3	2

Table 7.21

Q15 Jealous

Level	n	%
1	129	87.8
2	15	10.1
3	3	2
4	0	0
5	2	1.3

Table 7.22

Q17 Daily Routine

Level	n	%
1	117	78.5
2	24	17.1
3	7	4
4	0	0
5	2	1.3
Missing	1	-

Table 7.23

Q17 Joining in activities

Level	n	%
1	107	71.8
2	30	20.1
3	8	5.4
4	2	1.3
5	2	1.3
Missing	1	-

Table 7.24

Q18 Rating of health

Level	n	%
1	52	34.9
2	74	43
3	31	20.8
4	2	1.3
5	0	0
Missing	1	-

Distributions across levels for those who were and were not off school in the last week.

Table 7.34

Q1 Worrying

Level	Off school in the last week		Not off school in the last week	
	n	%	n	%
1	18	74.3	74	71.2
2	5	17.9	35	28.9
3	3	10.7	7	5.8
4	2	7.1	3	2.5
5	0	0	2	1.7

Table 7.35

Q2 Sad

Level	Off school in the last week		Not off school in the last week	
	n	%	n	%
1	15	53.7	95	78.5

2	7	21.4	15	12.4
3	4	14.3	7	5
4	3	10.7	4	3.3
5	0	0	1	0.8

Table 7.36

Q3 Weak

Level	**Off school in the last week**		**Not off school in the last week**	
	n	%	n	%
1	18	74.3	84	79.4
2	5	17.9	27	21.5
3	4	14.3	5	4.1
4	1	3.7	4	3.3
5	0	0	2	1.7

Table 7.37

Q4 Angry

Level	**Off school in the last week**		**Not off school in the last week**	
	n	%	n	%
1	20	71.4	98	81
2	4	14.3	12	9.9
3	1	3.7	8	7.7
4	2	7.1	1	0.8
5	1	3.7	2	1.7

Table 7.38

Q5 Pain

Level	Off school in the last week		Not off school in the last week	
	n	%	n	%
1	12	42.9	70	58.3
2	7	25	35	29.2
3	7	25	11	9.2
4	2	7.1	2	1.7
5	0	0	0	0
7	0	0	2	1.7
Missing	-	-	1	-

Table 7.39

Q7 Frustrated

Level	Off school in the last week		Not off school in the last week	
	n	%	n	%
1	18	74.3	82	77.8
2	7	21.4	27	21.5
3	1	3.7	8	7.7
4	2	7.1	3	2.5
5	1	3.7	2	1.7

Table 7.40

Q7 Hurting

Level	Off school in the last week		Not off school in the last week	
	n	%	n	%
1	17	57.1	71	58.7
2	8	28.7	30	24.8

3	1	3.7	8	7.7
4	1	3.7	5	4.1
5	1	3.7	2	1.7
7	1	3.7	2	1.7
7	0	0	3	2.5

Table 7.41

Q8 School Work

Level	Off school in the last week		Not off school in the last week	
	n	%	n	%
1	17	57.1	88	72.7
2	7	25	24	19.8
3	4	14.3	2	1.7
4	1	3.7	5	4.1
5	0	0	1	0.8
7	0	0	1	0.8
7	0	0	0	0

Table 7.42

Q9 Upset

Level	Off school in the last week		Not off school in the last week	
	n	%	n	%
1	17	70.7	94	77.7
2	7	25	18	14.9
3	2	7.1	4	3.3
4	2	7.1	4	3.3
5	0	0	1	0.8

Table 7.43

Q10 Tired

Level	Off school in the last week		Not off school in the last week	
	n	%	n	%
1	7	25	38	31.4
2	10	35.7	58	47.9
3	4	14.3	12	9.9
4	4	14.3	8	7.7
5	3	10.7	5	4.1

Table 7.44

Q11 Annoyed

Level	Off school in the last week		Not off school in the last week	
	n	%	n	%
1	17	57.1	83	78.7
2	7	25	27	22.3
3	0	10	7	5
4	5	17.9	2	1.7
5	0	0	2	1.7

Table 7.45

Q12 Scared

Level	Off school in the last week		Not off school in the last week	
	n	%	n	%
1	21	72	105	87.8
2	4	14.3	10	8.3
3	2	7.1	50	4.1
4	1	3.7	0	0

5	0	0	1	0.8

Table 7.46

Q13 Sleep

Level	Off school in the last week		Not off school in the last week	
	n	%	n	%
1	10	35.7	71	58.7
2	8	28.7	29	24
3	0	0	9	7.4
4	3	10.7	5	4.1
5	2	7.1	5	4.1
7	0	0	1	0.8
7	3	10.7	0	0
8	2	7.1	1	0.8

Table 7.47

Q14 Embarrassed

Level	Off school in the last week		Not off school in the last week	
	n	%	n	%
1	20	71.4	99	81.8
2	5	17.9	19	15.7
3	2	7.1	1	0.8
4	0	0	0	0
5	1	3.7	2	1.7

Table 7.48

Q15 Jealous

Level	Off school in the last week		Not off school in the last week	
	n	%	n	%
1	23	82.1	107	87.7
2	3	10.7	12	9.9
3	2	7.1	1	0.8
4	0	0	0	0
5	0	0	2	1.7

Table 7.49

Q17 Daily Routine

Level	Off school in the last week		Not off school in the last week	
	n	%	n	%
1	23	82.1	94	77.7
2	4	14.3	20	17.5
3	1	3.7	5	4.1
4	0	0	0	0
5	0	0	2	1.7

Table 7.50

Q17 Joining in activities

Level	Off school in the last week		Not off school in the last week	
	n	%	n	%
1	17	70.7	90	74.4
2	7	25	23	19
3	3	10.7	5	4.1
4	1	3.7	1	0.8

5	0	0	2	1.7

Table 7.51

Q18 Rating of health

Level	Off school in the last week		Not off school in the last week	
	n	%	n	%
1	7	21.4	47	38
2	14	50	50	41.3
3	7	25	24	19.8
4	1	3.7	1	0.8
5	0	0	0	0

Table 7.56: Which questions were judged to be the same and which wording was preferred

Combination	N stating they were the same	Wording preferred (N)	Comments
Worrying	9	3	scared is confusing
Scared		7	but close
Worrying	7	2	
Sad		0	
Keep both		5	
Worrying	1	1	
Frustrated		0	
Worrying	1	1	
Hurting		0	

Worrying	1	0	
Upset		0	
Keep both		1	
Tired	1	1	
Annoyed		0	
Tired	15	3	
Sleep		0	
Keep both		11	
Not sure		1	
Tired	7	3	
Weak		2	
Keep both		1	
Frustrated	1	0	
Tired		0	
Keep both		1	
Annoyed	4	0	
Jealous		1	
Keep both		3	
Embarrassed	3	1	
Jealous		2	
Sad	1	0	
Weak		1	

Sad	1	0	
Frustrated		0	
Keep both		1	
Sad	42	20	upset has a stronger meaning than sad, sad could be patronizing
Upset		20	
Keep both		2	
Weak	1	0	
Scared		1	
Weak	1	0	
Sleep		1	because when weak, want to go to sleep
Weak	1	0	
Pain		0	
Keep both		1	
Angry	13	4	
Annoyed		7	
Keep both		2	
Angry	7	0	
Annoyed		2	
Frustrated		3	
Keep 4 and 7		1	
Angry	1	0	
Pain		0	

Frustrated		0	
Keep all		1	
Angry	7	3	
Frustrated		2	
Keep both		1	
Not sure		1	
Pain	39	25	but make it easier by saying e.g. pain in your body, more obvious, but hurt is the better word it is the better way of putting it
Hurting		12	pain more of a formal word, easier for younger children
Different		1	hurt is continuous, pain is intermittent
Not sure		1	
Frustrated	27	7	
Annoyed		13	
Keep both		4	
Not sure		2	
Frustrated	2	0	
Scared		0	because when weak want to go to sleep
Keep both		2	
Frustrated	1	0	
Jealous		0	
Keep both		1	

Hurting	1	0	
Jealous		0	
Keep both		1	
School Work	1	0	
Daily Routine		0	
Keep both		1	
Upset	2	0	
Scared		2	

Table 7.57: Other comments by children

Question 1, they are worried because of their weight
Question 18 - fair is a funny word to use for children
For question 18 ask children to put why they feel excellent, or good etc, about their health. Add - here are some things you might be feeling today before the questions
The boxes are small, the hurting question is a bit abstract/not specific - what hurts? Put the thank you at the end. Put the writing in colour
can't do school work - commented that you wouldn't be in school if this was the case.
Make the writing a bit bigger
It covers everything
Enjoyed it
Fun, writing the right size
Good to have jealous in as they know that when they have a health problem they feel jealous of others who don't have it
It is a good questionnaire
It's quite good

It's really good
It looks like it is hard and will take ages but it doesn't when you do it
Make it clearer how to answer the questions
It's not necessary to have so many levels, e.g. some and few, a bit and a little bit
Noticed levels are repetitive
Questionnaire is quite good because gives you a chance to say how you feel
Questions must have the explanation - i.e. q17
Quite fun
School work may be part of daily routine
Too many levels - a little bit and a bit are the same sort of thing
Very good. Think should include how children feel about things at home
Was good, liked it, length is O.K.

References

Brazier, J.E., Deverill, M., Green, C., Harper, R. & Booth, A. 1999. A review of the use of health status measures in economic evaluation. Health Technology Assessment, 3(9).

Hays, R. & Revicki, D. 2005. Reliability and Validity (including responsiveness). In Fayers, P. & Hays, R. (Ed.). Assessing Quality of Life in Clinical Trials. 2nd edition. Oxford University Press.

Health Utilities Index http://healthutilities.biz/
Accessed 10/07/2008

Matza, L. S., Swensen, A. R., Flood, E. M., Secnik, K. & Leidy, N. K. 2004. Assessment of Health Related Quality of Life in Children: A Review of Conceptual, Methodological, and Regulatory Issues. Value in Health. 7(1) 79-92.

National Statistics.
http://www.statistics.gov.uk/StatBase/Product.asp?vlnk=14729
Accessed 05/12/07

Testing the Descriptive System in a Clinical Paediatric Population

8a.1 Introduction

This chapter reports on the testing of the questionnaire in a clinical paediatric population as an addition to testing the questionnaire in a general paediatric population. It was important to test it in a patient population, where children with a much wider range and severity of health problems could be included. The aim was to test some of the psychometric properties, to determine whether children are able to self complete and to test the alternative wordings for dimensions. In addition the aim was to compare this to the performance in the general population and use the results of both studies to refine the descriptive system as necessary.

8a.2 Methods

8a.2.1 Overview

A paediatric clinical population was recruited from Sheffield Children's Hospital to test the questionnaire in a sicker population. Children were recruited from both surgical and medical wards, including day care patients. Once consent had been obtained from the parent/guardian, children were asked for their assent and asked to complete the questionnaire. They were then asked a series of questions about the questionnaire to help evaluate it and the researcher recorded details about how long they had taken to complete it and whether there were any problems or difficulties. Basic data was collected from parents/guardians on whether children had had previous hospital admissions and whether children had a long standing illness, to help test the validity of the questionnaire. The results were also compared to those obtained in Chapter 7, testing the questionnaire on a general paediatric population. Children were recruited between 03/09/2007 and 02/11/2007.

8a.2.2 Recruitment and sample

This work was carried out in collaboration with the Clinical Research Facility at Sheffield Children's NHS Foundation Trust in order to access a paediatric clinical population. Several different care areas within Sheffield Children's Hospital were recruited to the study so as to access children with a wide range of acute and chronic health problems. These were a medical ward, two surgical wards, the day care unit and the Clinical Research Facility (an outpatient research facility).

The medical ward cares for children age 4-16 years with a wide range of acute and chronic medical conditions. One surgical ward covers children with a range of renal, gastrointestinal and neuro-surgery, and the other surgical ward covers orthopaedic, limb reconstruction, and spinal surgery. The day care unit cares for children undertaking surgical and medical day procedures (including urology, gastroenterology, endocrinology, neurology, oncology, orthopaedic, general surgery, dental, ENT (ear nose and throat) and allergy patients as well as patients coming for investigational procedures such scopes, MRI and CT scans and blood tests) and the Clinical Research Facility cares for a range of children involved in research projects, including those with brittle bone disease. Children attending the hospital come from throughout South Yorkshire and some come from slightly further afield such as Lincolnshire.

Children were eligible for the study if they were being cared for on one of the recruited areas and they were age 7-11 years. Children whose parents did not speak English had to be excluded as they were not able to understand the information leaflet or consent process and resources were not available for translation. In addition any children that the medical team thought were unsuitable to be approached, for example if they were too ill, were not approached.

Parents of eligible children were first approached by a member of the clinical team responsible for the child's care, who gave the parents/guardians an

information leaflet for themselves and also for their child. The parent/guardian could then pass the leaflet to the child and if they were interested, they contacted the researcher, who was present in the hospital. They were given an opportunity to ask any questions they wished and consent was taken if they were happy to give consent for their child to be asked if they would like to take part. Once the parent/guardian had consented, the child was approached by the researcher and given an explanation of the study and the opportunity to ask any questions they wished. If the child was happy to take part, they completed an assent form by writing their name and the date. The child was then given the health questionnaire to complete, with the researcher on hand at all times in case any help was required. The researcher completed a questionnaire evaluation form (the same as the one used in the testing on the general population, apart from the question asking if they had been absent from school as this did not make sense in this setting) after they had finished. In addition, a data collection form was completed by the parents/guardians which asked for information about previous hospital stays of the child, both inpatient and outpatient, whether their child has any long standing illness or disability, whether this affects their activities and a rating of their child's health (Appendix 8a.A). This data was obtained in order to be able to get a rough proxy for how sick children were, so that they could be split into groups to test if the descriptive system can differentiate between the groups. Those who have had previous hospital stays or longstanding illness or disabilities could be hypothesised to have more serious problems and so to have lower scores. The question on long standing illness/disability was taken from the British General Household Survey (National Statistics a) and is a standard question, usually used in adults but adapted here as a proxy version for parents/guardians to answer for their children. Information was also collected on gender, age and ethnicity.

Ethics approval was obtained from South Sheffield Research Ethics Committee and R&D Approval from Sheffield Children's NHS Foundation Trust. The study was carried out between 03/09/07 and 02/11/07.

8a.2.3 Analysis

In a similar manner to the testing on the general population, two key psychometric criteria – practicality and validity were examined in this study. In addition, inter rater reliability was tested by comparing parent and child responses to question 18 (general rating of health). However, this question is not part of the descriptive system itself. It was not possible to look at reliability or responsiveness due to time constraints.

The practicality of the draft descriptive system was tested by looking at the mean time taken to complete, the response rate, the completion rate and whether the child read and understood the instructions.

Content validity was tested by looking at whether children identified things missing from the descriptive system. Face validity was tested by looking to see if children were able to read and understand the questions and looking at what they found difficult or needed help on and why. Construct validity was tested by comparing the distribution of the responses across the levels of the clinical population with the general paediatric population from chapter 7, using a Mann Whitney non parametric test. Construct validity was also examined by scoring the descriptive system and seeing if this correlated with the child's rating of their health. The descriptive system was scored in exactly the same manner as for the general population, that is, assuming that the questions can be scored so that being at the highest level is 1 and the next level down is 2, and generating a score for each patient by summing the levels across the eleven chosen questions. The score was also calculated for those who had had a previous hospital stay in the last 6 months, those who did not, those who had had a previous outpatient appointment in the last 6 months, those who did not, those who had a long standing illness that limits their activities and those who did not. These scores were tested to see if the difference was significant. A final test of construct validity was to look at the differences in mean score between the

general population (in chapter 7) and the clinical population. Again, this difference was tested to see if it was significant.

Inter rater reliability was tested by comparing the difference in parent and child responses to question 18 (general rating of health). The difference was calculated as the rating by the parent minus the rating by the child, therefore a difference of 0 means agreement, a positive difference means the parent rates the child's health higher and a negative difference means the parent rates the child's health lower. This question is not part of the descriptive system and the recall period was over the last week for parents and today for the children, therefore the recall period was not the same. However, it is still useful to look at the amount of agreement between parents and children.

Whether children were able to self complete was assessed by looking at the proportion that were able to self complete.

Finally, questions identified as the same by children were examined to see if there was a substantial preference for one form of wording.

8a.3 Results

8a.3.1 Recruitment and sample

154 child admissions were identified as eligible, 98 consented, 56 did not consent and 1 child was withdrawn (upon discovery the child was 6 years old and not 7). Four of the child admissions were the same two children consenting twice but in a different location on a different date, therefore data on 97 child admissions from 95 children was analysed. Table 8a.1 shows the number of children consenting and not consenting by ward area.

Table 8a.1: Consent by ward area

	DC n(%)	S1 n(%)	S3 n(%)	M2 n(%)	CRF n(%)	Total n(%)
Approached	89	17	19	26	3	**154**
Consented	52 (58)	13(76)	16 (84)	15 (58)	2 (67)	**98 (64)**

DC: Day Care

S1: Surgical Ward 1

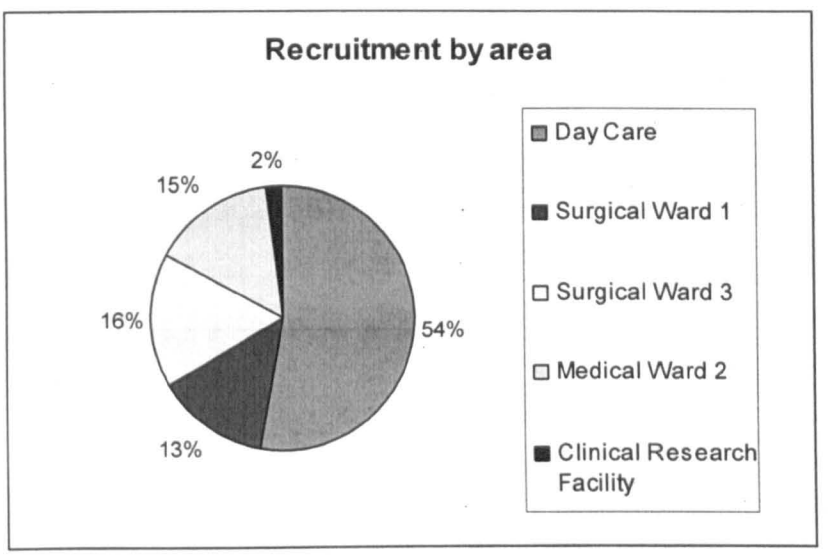
S2: Surgical Ward 3

M2: Medical Ward 2

CRF: Clinical Research Facility

Figure 8a.1 shows the number of patients recruited in each area.

Figure 8a.1: Recruitment by area



The characteristics of the sample who participated are shown in Table 8a.2 and are compared with the characteristics of 7-11 year old children of Great Britain according to the 2001 census for comparability (National Statistics b).

Table 8a.2: Characteristics of the sample

	N	%	Census data for Great Britain %
Male	57	58.8	51.21
7 years old	18	18.8	19.41
8 years old	17	17.7	19.62
9 years old	26	21.7	20.33
10 years old	13	13.5	20.55
11 years old	22	22.9	20.09
Age missing	1	-	-
Excellent health (parent rating)	39	40.2	-
Very good health (parent rating)	25	25.8	-
Good health (parent rating)	14	14.4	-
Fair health (parent rating)	5	5.2	-
Poor health (parent rating)	13	13.4	-
Other*	1	1	-
White	85	87.6	88.55
Mixed/dual heritage	5	5.2	2.75
Asian or Asian British	3	3.1	5.48
Black or Black British	1	1	2.48
Chinese	0	0	0.36
Other	3	3	0.37
Longstanding illness that doesn't limit activities	24	24.74	-
Longstanding illness that limits activities	39	40.21	-
Previous hospital stay	18	18.6	-
Previous outpatient visit	55	56.7	-

*parent rated child's health as 1 and 5 as it was excellent before they came in and has been poor for the last 3 days.

Table 8a.3 details the number of previous hospital and outpatient visits by patients. 15 children had had both a previous hospital stay and a previous outpatient visit.

Table 8a.3: Previous hospital and outpatient visits

	Previous hospital stay (number of times)	Previous outpatient visit (number of times)
Mean	2.22	4
Median	1	2
Min	1	1
Max	10	32

Tables 8a.4, 8a.5 and 8a.6 show the child's own rating of their health (question 18) broken down according to whether they had had a previous hospital stay in the last 6 months, a previous outpatient visit in the last 6 months and whether they have a long standing illness (broken down by whether it affects them or not). Those who had not had a previous hospital stay were more likely to rate their health as excellent or very good, however there was not much difference for the previous outpatient group.

Table 8a.4: Rating of health (previous hospital stay)

Previous hospital stay in last 6 months	N excellent (%)	N Very Good (%)	N Good (%)	N Fair (%)	N Poor (%)	N Total
No	21(26.6)	28(35.4)	19(24.1)	10(12.7)	1(1.3)	79
Yes	4(22.2)	4(22.2)	4(22.2)	4(22.2)	2(11.1)	18

Table 8a.5: Rating of health (previous outpatient visit)

Previous outpatient visit in last 6 months	N excellent (%)	N Very Good (%)	N Good (%)	N Fair (%)	N Poor (%)	N Total
No	11(26.2)	15(35.7)	9(21.4)	6(14.3)	1(2.4)	42
Yes	14(25.5)	17(30.9)	14(25.5)	8(14.5)	2(3.6)	55

Table 8a.6: Rating of health (long standing illness)

Long standing illness	N excellent (%)	N Very Good (%)	N Good (%)	N Fair (%)	N Poor (%)	N Total
No	8(23.5)	13(38.2)	7(20.6)	5(14.7)	1(2.9)	34
Yes (that doesn't affect them)	4(16.7)	9(37.5)	8(33.3)	2(8.3)	1(4.2)	24
Yes (that does affect them)	13(33.3)	10(25.6)	8(20.5)	7(17.9)	1(2.6)	39

8a.3.2 Practicality

Out of 154 patients identified, 98 consented giving a response rate of 64%. No information was collected about the characteristics of the non participating sample for ethical reasons.

The time taken to complete the questionnaire was low with a mean of 5.27 minutes overall and ranged from 3.59 minutes to 8.06 minutes across year group, decreasing with age as might be expected. The mean, median, minimum

and maximum time broken down by year group is shown in Table 8a.7 in Appendix 8a.B.

The completion rate by question is shown in Table 8a.8 below. There was very little missing data. Question 3 was missing by a 10 year old, question 6 a 7 year old and question 8, a 9 year old.

Table 8a.8: Completion rate by question, all cases

Question	Responses (n)	Missing (n)	Missing (%)
q1 (Worrying)	97	0	0
q2 (Sad)	97	0	0
q3 (Weak)	96	1	1
q4 (Angry)	97	0	0
q5 (Pain)	97	0	0
q6 (Frustrated)	96	1	1
q7 (Hurting)	97	0	0
q8 (School Work)	96	1	1
q9 (Upset)	97	0	0
q10 (Tired)	97	0	0
q11 (Annoyed)	97	0	0
q12 (Scared)	97	0	0
q13 (Sleep)	97	0	0
q14 (Embarrassed)	97	0	0
q15 (Jealous)	97	0	0
q16 (Daily Routine)	97	0	0
q17 (Joining in activities)	97	0	0
q18 (Rating health)	97	0	0

Table 8a.9 below shows how many children read the instructions themselves and is also shown by age group. Overall this was around half (51%) which is a lot lower than the school population. This generally increased with age, although the 9 year old children were the highest (68%).

Table 8a.9: Reading the instructions

Year	Read the instructions themselves	N	%
All cases	yes	49	51
7 years old	yes	4	22.2
8 years old	yes	7	41.2
9 years old	yes	17	68
10 years old	yes	7	53.8
11 years old	yes	14	63.6

Table 8a.10 below shows how many children were able to understand the instructions and is also shown by age group. This was high overall (94.8%) and increased with age apart from the 10 year old children which was a little lower at 92.3%.

Table 8a.10: Able to understand the instructions

Year	Able to understand the instructions	N	%
All cases	yes	92	94.8
7 years old	yes	17	94.4
8 years old	yes	16	94.1
9 years old	yes	24	96
10 years old	yes	12	92.3
11 years old	yes	22	100

8a.3.3 Content Validity

The percentage of children saying something was missing from the content of the questionnaire was fairly low (around 11%, varying from 5 to 27% by age group). Some children confused the questionnaire with notions of being healthy, for example they suggested including questions on eating healthy food and a child's weight. Other children suggested more daily functions, such as eating,

going to the toilet, getting in and out of bed and lifting things. Other suggestions were symptoms such as itching and breathing problems. Most of the daily functions should come under the daily routine question and the symptoms mentioned arose in the original interview work and tapped into the *pain, annoyed, sleep* and *worried* and *activities* dimensions respectively. Table 8a.11 below details what was thought missing from the content of the descriptive system.

Table 8a.11: What children identified as missing from the content of the descriptive system (n=1 for each item)

do people feel jealous of you
do you eat healthy food
eating, going to the toilet
feeling hungry (note- child was on NBM (nil by mouth) at the time)
get bored
getting in and out of bed, sitting up
itching - when have reaction to nuts
lifting heavy things
P.E. - breathing problems
weight - eating

8a.3.4 Face Validity

Table 8a.12 below shows the number and percentage of children who read the questions themselves (this was observed and recorded by the researcher). It also shows the breakdown by each year group. The proportion of patients who read the questions themselves was about half on average, but varied from 11% for 7 year old children to 76% for 9 year old children. For similar reasons to reading the instructions this was not necessarily because the children could not read, in most cases the children were too poorly or tired and asked for

someone to read the questions to them. Some children could not sit up very well and so they asked for the questions to be read to them.

Table 8a.12: Reading the questions

Year	Read all the questions themselves	N	%
All cases	yes	49	51.0
7 years old	yes	2	11.1
8 years old	yes	8	47.1
9 years old	yes	19	76.0
10 years old	yes	6	46.2
11 years old	yes	14	63.6

One child (aged 7) asked for question 6 to be read to them. 1 child (age 8) had the first three questions read to them and then read the rest themselves. 2 children (aged 10 and 11) had the first 4 questions read to them and then read the rest themselves. 41 children had all 18 questions read to them.

Table 8a.13 details whether help was needed on particular questions or not and is broken down by age group. Overall this was fairly high at 60.4%, but was lower for the oldest children at 40.9%.

Table 8a.13: Help needed on questions

Year	Help needed on questions	N	%
All cases	yes	58	60.4
7 years old	yes	13	72.2
8 years old	yes	12	70.6
9 years old	yes	15	60
10 years old	yes	8	61.5
11 years old	yes	9	40.9

About 60% of children needed some help with questions. This generally decreased by age. The main areas children needed help on were meanings of words, including *weak, frustrated, embarrassed* and *affected* (in both the context of sleep and schoolwork). They also needed more explanation for some questions, including which part of you hurts (q7), why you might feel jealous (q15) and more explanation on daily routine (q16). Some children were confused about schoolwork as they were in hospital at the time and asked how they should answer they question or put that they could not do their schoolwork. Some children also needed help deciding which level they were on q18. The nature of this help is detailed in table 8a.14 below and is listed for each question.

Table 8a.14: Nature of help required

Question	Nature of help (N)
1 (Worrying)	worrying - why would you be worried.
3 (Weak)	weak - checked what level they were with Mum - had reasoned it out themselves though.
	weak what it meant (8)
	weak what it said
6 (Frustrated)	Frustrated - meaning of word (14)
	frustrated - what it said and meant.
	what frustrated meant - explained and understood and reasoned out well.
	Frustrated - checked meaning , had got it slightly wrong, but understood on explanation and reasoned well reading and explanation.
7 (Hurting)	what hurting? Which part? (3) Hurt - checked with Mum
8 (School Work)	what affected meant. (9) school work - how affected. schoolwork - checked, did some today in hospital and could

	manage fine. Was confused because not at school. school work - wasn't sure what level - asked Mum - Mum reminded them they knew how their schoolwork was. School work - bit more explanation checked with Mum which answer to put. (2) Schoolwork - is half term at the moment. Affected - what said and meant. Can't do as in hospital (5)
11 (Annoyed)	what annoyed meant annoyed, meaning and why you would feel annoyed.
12 (Scared)	Is now post op and not scared but was a bit, so put what was this morning pre op
13 (Sleep)	help from Mum on level. (2) reminded them it was last night. (2) what affected meant (5) sleep couldn't remember last night, started thinking more generally
14 (Embarrassed)	what embarrassed meant (4) what embarrassed said, knew meaning (2) Embarrassed - said and meant.
15 (Jealous)	jealous, needed explanation (4) jealous - said and meant Jealous - why is it in there? (2)
16 (Daily Routine)	what it meant, understood on explanation. length was harder daily routine - which level Daily routine, what it meant (3) said and meant needed extra explanation
17 (Join in activities)	what was about clarified it was for today as on Day Care not sure if could

18 (Rating Health)	which level they were, and what fair meant what rate meant which level (5) didn't know what level (2) fair - what it meant. understanding
General	KS helped generally Patient can't speak, has cerebral palsy, can do yes and no, - Dad helped to complete Dad helped them choose on some of them general explanations from KS generally needed prompting. some explanation and help to keep focused on questions. some help from KS choosing levels some needed explaining in simpler terms.

Approximately 25% of children had question(s) that they did not understand. This generally decreased with age apart from the 7 year olds (lower than the 8 year olds) and the 10 year olds where no one had any questions they did not understand. This is shown in Table 8a.15 below broken down by year group.

Table 8a.15: Questions not understood

Year	Questions not understood	Frequency	%
All cases	Yes	23	25
7 years old	Yes	5	27.8
8 years old	Yes	6	35.3
9 years old	Yes	7	28
10 years old	Yes	0	0
11 years old	Yes	5	25

Table 8a.16 shows what proportion of children thought questions were difficult and is broken down by age group.

Table 8a.16: Questions found difficult

Year	Questions found difficult	N	%
All cases	Yes	23	25
7 years old	Yes	6	33
8 years old	Yes	5	29.4
9 years old	Yes	9	36
10 years old	Yes	2	15.4
11 years old	Yes	3	15

About 25% of children said they found one or more questions difficult. This varied from about 30% for the younger years to 15% for the older years. This was in addition to any notes made by the researcher (KS) as to whether they had needed any help. Most often the difficulty was in deciding which level to choose. Children were still able to answer the questions, but said it was difficult to choose. Table 8a.17 details which questions were found difficult and why.

Table 8a.17: Which questions were found difficult and why

Question	What was difficult
5 (Pain)	not sure which level
7 (Hurting)	wasn't sure what level
8 (School work)	schoolwork - thinking of level
10 (Tired)	tired - when are you supposed to be thinking about - time of day

14 (Embarrassed)	hard to say if you feel embarrassed or not
16 (Daily Routine)	didn't know which one to put what routine means which level to choose wasn't sure which level, reading them, took a bit longer to get
17 (Join in activities)	reading them, took a bit longer to get as on day care, not sure if could or not forgot was about today hard to choose
18 (Rating health)	not sure what level which level they are what level couldn't choose between good and fair

8a.3.5 Construct Validity

Across the distribution of response options, the hospital (more sick) population generally had a greater proportion of people at the lower levels and a lower proportion at the higher levels than the school population, especially for the pain, school work, daily routine and activities questions. The exceptions were q4 (angry) and q14 (embarrassed) where they were fairly similar and q11 (annoyed) where it was mixed across the levels who was higher and who was lower. The distribution of responses is shown for each question in tables 8a.18 to 8a.35 in Appendix 8a.B. It is shown for the hospital population and school population and for both these populations combined. The results of the Mann Whitney test for each question showed significant differences for q1 (worry), q3 (weak), q5 (pain), q7 (hurt), q8 (school work), q12 (scared), q16 (daily routine), q17 (activities) and q18 (rating of health). These were significant at p=0.05.

Table 8a.36 below shows the score by category of patient. Although there are small differences in the direction expected, none of these differences are significant which may be because the sample sizes were too small to detect any difference.

Table 8a.36: Score by category of patient

Category	Mean	Median	Min	Max	n
No previous hospital admission	20.65	18.5	11	43	78
Previous hospital admission	22.67	18	11	47	18
No previous outpatient visit	20.20	19	11	41	41
Previous outpatient visit	21.65	18	11	47	55
No long standing illness	21.33	19	11	44	33
Long standing illness that affects activities	22.15	20	11	47	39

Table 8a.37 below shows the mean, median, minimum and maximum score for each rating of health. The score increases as the level of health decreases which demonstrates that the questionnaire can distinguish between different levels of health and therefore gives some evidence of construct validity.

Table 8a.37: Score by rating of health

	Mean	Median	Min	Max	n
1	16.84	15	11	31	25
2	18.63	18	11	31	32
3	21.13	19	11	41	21
4	31	30	16	44	13
5	37.67	42	24	47	3

The mean score for the general population was 17.03 and for the clinical population was 21.03. This difference was significant at p=0.01 indicating that the measure can discriminate between these two groups.

8a3.6 Reliability

Table 8a.38 shows the difference between child and parent ratings of health. 32.99% of parents and children agreed with their ratings. 40.21 of parents rated their child's health lower than the child and 26.80% of parents rated their child's health higher than the child. It should be noted that the recall period is different, which limits the reliability of the test somewhat as discussed in chapter 7.

Table 8a.38: Difference between child and parent ratings of health

Difference in rating	N	(%)
-3	5	5.15
-2	9	9.28
-1	25	25.77
0	32	32.99
1	13	13.40
2	7	7.22
3	3	3.09
4	3	3.09

Table 8a.39 details what percentage of children were able to self complete, in that they were able to decide what level they were and tick the appropriate box on the questionnaire. It is broken down by age group. It is high overall at 85.4%, but does not really show a pattern of increasing with age. The most common reason for not being able to self complete was the child's medical condition rather than their ability. For example, if both arms were in plaster, or they had a drip in the hand they wrote with, they could not self complete.

Table 8a.39: Able to self complete

Year	Able to self complete	N	%
All cases	yes	82	85.4
7 years old	yes	15	83.3
8 years old	yes	16	94.1

9 years old	yes	22	88
10 years old	yes	11	84.6
11 years old	yes	17	77.3

8a.3.7 Wording

Approximately half of children judged one or more questions to be the same as another question. This varied by year group, ranging from 38.9% to 72.7%. Table 8a.40 below shows the break down by year group.

Table 8a.40: Questions the same

Year	Questions the same	N	%
All cases	Yes	48	52.2
7 years old	Yes	7	38.9
8 years old	Yes	8	47.1
9 years old	Yes	16	64
10 years old	Yes	8	72.7
11 years old	Yes	9	45

The largest pair of questions identified as the same was pain and hurting, with the children being fairly evenly split over which wording was preferred. Some were more clear, for example for sad and upset, sad was strongly preferred and for frustrated and annoyed, annoyed was strongly preferred. Several children identified tired and sleep as being the same but on reflection most children said they were not actually the same. The full list of questions identified as the same and comments by the children are given in Table 8a.41 in Appendix 8a.B

There were not many further comments made by children, but those that did comment generally said that they liked the questionnaire and thought it was good. The comments are given in Table 8a.42 below.

Table 8a.42: Other comments by children

Comment	N
boxes are too small	3
good for a wide range of different children	1
is a good idea, find out how can help children	1
it was good, I enjoyed it	1
scared is like shocked	1
think its good	1
was very good	1

8a.3.8 Other Observations by the Researcher

Other observations noted by the researcher were that children sometimes rushed the instructions and had to be reminded to read them. The schoolwork question was often difficult for children to answer when they were in hospital, unless they were staying for a few days and the hospital teacher brought schoolwork for them, as they were not doing any schoolwork and so were confused how to answer this question, although some ticked the last level 'I can't do my schoolwork.'

The question on sleep (q13) would be better in the past tense if it is about last night, as this would really help children to stop thinking about sleep in general and other nights.

Q18 (rating your health) was hard for children as they have difficulty distinguishing between health and healthy, many did not understand the word fair and a lot of children had problems choosing which level they were.

In general there seem to be too many levels for children to distinguish between for most questions, especially those questions with 7 or 8 levels (questions 7, 8 and 13). For adults, the maximum number of levels they can usually cope with is

7 (Streiner 1995) so for children it is likely to be less than this. Some children forgot to keep in mind it was about today and was because of their health.

Rating their child's health in the past week for adults is hard as often they are well, have an operation/procedure, feel worse, then feel slightly better so they end up taking an average. This was often the case in this setting as children had recently been admitted to hospital. There were many comments from parents that making the child's question (q18) about today is better as it is more reliable and children change quite a bit especially in the hospital environment.

Some children appeared to have adapted to a chronic condition as they based their answers or their own experienced scale and were judging themselves today compared to how they felt on other days. This was observed by the children thinking out loud and talking to the researcher about their answers.

8a.4 Discussion

Recruitment to the study was successful and a good balance of medical and surgical patients was achieved. The highest number of patients came from the day care unit which was because they had the highest throughput of patients to sample from. Consent rates varied from 58 to 84% between ward areas and although data was not formally recorded on reasons for non consent, reasons given at the time included the child being too poorly or not feeling up to it, there not being time before the patient was discharged or the child was too anxious about forthcoming surgery. In day care particularly, there was a short time to discharge post operation and patients were generally not up to it then and were keen to go home. Sometimes children were called for procedures or pre surgical assessment early and so did not have a chance to take part. One or two parents could not understand English and so it was not viable to be able to take consent. Although the sample was sicker than the general population in chapter 7, patients who were really sick were not included and therefore the

questionnaire has not been tested on a really sick population, however it would have been unethical to do this, especially as the children are young.

The sample was reasonably balanced across age and gender, however there were more boys than girls. The spread across health rating by parents was skewed, with more children at the higher levels which was what you might expect given the population studied, but it was not as skewed as the general population sampled in schools (chapter 7), hence this clinical population was not in as good health. The ethnic mix was fairly representative of the national population (of Great Britain) with a slight under representation of white children and a slight over representation of mixed race children.

The balance between acute and chronic conditions was good, as about 65% of children had a longstanding illness and in about 62% of these cases it limited their activities. Approximately 18% of children had had a previous hospital stay and just over half of patients had previously visited outpatients in the previous 6 months.

Overall the descriptive system performed well in this clinical population, with a good response rate, (64%) especially given the nature of the sample and the environment in which they were recruited. There were excellent completions rates, with very low numbers of missing data (only three missing items across the whole dataset) although this may be partly due to parents assisting with completion. Although the child would self complete, the parent would watch and point out if they missed a question, although this did not happen often. It also has a low mean time to complete (about 5 minutes), decreasing with age as you would expect. The percentage of children able to self complete was high, and whilst a lot of children had the instructions read to them, nearly all children were able to understand the instructions. There was only one level on a question where it was not used at all by the hospital population and this was level 4 on q15 (I feel quite jealous). This level was also not used by the school population. The practicality was therefore good.

The descriptive system demonstrated good content validity in that only about 11% of children felt there was something missing from the content and what they identified as missing were either specific symptoms, which feed into existing dimensions, or daily activities which could feed into the daily routine question.

The descriptive system demonstrated reasonable face validity in that although about half of the children had one or more questions read to them, most children were able to understand the questions. About a quarter of children overall found one or more questions difficult and most of the difficulty was deciding which level they were on the questions.

Good construct validity was demonstrated as the level of health as rated by the child correlated well with the score for the descriptive system, in that the lower the children rated their health, the higher their score (on average). Whilst this scoring method is very crude and is not based on any preference data, you would still expect to see this correlation as it demonstrates that the descriptive system can discriminate between different levels of health. Similarly, there are differences in scores between patients based on whether they had previous hospital admissions or not and whether they had a long standing illness. Those who had a previous hospital stay, outpatient visit or had a longstanding illness affecting their activities all had lower mean scores than those who did not, however these differences were not significant which may be due to the small sample size. The distribution of response options showed the hospital population generally had a greater proportion of people at the lower levels and a lower proportion at the higher levels compared to the school population and this was strongest for the pain, school work, daily routine and activities questions. There were some exceptions however (angry and embarrassed which showed no difference) and annoyed, where it was mixed across the levels who was higher and who was lower. It may be that children find *angry* a strong concept and do not necessarily associate it with health and it may have occurred for other reasons in both populations alike. Children also sometimes

struggled with the *embarrassed* question as they were not sure why it related to their health. The results of the Mann Whitney tests showed that 9 out of 18 questions showed significant differences and whilst the sample size was reasonable for testing, it would be desirable to undertake these tests on a much larger sample.

Unfortunately the numbers were too low to compare the difference in distributions by different categories of patients (those who had and had not had a previous hospital admission in the last 6 months; those who had and had not had an outpatient appointment in the last 6 months and those who have a long standing illness that affects their activities compared with those who do not have a long standing illness). If the sample had been larger this would have been a very useful way of further testing construct validity and is something that can be addressed in future work testing the measure.

There was reasonable agreement between the child and parent rating, although it should be noted that the recall period was different, children were asked to think about today and parents to think about the last week. Approximately one third of parents and children agreed on ratings and about 40% of parents rated their child's health lower than the child rated it, indicating that parents are more likely to see their child's health problems as worse than the child does.

Of the questions identified as the same there was a strong preference for *sad* instead of *upset* and children were split between which was best out of *pain* and *hurting*.

Comparison with results of testing on the general population

The main differences in comparison with the testing on the general population in chapter 7, are that the mean time to complete is longer in the hospital population, the response rate is lower in the hospital population, and more children read the instructions and questions for themselves in the school population. More children struggled with the school work question in the

hospital setting, but generally the type of help asked for was similar across the two populations. Finally, more children in the school population judged questions to be the same.

The proportion of children who read the instructions themselves was just over half in the clinical population and was lower than in the schools. A lot of this was due to the fact that parents were present and children often asked them to read for them, or if they were feeling unwell or could not sit up in bed properly, they asked the researcher (KS) to read them. This was not because they could not read themselves.

The differences shown in the distributions across response options also gave some good evidence of construct validity for the measure as it was able to discriminate between a general and clinical population.

Whilst the testing on the general (Chapter 7) and clinical populations provides some evidence on psychometric performance, further testing is needed on much larger samples to provide better evidence of the performance of the measure.

8a.5 Conclusions

Overall, this work has demonstrated that in its current form, the descriptive system demonstrates good practicality and validity in this clinical paediatric setting, although it is sometimes necessary to have an adult present to help with reading the instructions or questions, especially if the child is quite unwell. The next stage is to refine the descriptive system based on the results of the testing in both the general and clinical populations.

Appendix 8a.A: Data Collection Form

DATA COLLECTION FORM

Patient study number ☐

Has your child had a previous hospital stay in the last 6 months?
Yes ☐ How many times? _____
No ☐

Has your child had a previous visit to day care/outpatients in the last 6 months?
Yes ☐ How many times? _____
No ☐

Does your child have any long-standing illness, disability or infirmity? By long-standing I mean anything that has troubled them over a period of time or that is likely to affect them over a period of time?
Yes ☐
No ☐

Does this illness or disability (do any of these illnesses or disabilities) limit their activities in any way?
Yes ☐
No ☐

Overall, how would you (parent/guardian) rate your child's health during the past week?
☐ Excellent
☐ Very good
☐ Good
☐ Fair
☐ Poor

Gender of your child:
☐ Male
☐ Female

Age of your child:
_____ years

Ethnic Origin of your child:
☐ White
☐ Mixed/Dual heritage
☐ Asian or Asian British
☐ Black or Black British
☐ Chinese
☐ Other ethnic group (please specify)_____

Appendix 8a.B

Table 8a.7: Time to complete by age

	Mean (minutes)	Median (minutes)	Minimum (minutes)	Maximum (minutes)
All cases	5.27	5	2	18
7 years old	8.06	6	5	18
8 years old	5.71	6	2	11
9 years old	4.60	4	2	8
10 years old	4.77	4	3	8
11 years old	3.59	3	2	8

Tables 8a.18 to 8a.35: Distributions across levels for the combined and separate populations.

Q1 Worrying

Level	Total combined population		Hospital population		School population	
	n	%	n	%	n	%
1	143	57.9	50	51.5	93	62
2	65	26.3	25	25.8	40	26.7
3	23	9.3	13	13.4	10	6.7
4	6	2.4	1	1	5	3.3
5	10	4	8	8.2	2	1.3

Q2 Sad

Level	Total combined population		Hospital population		School population	
	n	%	n	%	n	%
1	174	70.4	63	64.9	111	74
2	39	15.8	18	18.6	21	14
3	20	8.1	10	10.3	10	6.7
4	10	4	3	3.1	7	4.7
5	4	1.6	3	3.1	1	0.7

Q3 Weak

Level	Total combined population		Hospital population		School population	
	n	%	n	%	n	%
1	160	65	57	59.4	103	68.7
2	47	19.1	16	16.7	31	20.7
3	21	8.5	12	12.5	9	6
4	13	5.3	8	8.3	5	3.3
5	5	2	3	3.1	2	1.3
Missing	1	-	1	-	-	-

Q4 Angry

Level	Total combined population		Hospital population		School population	
	n	%	n	%	n	%
1	198	80.2	79	81.4	119	79.3
2	24	9.7	8	8.2	16	10.7
3	11	4.5	2	2.1	9	6
4	6	2.4	3	3.1	3	2
5	8	3.2	5	5.2	3	2

Q5 Pain

Level	Total combined population		Hospital population		School population	
	n	%	n	%	n	%
1	126	51.2	43	44.3	83	55.7
2	66	26.8	24	24.7	42	28.2
3	30	12.2	12	12.4	18	12.1
4	12	4.9	8	8.2	4	2.7
5	4	1.6	4	4.1	0	0
6	8	3.3	6	6.2	2	1.3
Missing	1	-	-	-	1	-

Q6 Frustrated

Level	Total combined population		Hospital population		School population	
	n	%	n	%	n	%
1	161	65.4	61	63.5	100	66.7
2	51	20.7	18	18.8	33	22
3	16	6.5	7	7.3	9	6
4	8	3.3	3	3.1	5	3.3
5	10	4.1	7	7.3	3	2
Missing	1	-	1	-	-	-

Q7 Hurting

Level	Total combined population		Hospital population		School population	
	n	%	n	%	n	%
1	135	54.7	47	48.5	88	58.7
2	62	25.1	24	24.7	38	25.3
3	15	6.1	6	6.2	9	6
4	12	4.9	6	6.2	6	4

5	9	3.6	6	6.2	3	2
6	7	2.8	4	4.1	3	2
7	7	2.8	4	4.1	3	2

Q8 School Work

Level	**Total combined population**		**Hospital population**		**School population**	
	n	%	n	%	n	%
1	160	65	56	58.3	104	69.3
2	41	16.7	10	10.4	31	20.7
3	13	5.3	6	6.3	7	4.7
4	11	4.5	5	5.2	6	4
5	3	1.2	2	2.1	1	0.7
6	4	1.6	3	3.1	1	0.7
7	14	5.7	14	14.6	0	0
Missing	1	-			-	-

Q9 Upset

Level	**Total combined population**		**Hospital population**		**School population**	
	n	%	n	%	n	%
1	174	70.7	63	64.9	111	74.5
2	47	19.1	22	22.7	25	16.8
3	8	3.3	2	2.1	6	4
4	11	4.5	5	5.2	6	4
5	6	2.4	5	5.2	1	0.7
Missing	1	-	-	-	1	-

Q10 Tired

Level	Total combined population		Hospital population		School population	
	n	%	n	%	n	%
1	79	32.1	34	35.1	45	30.2
2	102	41.5	34	35.1	68	45.6
3	25	10.2	9	9.3	16	10.7
4	16	6.5	4	4.1	12	8.1
5	2	9.8	16	16.5	8	5.4
Missing	1	-	-	-	1	-

Q11 Annoyed

Level	Total combined population		Hospital population		School population	
	n	%	n	%	n	%
1	169	69	70	72.2	99	66.9
2	48	19.6	14	14.4	34	23
3	12	4.9	6	6.2	6	4.1
4	9	3.7	2	2.1	7	4.7
5	7	2.9	5	5.2	2	1.4
Missing	2	-	-	-	2	-

Q12 Scared

Level	Total combined population		Hospital population		School population	
	n	%	n	%	n	%
1	189	76.8	63	64.9	126	84.6
2	31	12.6	17	17.5	14	9.4
3	14	5.7	7	7.2	7	4.7
4	3	1.2	2	2.1	1	0.7
5	9	3.7	8	8.2	1	0.7

Missing	1	-	-	-	1	-

Q13 Sleep

Level	Total combined population		Hospital population		School population	
	n	%	n	%	n	%
1	132	53.4	51	52.6	81	54.4
2	54	21.9	17	17.5	37	24.8
3	19	7.7	10	10.3	9	6
4	12	4.9	4	4.1	8	5.4
5	12	4.9	5	5.2	7	4.7
6	4	1.6	3	3.1	1	0.7
7	7	2.8	4	4.1	3	2
8	6	2.4	3	3.1	3	2
Missing	1	-	-	-	1	-

Q14 Embarrassed

Level	Total combined population		Hospital population		School population	
	n	%	n	%	n	%
1	196	79.7	77	79.4	119	79.9
2	35	14.2	11	11.3	24	16.1
3	8	3.3	5	5.2	3	2
4	2	0.8	2	2.1	0	0
5	5	2	2	2.1	3	2
Missing	1	-	-	-	-	-

Q15 Jealous

Level	Total combined population		Hospital population		School population	
	n	%	n	%	n	%

1	216	87.8	87	89.7	129	86.8
2	19	7.7	4	4.1	15	10.1
3	5	2	2	2.1	3	2
4	0	0	0	0	0	0
5	6	2.4	4	4.1	2	1.3

Q16 Daily Routine

Level	**Total combined population**		**Hospital population**		**School population**	
	n	%	n	%	n	%
1	168	68.3	51	52.6	117	78.5
2	51	20.7	27	27.8	24	16.1
3	11	4.5	5	5.2	6	4
4	9	3.7	9	9.3	0	0
5	7	2.8	5	5.2	2	1.3
Missing	1	-	-	-	1	-

Q17 Joining in activities

Level	**Total combined population**		**Hospital population**		**School population**	
	n	%	n	%	n	%
1	151	61.4	44	45.4	107	71.8
2	49	19.9	19	19.6	30	20.1
3	21	8.5	13	13.4	8	5.4
4	16	6.5	14	14.4	2	1.3
5	9	3.7	7	7.2	2	1.3
Missing	1	-	-	-	1	-

Q18 Rating of health

Level	Total combined population		Hospital population		School population	
	n	%	n	%	n	%
1	77	31.3	25	25.8	52	34.9
2	96	39	32	33	64	43
3	54	22	23	23.7	31	20.8
4	16	6.5	14	14.4	2	1.3
5	3	1.2	3	3.1	0	0
Missing	1	-	-	-	1	-

Table 8a.41: Which questions were judged to be the same and which wording was preferred

Combination	N stating they were the same	Wording preferred (N)
Worrying	5	3
Scared		2
Sad	17	15
Upset		2
Sad	1	1
Scared		0
Weak	1	0
Hurting		1
Weak	1	0
Pain		1
Weak	1	1
Tired		0

Angry	7	4
Annoyed		3
Angry	5	4
Frustrated		1
Angry	1	0
Upset		1
Pain	19	9
Hurting		10
Not sure		1
Frustrated	1	0
Upset		1
Frustrated	9	1
Annoyed		8
School work	1	0
Activities		1
Upset	1	1
Scared		0
Tired	9	3
Sleep		1
Not the same		5
Annoyed	1	1
Scared		0
Angry	1	0

Frustrated		0
Annoyed		1

References

National Statistics a.
http://www.statistics.gov.uk/downloads/theme_compendia/GHS05/GHS2005_AppxE.pdf
Accessed 28/11/07

National Statistics b.
http://www.statistics.gov.uk/StatBase/Product.asp?vlnk=14629
Accessed 05/12/07

Streiner, D.L. & Norman, G.R. 1995. Health Measurement Scales. A Practical Guide to their development and use. 2nd ed. Oxford University Press. Oxford.

Refinement of the Descriptive System

8b.1 Introduction

This chapter describes how the descriptive system was refined in light of the results presented in chapter 7 and 8a and also to meet the constraints of a preference based measure. The aim was to produce a descriptive system that was amenable to valuation and had taken into account the results of the testing phases.

8b.1.1 Constraints of a preference based measure

To be amenable to health state valuation, there is a limit to the number of dimensions that can be included in the descriptive system. Previous valuation work of generic instruments has varied in the number of dimensions contained in the descriptive system, from 8 in the HUI3, 6/7 dimensions in the HUI2, 6 in the SF-6D, 5 in the EQ-5D and 15 in the 15D (although this used multi attribute utility theory to estimate the valuation function) (Brazier 2007). The conventional view that is often cited in this type of valuation work is that individuals can process around 7 (plus or minus two) separate pieces of information in making a single decision (Miller 1956). Viewing each dimension as a separate piece of information would mean that ideally a descriptive system should contain around 7 dimensions, certainly no more than 9 and in general, the fewer the better as then individuals have less information to process.

In its current state, the descriptive system developed in this work contains 11 dimensions, which is clearly too many to be suitable for valuation. Some of these dimensions have alternative forms of wording, therefore prior to any dimension reduction, the first stage of reducing the draft descriptive system was to select the final wording choice for each of the 11 dimensions.

8b.1.2 Selecting the final wording for each dimension

Based on the preferences and comments of the combined general and hospital population reported in the previous chapters and the observations of the researcher, decisions about the wording to keep are summarized in Table 8b.1 below. Preferences for each alternative wording were calculated as the total number of times one wording was preferred when the alternative wording for that dimension was directly stated by the children as the same. Comments are taken from where children asked for help on particular questions and other comments they made when stating questions as the same. Other observations by the researcher are also listed and taken into account when making a final decision over the wording choice. The general principle was to go with the preferences of the children for the wording unless there was other evidence from comments or observations by the researcher that indicated to the contrary.

Table 8b.1: Decisions over the alternative wordings

Dimension to keep (and wording rejected)	Preferences of the children (n)	Comments by children	Observations by the researcher
Worried (Scared)	Worried(6) Scared(8)	Scared is confusing	Scared was often seen as a more extreme form of being worried. A few children also didn't see how it might relate to your health.
Sad (Upset)	Sad(35) Upset(22)	Upset has a stronger meaning than sad, sad could	

		be patronizing	
Annoyed (Frustrated) (Angry)	Annoyed (34) Angry(15) Frustrated (14)	Lots of children had problems with the meaning of frustrated	
Pain (Hurt)	Pain(34) Hurting (22)	Pain is more formal, hurt is easier for younger children	Lots of children asked "what hurts" for the hurt question and some thought it was emotional hurt. The pain question worked much better
School work			
Daily Routine			
Tired (Weak)	Tired (3) Weak(3)	Lots more help was required on the weak question in terms of what it meant	
Joining in activities that want to			
Sleep			
Embarrassed			
Jealous			

8b.2 Reduction of the dimensions

Having chosen the best wording for the final 11 dimensions, the next stage was to reduce the number of dimensions as 11 is too many for valuation. A decision had to be made about the reasoning for excluding the dimensions and there is a trade off with the breadth of coverage of the measure and the fact that excluding dimensions may run the risk of excluding a dimension that children might value very highly. Therefore the original qualitative interview work was examined again. The two weakest dimensions were *jealous* and *embarrassed* which were also the two dimensions not in common across the age groups. These were the weakest dimensions in the sense that they were not strongly put forward by children in comparison to other dimensions such as pain and feeling sad which emerged clearly for many different reasons, in many different contexts and situations. Whilst the aim of the qualitative analysis was to identify the breadth of ways in which children's lives were affected by their health and each dimension emerging is equally valid, there was far more substantive data behind the other 9 dimensions.

In addition, both came up as problems with reading and understanding in the testing phases in schools and hospital. In addition, although other dimensions also did not show significant differences, *jealous* and *embarrassed* had no significant difference between the general population and patients. Several children asked why jealous was included in the questionnaire and why your health might cause you to be jealous; they did not understand why it related to health. Jealous was also the only dimension to contain an unused level (level 4). This dimension was also the dimension where the least problems were reported as 87.8% of children reported no problem with this dimension. In the qualitative interviews, sometimes children said that feeling embarrassed also led to them feeling sad, so it is likely that some of this impact would be picked up in the sad dimension anyway. Due to all the reasons outlined above, jealous and embarrassed were excluded from the descriptive system. By comparison with the other dimensions, they were weaker overall, leaving nine dimensions.

To exclude any more dimensions was very difficult as the remaining 9 dimensions were all very strong (from the original qualitative work) and performed well in the testing phases. There were also no obvious ways in which two dimensions could be combined together. The original qualitative work undertaken to develop the descriptive system produced a set of dimensions which reflect how health affects children's lives but does not give information about how important these dimensions are or any notion of preference by children for the dimensions, and so there is a danger that if any further reduction is done, dimensions that are really important to children may be removed. Whilst there has had to be some reduction, by removing *embarrassed* and *jealous*, these were weak dimensions originally and so the vast majority of ways in which children's health is affected is still contained within the descriptive system. To reduce the number of dimensions further would be a big compromise on this strength of the descriptive system.

Because the statements in each level are so short, the overall amount of information contained in a health state with these 9 dimensions was felt to be no more than that contained in other health states from other generic descriptive systems in previous valuation work, such as the HUI2 (6/7) and the HUI3 (8). Some descriptive systems also combined dimensions, such as the EQ-5D combining anxiety and depression, which is actually two pieces of information, but counts as one dimension.

No further reduction of the dimensions was therefore undertaken prior to the valuation survey.

8b.3 Other changes to the descriptive system

In addition to the reduction of the number of dimensions, there were other minor changes to the descriptive system based on the results of the testing in schools and in hospital.

8b3.1 Recall period

Some children said that they started thinking about other days, especially as they got further on in answering the questions, when the instructions were less fresh in their minds. In order that they focused on the current day, the recall period 'today' was added to the end of each level and 'last night' was added to the end of each level on the sleep question. This also meant that the instructions could be shorter, without so much instruction to focus on today.

8b3.2 Perspective

The instructions asked children to think about how their health was affecting them today and answer the questions on this basis. When questioned at the end, many children said that they had started thinking about other reasons, for example, they were tired because they had stayed up late, or they were worried because they had a test at school coming up. Generally children found it difficult to remain focused on this for all the questions.

The final measure is most likely to be used in clinical trials and other large scale studies rather than for individual clinical decision making. Therefore it does not really matter why children are choosing the options, as in a randomised controlled trial in theory these reasons will cancel each other out across the arms of the trials and the only differences showing up will be those that are a difference due to the intervention, which is what is required. In addition, it is very difficult for children to disentangle exactly whether their choice of level is based on their health as *feeling worried* because you have a test coming up could just be part of a more general problem, in that you are more prone to feeling worried, which may be part of your overall mental health. Rather than leave children to make these difficult judgements and because it is unlikely to matter for the purposes of the studies where the final measure will be used, the instructions were simplified to remove this and changed to simply ask children how they are today. In this way there will also be no inconsistencies between those who can reason out because of their health and those who can't.

8b3.3 Wording changes

The title of question 1 was changed from worrying to worried (to match up with the style of the other questions).

Some children noted that the explanation was really necessary for the daily routine question and some children asked for more clarification. One child commented that they thought washing was 'doing the washing' and some children asked whether this included having a shower. To make it clearer, the explanation given in brackets for the daily routine question was altered from:

Daily routine (things like eating, washing, getting dressed, getting ready)
to
Daily routine (things like eating, having a bath/shower, getting dressed)

To make q17 (joining in activities) clearer, the wording was changed from:
Able to join in the activities that you want to (things like playing out with your friends, doing sports, joining in things)
to
Able to join in activities (things like playing out with your friends, doing sports, joining in things)

One of the most common problems identified was the use of the word affected, which occurred in both q8 (school work) and q13(sleep). Whilst this term often arose in the original interview work, some of the younger children struggled with the meaning during the testing phase. The two questions containing this term were therefore changed to the following alternative wording, using the scales from q16 (daily routine) and q17 (activities) which had worked well in the testing phases.

<u>Alternative wording for q8 (school work)</u>
I have no problems with my school work
I have a few problems with my school work

I have some problems with my school work
I have many problems with my school work
I can't do my schoolwork

Alternative wording for q13 (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

This also had the effect of reducing the number of levels in the two longest questions in the descriptive system to five, which address another of the comments made by children, that there were too many levels, especially on these two questions.

The school work question sometimes posed a potential problem as children noted they may not have any school work on the day of completion, as it may be a weekend, or the school holidays. There was no easy solution to this problem, however to improve this, the term homework was added to the dimension name as children often have home/schoolwork to do at the weekends or in the holidays. In addition, reading and writing were listed as activities as part of the explanation.

Many children suggested that including an example question would be helpful and so one was added, using the question 'upset', which was one of the original alternative wordings.

8b3.4 Stylistic changes

At the suggestion of some children the size of the answer boxes was increased and more colour was used to make it more attractive. Children liked the use of the clip art pictures on the front, however these are copyrighted to Microsoft

and so a selection of drawings done by children of researchers in ScHARR were used as the basis for creating new pictures.

The descriptive system was given the name Child Health Utility 9D (CHU 9D) to represent the 9 dimensions contained within it.

To aid with the subsequent valuation of the descriptive system, the dimensions were reordered so that all the emotion dimensions were listed together, then all the physical ones, then all the joining in/cognitive/social ones. This should help respondents process the information contained within each health state, by grouping similar dimensions together.

8b.4 Conclusion

The final descriptive system for the CHU9D is shown in Appendix 8b.A and the final questionnaire how it will appear for children (including the example and the colour and pictures) in shown in Appendix 8b.B. The overall reduction in dimensions was small (from 11 to 9) and whilst the number remaining is still reasonably large for valuation work, the overall amount of information contained in a health state should be feasible for valuation. The strength of the minimal amount of reduction done is the presentation of the breadth of coverage of the descriptive system which is a key strength for a generic measure, however it may be that following the pilot valuation, further changes to the descriptive system may be necessary.

Appendix 8b.A: Final Descriptive System – Child Health Utility 9D

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. 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

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. 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

6. 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

7. 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

8. School Work/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

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

Appendix 8b.B: Final questionnaire for children

The University Of Sheffield.

Child Health Utility 9D

Instructions

These questions ask about how you are **today.** For each question, read all the choices and decide which one is most like you **today**.

Then put a tick in the box next to it like this ☑. Only tick **one** box for each question.

Example

Today I feel quite upset so I will tick this box.

Upset

- ☐ I don't feel upset today
- ☐ I feel a little bit upset today
- ☐ I feel a bit upset today
- ☑ I feel quite upset today
- ☐ I feel very upset today

Now think about and answer the rest of the questions below

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 (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

References

Brazier, J.E., Ratcliffe. J., Salomon, J. & Tsuchiya, A. 2007. Measuring and Valuing Health Benefits for Economic Evaluation. Oxford University Press.

Miller, G.A. 1956. The magical number seven plus or minus two: some limits on our capacity to process information. Psychological Review 63, 81-7. Cited in: Fischer, G.W., 1978. Utility Models for multiple objective decisions: do they represent human preferences? Decision Sciences 10, 451-479.

Valuation Study

9.1 Introduction

This chapter reports on the final phase of the research, which was to generate preference weights for the final descriptive system. This was a feasibility study and the objectives were to test the feasibility of valuing this classification and generate preference weights for every health state defined by the descriptive system.

9.2 Methods

9.2.1 Overview

This study was a feasibility study to estimate preference weights for all the health states defined by the descriptive system. Valuation interviews were undertaken with the UK adult general population to obtain preference weights for a sample of the health states in the system. Regression modelling was then carried out to estimate a model to predict a value for every health state defined by the system. A range of models were tested and were evaluated based on their predictive performance.

9.2.2 Valuation technique

Currently, the recommendation is to obtain health state preferences using either the Standard Gamble (SG) or Time Trade Off (TTO) elicitation methods (Gold 1996) (Brazier 1999). In addition, current guidance from the National Institute for Health and Clinical Excellence (NICE) recommends for its reference case that a choice based method be used, such as the SG or TTO. These approaches are preferred because they are choice based and involve respondents making a choice involving a sacrifice (risk of death for standard gamble and years of life for time trade off). They were preferred based on the recommendations of the advisory panel.

NICE also recommends that the preferences are obtained from a representative sample of the public. It does not actually specify that this sample of the public must be adults, but from previous experience of valuation work we know this type of work is possible with adults as it has been done many times, however it may be in the future that we want to consider using children's preferences (NICE). In order to be consistent with NICE and due to its successful application in the valuation of other preference based instruments, including the SF-6D (Dolan 1997), the Health Utilities Index 2 (Brazier 2002) and the ADQoL (Stevens 2005) the SG method was chosen using a sample of the UK adult general population. In addition to using SG, ranking was also used as an additional method of valuation. There has been recent interest in using this technique for health state valuation (Brazier 2007) and it may have potential for use in future valuation work to try and obtain the preferences of children and so the feasibility of using ranking methods to value the descriptive system was also tested.

The standard gamble approach asks respondents to make a choice between a certain intermediate outcome and the uncertainty of a gamble with two possible outcomes:

- Choice A: 100% chance of the health state being valued
- Choice B: A chance of perfect health with probability p and a chance of dead with probability 1-p.

The value of p is varied until the respondent is indifferent between the two alternatives A and B. The point of indifference is the utility value of the health state.

9.2.3 Sample

Due to resource constraints, there was a limit to the number of interviews that could be undertaken. There was enough money to contract a survey company to undertake the interviews, however there was a trade off to be made between

the number of interviews achieved and geographical spread. As this was a feasibility study and previous research has demonstrated that geographical location does not make a difference to health state values (The MVH Group), a decision was made to undertake local face to face interviews as this would mean a much bigger sample size, which was felt to be more important than achieving geographical spread. A spread across age, gender and ethnicity and social class could still be achieved however. A sample size of 300 was possible with available resources and compares favourably with the 200 used in the UK valuation of the HUI2 (McCabe 2005) and the ADQoL (Stevens 2005), although approximately half of the sample size used in the SF-6D valuation (Brazier 2002).

An interview team of three people was contracted to undertake the interviews from the Centre for Research and Evaluation (CRE) at Sheffield Hallam University. CRE have worked with ScHARR on a number of occasions on previous valuation studies and their interview team are very experienced in this type of work, particularly in the use of SG and TTO methods. CRE also undertook the sampling, management of the interviews and entered the data.

A random street sample was selected from addresses in Sheffield and Huddersfield using AFD Names and Numbers software which provides access to UK names and addresses for over 39 million people. The sampled households were all posted a letter, inviting them to take part in the research, (Appendix 9.1). When interviewers called at the household, participants were given an information leaflet to read with further information (Appendix 9.2). After reading this and asking any questions they wanted, participants who agreed to take part were asked to sign a consent form (Appendix 9.3) and an interview was arranged in their own homes at their convenience.

9.2.4 Selection of health states

As there are 1,953,125 unique health states (levels^{dimensions} =5^{9}) defined by the CHU9D descriptive system, it was infeasible to value them all. Instead, a sample

was selected to be valued and a model estimated using these values, to predict a value for every health state defined by the system. This approach has been used successfully in a number of previous valuation surveys (Dolan 1997) (Brazier 2002) (McCabe 2005). Previous experience has shown that respondents can manage about 9 valuation tasks in an interview, (Brazier 2002) (Dolan 1997) so with 300 interviews, this gave 2700 potential observations.

An orthogonal array of health states was generated using the Orthoplan feature of SPSS (Version 12) which generates a design for an additive model using the minimum number of health states required to estimate a model to predict all health states in the descriptive system. This found the minimum number of states required for a 9 dimensional system with 5 levels per dimension was 64. The design generated included 2 duplicate states and also the best state in the descriptive system (level 1 on each dimension, termed state 111111111) twice. As the design of the SG method assumes the best state to be equal to 1, it was not possible to use this as an intermediate state for valuation, therefore two substitute states were created to replace these best states, with all dimension levels at the top apart from 1 dimension, so as to keep the replacement health states as close as possible to the top of the descriptive system, keeping as close to the orthoplan design as possible. Others working in the field have found that it is common to get duplicate states in this size of descriptive system and so more observations were obtained on the duplicate states, rather than substituting more states. The included states are listed in Table 9.1, showing the level on each dimension. Each health state can be represented by a 9 digit number, with each digit representing the level on each of the dimensions. The 64 states were divided into 8 sets of 8, trying to balance the severity of states in each set (by looking at the levels on each dimension) and making sure the 2 duplicate states were separated. The worst health state (Pits, 555555555) was added to each set, giving a total of 9 health states in each set. The interviewers rotated round the sets so that each state got an equal number of observations and each respondent only had 9 SG valuation tasks to do.

As adults were valuing the health states, the schoolwork/homework dimension was changed to work. This means that the health states that the adults are now considering for valuation make sense as adult health states. Changing schoolwork/homework to work for adults changes one of the dimensions of the descriptive system. Whilst this is perhaps the closest in meaning that this can be for an adult, there are differences in how health may affect work and so the implications of this mean that the health state being valued is not quite what is intended. This is something that can be tested in future valuation work, for example valuation work with adults that considers the actual child state from the perspective of the child, instead of a non child state from their own perspective. The descriptive system is shown in Appendix 9.4 and an example health state is shown in Appendix 9.5.

9.2.5 Valuation interviews

The respondent was first asked to self complete the descriptive system in order to familiarise themselves with it and also to understand the range of the levels within each dimension. As the respondents were adults, the description of the activities dimension was altered slightly to make it meaningful for them, changing 'playing out with your friends' to 'seeing friends'.

Respondents were then asked to rank the 9 health states from the set being used, plus the best state in the descriptive system (111111111) and dead. This was followed by SG valuations of all 9 health states. The version of the SG script used was the same as that used in the original and UK HUI2 valuations (McCabe 2005) (Torrance 1996), which is the ping pong version developed by the team at McMaster University. A Chance Board prop was also used, which displays the probabilities both numerically and in the form of a pie chart. This method uses increments of 10% except at the top and bottom ends where it is more sensitive between 90 and 100 and 0 and 10 and uses increments of 5%. The perspective the respondent was asked to imagine was that they would be in the health state

described for the rest of their life. This is in contrast to the approach used in the valuation of the HUI2 (McCabe 2005) (Torrance 1996) where the respondent was asked to imagine that they were a child of 10 years old and would be in the health state described until the age of 70 and would then die. Experience in the UK valuation of the HUI2 (McCabe 2005) showed that respondents found this difficult to do and some were bringing their experiences of adulthood into the valuation and some were trying to remember what it was like when they were 10. Others also thought of an imaginary child of 10. For this reason, the perspective was chosen to be simple and the respondent was asked to imagine themselves in this health state for the rest of their life. The health state was valued against perfect health (as state 1,1,1,1,1,1,1,1,1,1 in this descriptive system is assumed to be equivalent to perfect health) and dead. If respondents rated a health state worse than dead in the ranking exercise, they did a worse than dead form of SG valuation where a certain choice of dead was offered as a choice against perfect health with probability p and the health state with probability (1-p). The utility value is –p at the point of indifference. This differs from the original HUI2 valuation as they chose not to value health states worse than dead. The methodology for states worse than dead used here is that undertaken in the valuation of the SF-6D (Brazier 2002) and the UK HUI2 (McCabe 2005) valuation and is based on the transformation by Patrick et al (Patrick 1994).

Finally, basic socio demographic information was collected and questions asking respondents how difficult they found the tasks. After completion of the interview, the interviewer was asked to assess how well they felt the respondent had concentrated and understood the tasks. The start and end times of the rank and SG tasks were also recorded. The script of the interview is included as Appendix 9.6. To obtain information on the socio-economic characteristics of respondents, the database supplying the names and addresses (AFD) (Names and Numbers Manual) also provides censation data on the households (based on the latest census data), including affluence, which categorizes people into five categories (wealthy, prosperous, comfortable, striving and struggling). The figures for the sample were compared to the UK population as a whole.

9.2.6 Exclusion criteria

Based on the principles used in previous valuation work, for the SG modelling, respondents were excluded if they valued all health states the same (Dolan 1997) (Stevens 2005) (McCabe 2005) and observations were excluded if the data was unusable. This was where respondents could not decide whether a health state was better or worse than dead at the beginning of a valuation task (having already ranked the health state), where respondents chose 100% chance of dead over 100% chance of the health state even though they had confirmed the health state was better than dead and where respondents chose 100% of dead rather than 100% chance of perfect health. If respondents valued all health states the same, this was taken to be an indication of misunderstanding the task. Whilst imposing restrictions on the data by excluding respondents, it was not possible to incorporate unusable data. Excluding respondents who valued all health states the same is common practice in this type of research (Brazier 2002) (McCabe 2005).

9.2.7 Modelling

The aim was to estimate a model for predicting health state values for every health state defined by the descriptive system. The approach taken in previous valuation work of generic measures was followed (McCabe 2005) (Brazier 2002).

The basic model structure for the model was:

$$Uij = g(\beta xij) + \varepsilon ij$$

Where:

i = 1,2,.....,n represents individual health states in the descriptive system;

j= 1,2,.....,m represents individual respondents;

Uij = the standard gamble value for health state i valued by respondent j;

g = appropriate functional form;

x = a vector of dummy variables for each level of each dimension in the descriptive system.

εij = the error term.

Personal characteristics were not included in the model as the model is intended to be used as a societal model of preferences and not adjusted for individual characteristics. Whilst there may be personal characteristics that prove important and can be estimated as respondent level covariates, these will not be used when applying the algorithm in practice as the aim was to estimate a utility function for the UK population as a whole.

To estimate the model, dummy variables are created for each level on each dimension with level 1 acting as the baseline for each dimension. The dummy variables take a value of 1 if the health state has the dimension at the level the dummy is representing and 0 otherwise. In a simple linear model, the intercept represents the estimated value of the best state (1,1,1,1,1,1,1,1,1,1) and the value for all other health states are derived by summing the coefficients of the appropriate dummies. Models estimated in this way have utility as the dependent variable. An alternative is to assume that because this is a generic measure, the best state can be assumed to be perfect health and so takes a value of 1. To achieve this, the constant can be forced to unity by estimating the model with no intercept and the dependent variable becomes disutility (U_{ij} -1). The value of a health state then becomes 1 plus the sum of the coefficients on the relevant dummy variables. Both the utility and disutility forms of model were estimated.

There are numerous possible interaction terms in the descriptive system and modelling them all would require a much larger dataset to prevent the risk of finding statistical significance due to chance (McCabe 2005). In addition, previous valuation work has found that interactions do not improve the models and often increase the number of inconsistencies (The MVH Group) (Brazier

1999) (Torrance 1996). As the orthogonal design of the survey was for main effects only, only very basic interactions were estimated. Therefore two interaction terms were added, following the previous methods of the UKHUI2 and SF-6D valuations. A dummy variable (MOST) taking the value of 1 if a health state had any level at level 1 otherwise 0 and a dummy variable (LEAST) taking the value of 1 if a health state had any level at level 5, otherwise 0 were added.

Other specifications and transformations such as Tobit models could be considered, however previous valuation work has shown that these do not improve the modelling at all (McCabe 2005, Brazier 2002). Another recent approach that has been applied is that of Bayesian non-parametric modelling. This has been successfully applied in both the valuation of the SF-6D (Kharroubi 2007) and the UK valuation of the HUI2 (Kharroubi forthcoming). This is a complex approach and requires a balanced design approach to the sampling of the health states which is different from the orthogonal array used here and so was not considered in this thesis.

9.2.8 Specification of the models

The choice of model specification depends upon the type of data used. Standard OLS regression assumes a zero mean, constant variance error structure, with independent error terms, i.e. $cov(\varepsilon ij, \varepsilon i'j)=0$, $i \neq i'$ (Gujarati 1995). This assumption means that the 2700 observations from 300 respondents are treated as though 2700 respondents provided them.

An alternative specification is the Random Effects (RE) model, which allows for the fact that the error term may not be independent of the respondent, and separates out within and between respondent error terms.

$\varepsilon ij = uj + eij$

Where:

uj = the respondent specific variation (assumed to be random across individuals)

eij = the error term for the ith health state valuation of the jth individual (assumed to be random across observations)

The RE model also assumes that the allocation of health states to respondents is random i.e. cov(uj, eij)=0 (McCabe 2005).

A different specification is the fixed effects model, which also allows for the importance of individual effects but does not assume these are random, instead the respondent specific variation is estimated along with the coefficients on the explanatory variables.

To test whether individual effects were important the Breusch-Pagan test was used and if they were, the Hausman test was used to determine whether fixed or random effects were appropriate (Brazier 2002).

In addition to the individual level models described above, mean and median aggregate level models were estimated using the mean and median values for each health state. It could be argued that these models do not make the most efficient use of the data, however they use the information that is perhaps of most interest to policy makers, the central estimate for each health state (McCabe 2005).

One further type of modelling was undertaken, which made use of the ordinal preference data obtained from the ranking exercise. This type of rank modelling is a more recent development in health state valuation modelling and has been successfully applied in the major generic PBMs (Brazier 2007). The basic foundation for this type of modelling (estimating cardinal values from ordinal data) is based on Thurstone's law of comparative judgement (Brazier 2007). The modelling process followed that undertaken by McCabe in estimating a rank model for the HUI2 (McCabe 2006). The ranking task was designed to include the state 'dead' so that the modelling could be normalised to produce a utility of 0 for dead. By including dead in the regression model, the estimated coefficient

can be used to rescale the results onto a scale with dead as 0 (Brazier 2007). Of all the valuation techniques, ordinal methods are the most likely to be able to be undertaken by children and given that in the future, it would be desirable to obtain children's valuations for this descriptive system, ranking models are an important part of the feasibility testing.

The model is a rank-ordered logit model taking the following general form:

$$Uij = \beta xij + \theta D + \varepsilon ij$$

Where x is a vector of dummy explanatory variables for each level of each dimension, for example x45 is dimension 4(tired) at level 5 (I feel very tired). The dummy variables take a value of 1 if the dimension is at this level and a value of 0 otherwise. As in the SG modelling, level 1 acts as a baseline for each dimension. D is a dummy variable for dead which takes the value of 1 for this state and 0 otherwise. Perfect health is constrained to equal 1 and the value of a health state is calculated by subtracting the sum of the coefficients for each of the dummy variables from 1. As the model is not directly estimating utility on the 0-1 (dead- perfect health) scale required for health state valuation, the coefficients have to be rescaled using the formula $\beta rij=\beta ij/\theta$; where βrij is the rescaled coefficient and θ is the coefficient for dead. By rescaling, the model produces values on the dead (=0) - perfect health (=1) scale.

9.2.9 Assessment of models

Several measures were used to assess model performance. Firstly, coefficients of the models were examined to see if they were significant and had the expected negative (for utility) or positive (for disutility) sign. As the dummies represent progressively worse problems on each dimension starting from a baseline of no problems, the coefficients were expected to be increasing in absolute size. Logical inconsistencies in the coefficient values were looked for, in that the lower the level, the larger the decrement should be, for example pain

level 5 should have a higher decrement than pain level 4. The adjusted R^2 for each model was also reported (where appropriate).

Models were also assessed on the basis of their predictive performance, i.e. how well they predicted observed mean values. To do this, a number of measures were used. The mean absolute error (MAE) and root mean square error (RMSE) were calculated, which are both summary measures giving an indication of the prediction errors of a model, with the RMSE giving more weight to larger errors. In addition, the percentage of health states predicted to within 0.1 and 0.05 (absolute value) of the observed mean value are reported. The value of 0.1 was chosen as it is the value used in previous valuation studies (McCabe 2005) and the value of 0.05 was chosen as it has been considered an important difference in many contexts (O'Brien 1994).

Finally, the predicted health state values were plotted against the observed health state values to look for any patterns in the errors. Both the SG and the rank models were tested against the observed mean SG values after the exclusion criteria were applied. A test of the null hypothesis that the mean prediction error was 0 was undertaken for each of the models in order to determine whether there was any bias in the predictions. A Ljung box test was also carried out to test whether there was any non randomness in the prediction errors, i.e. if the error was systematically related to the severity of the health state (Ljung 1979). Errors were ordered by actual mean health state valuation.

The rank model was tested for a key assumption of this type of modelling which is the independence from irrelevant alternatives which states that the ordering of a given pair of items does not depend on the other alternatives available (Brazier 2007). This uses a Hausman test comparing datasets where one alternative is dropped, with the full dataset. The test is the same as that

undertaken by McCabe et al (McCabe 2006) in their rank modelling of the HUI2 and SF-6D. The Hausman test compares the maximum-likelihood estimator of beta based on the full dataset with maximum likelihood estimators of beta based on a dataset where one alternative is dropped. Here, the cards ranked first were dropped and the model re estimated, then the cards ranked second, then third and so on. Under the assumption, the betas from the two models being compared should be approximately the same. If they are significantly different then the assumption is violated. All modelling and analysis was carried out using STATA version 10.

9.3 Results

1245 addresses were mailed to and of these, 1195 were approached in person at the door. Out of those approached, 534 (45%) were not in/no contact was made, 320 (27%) refused and 300 (25%) agreed to be interviewed. Therefore the response rate was 25%. Information on the characteristics of the non-consenting individuals was not available as ethical constraints did not allow collection of this data.

In total 300 interviews were carried out. For the SG modelling, 52 observations were excluded as unusable and 17 respondents were excluded as they valued all health states the same. In addition, 1 respondent was excluded as they valued the Pits state as 1 and all other health states at 0.95. This led to a dataset with 2478 observations from 282 respondents (6% of respondents were excluded). For the rank modelling, no exclusions were made.

The characteristics of the included and excluded populations are shown in Table 9.2. Compared to the included population, the excluded population had a higher % of men, more left school at 16, more found the SG exercise very difficult and the interviewer rated the understanding and concentration in the ranking and SG tasks lower. In addition, 28% of the excluded population said the SG task was very easy, compared to 8.2% of the included population, perhaps indicating that

they had misunderstood the task as the SG task is not generally seen as very easy. χ^2 tests or Fischer's exact tests (as appropriate given the n) to look for any difference between the groups showed no difference in gender, employment, education or ethnicity. Differences were found between those included and excluded for respondents having difficulties with the SG task (Fischer's exact 0.035), the interview rating of the respondents understanding of the ranking task (Fischer's exact 0.018), the effort and concentration put into the ranking task (Fischer's exact 0.00) and the effort and concentration put into the SG task (Fischer's exact 0.00).

The socio economic characteristics of the whole sample were as follows (data was missing for 1 person) and the UK figures are also given.

Category	**Sample %**	**UK%**
A, Wealthy	32.8	23.4
B, Prosperous	7.7	20.9
C, Comfortable	19.1	19.7
D, Striving	33.8	21.3
E, Struggling	6.7	13.6

There are more wealthy people in the sample compared to the UK, but significantly less in the prosperous category. There are also more in the striving category in the sample than the UK population.

Descriptive statistics for the health states from the included respondents are shown in Table 9.3. Each state has been valued 35 times on average (minimum 32, maximum 39), apart from the 2 duplicate states (222222212 and 333333313) which were valued 68 and 72 times respectively. In addition, the Pits state (555555555) was valued 235 times.

The mean health state values range from 0.387857 to 0.931579. The median mostly exceeds the mean (66.7% of cases). There were 23 negative valuations (0.93%). Figure 9.1 shows the distribution of health state values.

The interaction terms made things worse and did not improve the modelling as they increased the number of inconsistencies and decreased the number of significant coefficients. They are just reported in Appendix 9.7. Overall, the disutility models were much better than the utility models in terms of the number of significant coefficients (higher) and the number of inconsistencies (lower), therefore only the disutility models are reported in full and the utility models are in Appendix 9.8. The Breusch Pagan test suggested that individual effects were present in the data (chi2= 2388.23, p=0.00). The Hausman test did not work as the model fitted failed to meet the asymptotic assumptions of the Hausman test. This was probably due to a misspecification problem, in that the random effects model was not efficient for the data. A more general test (seemingly unrelated estimation) was tried but this also failed. Therefore both random and fixed effects models were estimated and judged on the basis of their predictive performance, but bearing in mind that they were probably not the best models.

Going on the number of significant coefficients and the number of inconsistencies, the best three models from all the disutility models were the OLS, RE and mean disutility models with the constant restricted to 1. They are summarized in Table 9.4. together with the model estimated from rank data.

All coefficients are significant in the rank model, have the expected sign and there are 8 inconsistencies. It is shown alongside the three best restricted disutility models in Table 9.4. The results of the Hausman test for the independence from irrelevant alternatives are shown below. Significant results are shown in bold. It was not possible to test the models estimated without health states ranked second and tenth as the model violated the assumptions. A more general test did not work either. The models are sensitive to excluding those health states at the top and towards the bottom of the rankings as those

ranked first, eight and ninth and significant, hence we reject the equality and there is some evidence that the assumption does not hold.

Hausman test for the independence from irrelevant alternatives

Alternative dropped	Hausman	Prob>chi2
1	211.04	**0.0000**
2		
3	38.60	0.3970
4	20.80	0.9854
5	15.56	0.9992
6	22.20	0.9741
7	25.65	0.9201
8	87.82	**0.000**
9	196.00	**0.000**
10		
11	4.40	1.000

For the OLS model, all the coefficients have the expected positive sign and there are 30 out of 36 coefficients significant at the 0.1 level. There are 14 inconsistencies which reduces to 10 if you remove those that are not significant at the 0.1 level. The RE model also has all coefficients with the expected positive sign and there are 33 out of 36 significant at the 0.1 level. There are 11 inconsistencies which reduces to 10 if you remove the one not significant at the 0.1 level (tired 5). The mean model has all coefficients with the expected positive sign and 28 out of 36 are significant. There are 14 inconsistencies which reduces to 8 removing those that are not significant at the 0.1 level. Finally, the rank model has all 36 coefficients significant at the 0.1 level and they are all the expected positive sign. There are 8 inconsistencies.

In terms of predictive performance, the OLS and mean models perform best with 100% of errors within +/-0.1, whilst the RE model is still high at 98.4 and the rank model is the worst, at 90.5. When the accuracy is increased to within 0.05, the mean model performs best, at 98.4% with the OLS next at 90.5%. The RE model is much lower at 77.8% and the rank model is the worst at 65.1%.

The MAE is lowest for the OLS model at 0.0261, closely followed by the mean model at 0.0263. The MAE of the RE model is higher at 0.0313 and the rank model has the highest at 0.0461. The RMSE is lowest for the mean model, then the OLS, then the RE and finally is highest for the rank model at 0.0573.

Figures 9.2 – 9.5 show plots of the observed values and the predicted values for each of the four reported models. They are ordered by the observed mean value. There does not appear to be any systematic pattern in the errors apart from the rank model which under predicts at the higher end (i.e. the health states with a higher observed mean value). The results of the Ljung box tests for each of the models (1 to 4) are shown below.

	OLS(1)	RE(2)	Mean(3)	Rank(4)
Test statistic	4.9116	8.9025	5.3772	7.9305
Prob>Chi2(8*)	0.7670	0.3506	0.7166	0.4403

* The number of lags is the square root of n which is conventional for this type of test.

None of the test statistics are significant, therefore none of the models show evidence of autocorrelation in the prediction errors.

The RE model appears to be the only model that gives biased predictions, as indicated by the t test of the null hypothesis that the mean prediction error is 0.

9.3.1 Further modelling

Despite the very good predictive performance, the models still have inconsistencies in them and some coefficients are not significant. Therefore further modelling was undertaken to estimate a parsimonious consistent regression model using the general to specific approach. This approach was used in the valuation of the SF12 and later SF-36 models (Brazier 2004). These models were constructed by combining levels where inconsistencies were

present and removing levels not significant at $p<0.1$. This was done on the two best performing models (the mean and OLS restricted). These two models are shown in Table 9.5 and graphs of their predictive performance are shown in Figures 9.6 and 9.7. All coefficients are significant at $p<0.1$ and all but 1 coefficient in both models are significant at $p<0.05$. There are no inconsistencies in these models. The models are also consistent in that the same levels had to be combined, apart from the OLS model which still has sleep4, whereas the mean model has sleep234 combined. The dimensions worry, annoyed and tired all had all levels (except level 1) combined. Levels 4 and 5 were combined for sad, levels 2 and 3 for pain, levels 2 and 3 for work and 4 and 5 for work and levels 2, 3 and 4 for activities. The predictive performance is not as good as the full models where levels were not combined. The MAE for the OLS model is 0.0343 and similar at 0.0349 for the mean model. These are higher than the full models. The RMSE are also similar, at 0.0426 for the OLS model and 0.0431 for the mean model. Both models predict well at 98.41% of predicted values within 0.1 of the observed mean, and the mean model is slightly better at 76.19% of predicted values within 0.05 of the observed mean compared to the OLS model, which predicts 73.02%. Neither model had biases in the prediction errors as indicated by the t test of the null hypothesis that the mean prediction error is 0. There do not appear to be any patterns in the prediction errors either from looking at the graphs in Figures 9.6 and 9.7. The results of the Ljung box test are shown below for both models. Neither model shows evidence of autocorrelation in the prediction errors.

	OLS reduced (5)	Mean reduced (6)
Test statistic	6.5165	6.7737
Prob>Chi2(8*)	0.5896	0.5612

* The number of lags is the square root of n which is conventional for this type of test.

9.4 Discussion

Health state values were successfully generated for the health states in the survey and a reasonable range of values was produced, although the mean value for the Pits state (0.337) was perhaps higher than what was expected. This compares with other generic descriptive systems with mean health state values which ranged from -0.543 to 0.878 (EQ-5D (Dolan 1997)), 0.10 to 0.99 (SF-6D (Brazier 2002)) and -0.07 to 0.79 (UKHUI2 (McCabe 2005)). The Pits state could be low due to the language used due to the nature of it being a paediatric descriptive system. Adults did not have any knowledge that the states being valued were child health states when undertaking the valuation tasks. Hence, when reading the descriptions, they may have placed less weight on the severity of the levels. For example, the level "I feel very worried today" may be seen by a child as really severe, however for an adult, who is perhaps thinking in terms of stronger language such as anxiety, this might not seem so severe, as they can imagine much worse levels, for example" I feel really anxious".

Generally, the modelling was successful and overall the disutility models performed much better than the utility models and the best performing of these were the models where the constant was restricted to equal 1. This fits in well with the practical application that is required of these models in calculating QALYs, in that a scale with perfect health =1 and dead =0 is required and there are strong theoretical arguments for restricting the intercept to unity (Brazier 2002). This model assumes that the best health state in the descriptive system (111111111) has a value of 1 and dead has a value of 0.

Overall, the mean model was the best in terms of predictive performance as it is the most accurate at predicting observed mean values, with the highest percentage predicted within +/-0.05 for all models, the lowest RMSE and a low MAE (nearly equivalent to OLS MAE which is the lowest). The mean model also has one of the lowest number of inconsistencies (8), the same as the rank model. The rank model was the worst in terms of predictive performance, being the worst on each measure of performance, however it is the only model with all

coefficients significant at the 0.1 level and has the same number of inconsistencies as the mean model. None of the models had any problems with autocorrelation in the prediction errors.

It should be noted that the SG models are being tested against the data they were estimated on, whereas the rank models are not, although the data comes from the same respondents. Therefore it is perhaps not surprising that the rank model is outperformed by all the SG models. However, the rank model is still a reasonably good model and there are similarities with the inconsistencies in the SG models, for example, sad5, tired4, tired5, pain3 and activities4. It is notable that the rank model performs well and is similar to the SG models in terms of what levels are inconsistent and this gives encouraging results for using this type of valuation technique in the future to access children's valuations.

Other studies that have used rank models have found that the results are not dissimilar to the SG models. The UK valuation of the HUI2 found the rank model increased the inconsistencies by 1 (McCabe 2006) and found the best SG model performed better on all tests, but was remarkably similar. The SF-6D found the rank model quite different to the best performing SG model, as the number of inconsistencies decreased however the predictive performance of the rank model was only slightly worse (McCabe 2006).

The inconsistencies were similar across the different model specifications, the most common were sad5, annoyed3, annoyed4, pain3, sleep3, work3, work5 and activities4. The exception was the rank model which was the only model to be consistent for the work dimension.

Estimating parsimonious consistent models from the 2 best performing full models worked, although several levels had to be collapsed. The results of this were similar across the 2 models which is reassuring. Part of the collapsing may be due to the fact that adults were valuing these health states and not children.

For example the dimension *worry* is perhaps not seen as very strong by an adult in contrast to the similar concept usually used in adult measures, anxiety. Similarly, it may be that adults see being *annoyed* as nothing particularly bad and so this dimension also had collapsed levels. Perhaps the most surprising dimension that had to combine levels was *tired*, however it may be that because there is also a *sleep* dimension, the adults valuing these health state focused in on that. Alternatively, it may be that the descriptive system is too big with 9 dimensions or perhaps there are too many levels and adults are employing simplifying heuristics when valuing the health states. Larger descriptive systems are more likely to result in doing this, such as just focusing in on key dimensions (Lloyd 2003). Undertaking a large valuation survey with children valuing the health states would provide more information on this issue and also using 'think aloud' techniques when people are valuing health states to gain a better understanding of what they are focusing on and whether they use any heuristics. There may also have been some implications with changing school work to work.

The results of the models could be used to refine the descriptive system, in that those levels or dimensions that are not significant could be dropped as they are not showing as being valued as important by adults. However, given that this is a feasibility study and no valuation work has yet been done with children on the descriptive system, it is not strong enough evidence to do this. In addition, it may be that dimensions or levels not valued by adults, may be valued by children. This would then lead to the interesting question of whose values should you use.

The preferred overall model is model 5, the OLS parsimonious consistent model. This model has all coefficients significant and has no inconsistencies. It is slightly better on predictive performance that the mean model.

A value for every health state in the descriptive system can be estimated by subtracting the sum of the coefficients of the relative on dummies from 1. For example, for health state 233243425, the value would be:

1-(0.0227+0.0445+0.0313+0.0479+0.1245+0.0212+0.0699+0.0487+0.1079)

=0.4814

The mean absolute error of the best model (model 5) is 0.0343 and this amount is unlikely to be considered meaningful in many contexts (Torrance 1996). As the aim of the model is to predict mean health values across patients in many different states and the error is random, this is an acceptable error when using the model in practice. Research by Walters and Brazier (2005) also found that the minimally important difference in utility score for the SF-6D was 0.041 (mean) and 0.074 (mean) for the EQ-5D.

The proportion of health states valued out of the entire descriptive system was very small at 0.003% (63/1,953,125). This compares with 1.4% for the SF-6D valuation and 0.64% for the UK HUI2 valuation (McCabe 2005). It may be that with a larger dataset and a larger number of health states being valued, some of these problems may be overcome. This study was a feasibility study and future research can test this.

One of the most important factors in this valuation study is that the population valuing the health states is adults, in contrast to the descriptive system which is for children. In addition, when valuing the health states, adults were not aware that these states were child health states and were asked to imagine themselves as they are now (as an adult) in this health state for the rest of their lives. Adults were chosen for the valuation survey as using children to undertake valuations is something that has not been done before. This does not mean it is not possible, rather that further research is needed to investigate whether it is feasible to obtain valuations from children. This valuation study was a feasibility study and

the intention was not to test new methods. The SG and TTO methods are cognitively demanding and it is uncertain whether children would be able to manage these tasks. There are also ethical issues that would be raised by asking to children to think about scenarios that involved a risk of death. In recent years, the use of ranking/ordinal methods to value health states has increased and this is perhaps a method that would be more appropriate and feasible to undertake with children. Perhaps the simplest way would be to present health states in pair wise comparisons as ranking many health states at once can be just as cognitively demanding. There is also the issue of the independence from irrelevant alternatives. Work has been done to value descriptive systems in this way using ordinal techniques, including the use of discrete choice experiment (DCE) techniques for estimating preference weights for a sexual quality of life questionnaire (Ratcliffe 2006) and for an asthma quality of life questionnaire (Brazier 2006). Even if you were able to elicit children's valuations for health states however, there is the question of whether these are the values that you want to use for making resource allocation decisions. Similarly, it would also be interesting to undertake preference elicitation work where the adult valuing the health state knew this was a child health state and see if this makes any difference to the values. This was the approach taken in the valuation of the HUI2 and the ADQoL (Torrance 1996) (Stevens 2005). Whether children's valuations should be used is a normative issue and there are arguments either way. It can be argued that children are not rational, informed and autonomous individuals (an ideal for health state valuation) and therefore should not undertake valuation tasks. However, it may be that some adults also do not fulfil this criteria and previous valuation work has demonstrated some evidence of this as respondents have been excluded on the grounds of irrational or inconsistent responses (Brazier 2002, McCabe 2005). Perhaps more importantly, some people may argue that society does not see children as legal agents, in that before the age of 18 they are not allowed to vote and hence not viewed as decision makers in society.

What is unknown is how children's valuations may differ (if at all) from adult valuations. If there is very little difference in values, then it perhaps becomes irrelevant whose values are used. However, if there are differences, then a decision would have to be made over which values are more appropriate. This is something that can only be determined empirically and what would be most interesting is perhaps the potential differences in strength of preference for the different dimensions of health. These important questions should be the subject of future research.

9.5 Conclusions

This research has demonstrated that it is feasible to value the descriptive system and preference weights have been generated for all health states defined by the system. A number of models have been estimated using both the SG and ordinal (rank) data. The best performing models were restricted disutility models, which restrict the constant term to 1 and have stronger theoretical arguments. The model recommended for use in assigning preference weights for the health states defined by the CHU9D is the OLS parsimonious model (model 5). The CHU9D is now able to be used to generate quality adjusted life years (QALYS) by using the system and combining it with length of life. The CHU9D offers an alternative to the HUI2 and can be used in the economic evaluation of paediatric health care interventions. Further research is needed to investigate the impact of children's preferences for the health states and whether ordinal methods can be used to achieve this.

Table 9.1: Health States Used

WORRIED	SAD	ANNOYED	TIRED	PAIN	SLEEP	DAILY ROUTINE	WORK	ACTIVITIES
3	3	2	2	4	5	1	2	1
2	1	3	1	2	5	4	3	3
2	2	2	2	2	2	2	1	2
2	2	5	3	4	1	3	3	1
1	5	3	3	2	4	1	2	2
5	5	5	5	5	5	5	1	5
4	2	3	2	1	5	3	4	1
5	4	2	3	2	1	3	3	1
1	2	3	5	2	3	4	5	1
1	2	2	5	3	1	3	3	4
4	4	4	4	4	4	4	1	4
4	1	5	3	3	2	1	5	2
3	3	3	3	3	3	3	1	3
3	4	1	1	2	5	3	2	2
2	1	2	3	3	5	1	5	4
2	1	4	2	3	3	5	2	1
5	1	2	4	3	3	2	2	1
3	1	3	5	1	4	2	3	2
1	1	5	4	2	2	3	2	3
2	3	2	1	1	4	3	5	5
5	2	1	2	3	4	1	4	3
1	1	2	2	5	4	3	2	3
5	2	1	3	1	3	4	2	2
5	1	3	1	4	2	2	3	3
2	3	1	3	2	4	5	3	1
2	4	1	5	3	2	1	4	3
2	3	3	4	5	1	1	1	2
2	2	2	2	2	2	2	1	2
4	5	2	1	2	3	1	3	3
1	3	2	1	4	3	5	4	2
3	2	2	4	1	3	1	3	5
3	1	3	2	1	2	5	3	4
3	3	3	3	3	3	3	1	3
4	3	1	3	5	2	2	3	1
1	4	2	3	1	2	5	1	3
1	3	1	4	3	5	2	3	2
2	4	3	1	3	1	2	2	5
1	5	4	2	3	1	3	3	2
1	3	5	1	2	3	2	4	4
2	3	5	2	1	1	4	2	3
3	5	2	1	3	2	4	1	1
4	3	2	5	1	1	2	2	3
4	2	3	1	3	1	5	2	2
4	1	1	2	2	3	3	1	5
1	4	3	2	5	3	2	5	1
3	2	1	4	2	1	5	5	3
1	1	1	1	1	1	1	1	1
5	3	3	2	2	1	1	1	4
3	2	1	1	5	2	3	2	4
2	2	4	1	5	3	1	3	3
1	2	4	3	1	5	2	1	3
3	1	2	3	5	1	4	4	2
3	3	4	5	2	2	1	2	1
2	5	3	4	1	2	3	4	1
3	4	5	2	1	3	1	3	2
1	3	1	2	3	2	4	3	5
3	2	5	1	3	4	2	1	1
3	1	4	3	2	1	2	4	5
5	3	4	1	1	2	3	5	2
2	5	1	3	1	3	2	2	4
1	2	3	3	4	2	1	2	5
2	1	1	5	4	3	3	1	2
1	1	1	1	1	1	1	1	1
3	5	1	2	4	1	2	5	3
1	1	1	1	2	1	1	1	1
1	1	1	1	1	1	1	1	2

Table 9.2 Characteristics of the population (full sample n=300)

		Total (n=300)	Included (n=282)	Excluded (n=18)
Age in years (mean)		49.01	48.98	50.72
% Male		40.8	40.21	50
Employment (%)	employment or self-employment	51.33	52.48	33.33
	retired	29	28.37	38.89
	housework	7.33	6.74	16.67
	student	4	3.55	11.11
	seeking work	1.67	1.77	-
	unemployed	1.67	1.77	-
	long-term sick	3.67	3.9	-
	other	1.33	1.42	-
Highest level of education (%)	secondary school (left school at 16 or before)	51.01	50.36	61.11
	further education (left school at 18)	16.11	16.79	5.56
	higher education (university or college)	28.19	27.86	33.33
	post-graduate education	4.7	5	-
Ethnicity (%)	White	97.99	98.22	94.4
	Mixed/dual heritage	1.34	1.42	-
	Asian or Asian British	0.67	0.36	5.6
Difficulty with ranking exercise (%)	very difficult	14.48	14.7	11.11
	quite difficult	31.31	31.18	33.33
	neither difficult or easy	21.21	21.15	22.22
	fairly easy	26.94	26.88	27.78
	very easy	6.06	6.09	5.56
Difficulty with standard gamble exercise (%)	very difficult	6.35	6.05	11.11
	quite difficult	21.07	22.06	5.56
	neither difficult or easy	19.73	20.28	11.11
	fairly easy	43.48	43.42	44.44
	very easy	9.36	8.19	27.78
Understanding on ranking exercise (%)	fully understood the task	80.94	80.07	94.44
	partially understood the task	18.06	18.86	5.56
	did not really understand the task	1	1.07	-
Understanding on SG task (%)	fully understood the task	82.61	82.21	88.89
	partially understood the task	16.39	17.08	5.56
	did not really understand the task	1	0.71	5.56
Interviewer rating of respondents understanding	understood and performed tasks easily	65.65	66.67	50
	some problems but	28.91	28.99	27.78

of ranking task (%)	seemed to understand			
	doubtful whether the respondent understood	5.44	4.35	22.22
Interviewer rating of respondents understanding of standard gamble task (%)	understood and performed tasks easily	71.43	72.46	55.56
	some problems but seemed to understand	22.79	22.46	27.78
	doubtful whether the respondent understood	5.78	5.07	16.67
Effort and concentration of respondent on ranking (interviewer assessed)	Concentrated very hard and put a great deal of effort into it	40.82	42.39	16.67
	Concentrated fairly hard and put some effort into it	50	50.72	38.89
	Didn't concentrate very hard and put little effort into it	8.5	6.52	38.89
	Concentrated at the beginning but lost interest/concentration before reaching the end	0.68	0.36	5.56
Effort and concentration of respondent on standard gamble (interviewer assessed)	Concentrated very hard and put a great deal of effort into it	41.5	43.12	16.67
	Concentrated fairly hard and put some effort into it	52.04	52.17	50
	Didn't concentrate very hard and put little effort into it	6.46	4.71	33.33

Table 9.3 Descriptive statistics for health states

State	Observations	Mean	Median	Std. Dev.	Min	Max
111111112	38	0.9316	0.975	0.1039	0.55	1
111121111	34	0.9206	0.95	0.1027	0.55	1
112254323	38	0.4888	0.45	0.2669	0.05	0.975
115422323	37	0.7426	0.75	0.2378	0.1	0.975
122531334	32	0.7195	0.75	0.2130	0.1	0.975
123342125	34	0.5757	0.65	0.2422	0.05	0.975
123523451	34	0.6838	0.75	0.2198	0.1	0.975
124315213	39	0.6654	0.65	0.2585	0.05	1
131232435	34	0.6103	0.65	0.2417	0.05	0.975
131435232	33	0.6538	0.65	0.2391	0.1	1
132143542	36	0.6028	0.65	0.2547	0.05	0.975
135123244	34	0.7331	0.8	0.2431	0.1	1
142312513	34	0.5904	0.55	0.2946	0	0.975
143253251	33	0.6061	0.65	0.2373	0	0.95
153324122	34	0.7404	0.8	0.2318	0.1	1
154231332	34	0.7441	0.75	0.1868	0.35	0.975
211543312	37	0.6946	0.75	0.2359	0.25	0.975
212335154	37	0.7291	0.75	0.2200	0.1	0.975
213125433	34	0.6191	0.55	0.2224	0.05	0.975
214233521	37	0.7264	0.85	0.2498	0.1	0.975
222222212	68	0.7699	0.85	0.1873	0.15	1
224153133	37	0.6649	0.75	0.2726	0.1	0.975
225341331	34	0.6846	0.675	0.1818	0.15	0.975
231324531	34	0.6993	0.75	0.2234	0.15	0.975
232114355	34	0.6125	0.65	0.2372	0.05	0.975
233451112	37	0.7264	0.75	0.2516	0.05	0.975
235211423	33	0.6568	0.75	0.2946	0.1	1
241532143	34	0.7051	0.65	0.2099	0.1	0.975
243131225	37	0.6764	0.75	0.2552	0.05	0.975
251313224	34	0.7551	0.75	0.2196	0.1	1
253412341	34	0.6221	0.775	0.2309	0.15	0.975
312351442	32	0.6047	0.65	0.1981	0.25	0.95
313212534	36	0.6535	0.55	0.2794	0.05	0.975
313514232	33	0.7811	0.675	0.1918	0.15	1
314321245	33	0.5879	0.8	0.2509	0.05	0.95
321152324	34	0.5537	0.65	0.2481	0.05	0.975
321421553	37	0.6811	0.55	0.2539	0.1	0.975
322413135	34	0.6787	0.7	0.2497	0.05	0.975
325134211	34	0.7625	0.7	0.1933	0.15	0.975
332245121	34	0.5559	0.75	0.2327	0.05	0.975
333333313	72	0.7236	0.575	0.2315	0.05	0.975
334522121	38	0.7105	0.75	0.2378	0.05	1
341125322	37	0.6608	0.75	0.2605	0.1	1
345213132	35	0.7021	0.75	0.2372	0.1	1
351241253	38	0.5105	0.75	0.2615	-0.1	0.975
352132411	37	0.7338	0.55	0.2491	0.05	0.975
411223315	34	0.6890	0.85	0.2479	0.05	1
415332152	37	0.7561	0.725	0.2310	0.1	0.975
423131522	34	0.6662	0.85	0.2309	0.05	0.975
423215341	34	0.6221	0.65	0.2639	0.05	0.975
431352231	33	0.6280	0.65	0.2728	0.05	1
432511223	34	0.7419	0.65	0.2279	0.05	1

444444414	37	0.5824	0.75	0.2629	0.05	0.975
452123133	37	0.7081	0.65	0.2374	0.1	1
512433221	34	0.7640	34	0.1652	0.35	0.975
513142233	34	0.6699	34	0.2334	0.05	0.975
521234143	34	0.6316	0.75	0.2416	0	0.975
521313422	34	0.7750	0.65	0.1526	0.35	0.975
533221114	34	0.6985	0.75	0.2409	0.05	0.975
534112352	37	0.7716	0.75	0.2419	0.1	0.975
542321331	38	0.6283	0.85	0.2818	0	0.975
555555515	35	0.3879	0.65	0.3070	-0.1	1
555555555	235	0.3368	0.45	0.3154	-0.75	1

Table 9.4 Models

	Model			
Coefficient	OLS (1)	Random Effects (2)	Mean (3)	Rank (4)+
Constant	1	1	1	1
worry2	0.0058	0.0117	0.0082	0.0206**
worry3	0.0363**	0.0292**	0.0380**	0.0342**
worry4	**0.0261**	0.0313**	**0.0250**	0.0417**
worry5	**0.0312***	0.0344**	**0.0324**	0.0964**
sad2	0.0405**	0.0335**	0.0430**	0.0457**
sad3	0.0435**	0.0377**	0.0458**	**0.0386****
sad4	0.0780**	0.0677**	0.0772**	0.0717**
sad5	**0.0688****	**0.0677****	**0.0699****	**0.0613****
annoy2	0.0380**	0.0271**	0.0398**	0.0377**
annoy3	**0.0316****	**0.0265****	**0.0334***	0.0382**
annoy4	**0.0248**	**0.0217***	**0.0233**	**0.0372****
annoy5	**0.0243**	0.0335**	**0.0257**	0.0572**
tired2	0.0668**	0.0390**	0.0679**	0.0377**
tired3	**0.0397****	**0.0276****	**0.0402****	0.0380**
tired4	**0.0355****	**0.0271****	**0.0353***	**0.0304****
tired5	**0.0380****	**0.0199**	**0.0376***	**0.0287****
pain2	0.0394**	0.0434**	0.0418**	0.0637**
pain3	**0.0241***	**0.0285****	**0.0259**	**0.0409****
pain4	0.1236**	0.1301**	0.1216**	0.1035**
pain5	0.1471**	0.1504**	0.1475**	0.1135**
sleep2	0.0319**	0.0248**	0.0344**	0.0315**
sleep3	**0.0091**	**0.0176***	**0.0107**	0.0330**
sleep4	0.0489**	0.0543**	0.0476**	0.0678**
sleep5	0.0955**	0.0910**	0.0971**	0.0699**
daily2	0.03525**	0.0411**	0.0372**	0.0382**
daily3	0.0595**	0.0592**	0.0610**	**0.0358****
daily4	0.0685**	0.0803**	0.0677**	0.0620**
daily5	0.0969**	0.1022**	0.0990**	0.0963**
work2	0.0485**	0.0519**	0.0413**	0.0443**
work3	**0.0454****	**0.0457****	**0.0379****	0.0523**
work4	0.0842**	0.0801**	0.0770**	0.0756**
work5	**0.0507****	**0.0578****	**0.0458****	0.1039**
activ2	0.0115	0.0122	0.0128	0.0314**
activ3	0.0634**	0.0535**	0.0646**	0.0484**
activ4	**0.0422****	**0.0336****	**0.0415****	**0.0396****
activ5	0.1148**	0.1018**	0.1163**	0.0766**
Dead	-	-	-	1**
N	2478	2478	63	3000
Inconsistencies (after removing insignificant ones)	14 (10)	11 (10)	14(8)	8(8)
% within +/-0.1	100	98.4	100	90.5
% within +/-0.05	90.5	77.8	98.4	65.1
MAE	0.0261	0.0313	0.0263	0.0461
RMSE	0.0312	0.0397	0.0309	0.0573
T test	-0.944	-4.522**	-0.505	1.660

+ rescaled coefficients

**significant at p<0.05

*significant at p<0.1

Inconsistencies are shown in bold type

Table 9.5: Parsimonious consistent models

	Model				
Coefficient	OLS (5)	P>t	Coefficient	Mean (6)	P>t
Constant	1		Constant	1	
worry2345	0.0227	0.047	worry2345	0.0251	0.082
sad2	0.0420	0.003	sad2	0.0438	0.018
sad3	0.0445	0.002	sad3	0.0460	0.013
sad45	0.0722	0	sad45	0.0728	0
annoy2345	0.0313	0.006	annoy2345	0.0326	0.025
tired2345	0.0479	0	tired2345	0.0482	0.001
pain23	0.0332	0.004	pain23	0.0349	0.02
Pain4	0.1245	0	Pain4	0.1225	0
Pain5	0.1426	0	Pain5	0.1461	0
sleep23	0.0212	0.08	sleep234	0.0280	0.059
sleep4	0.0506	0.004			
sleep5	0.0907	0	sleep5	0.0952	0
daily2	0.0371	0.009	daily2	0.0379	0.039
daily3	0.0612	0	daily3	0.0612	0.001
daily4	0.0699	0	daily4	0.0682	0.003
daily5	0.0930	0	daily5	0.0971	0
work23	0.0487	0	work23	0.0403	0.016
work45	0.0656	0	work45	0.0609	0.002
activ234	0.0368	0.001	activ234	0.0376	0.01
activ5	0.1079	0	activ5	0.1129	0
N	2478			63	
Inconsistencies	0			0	
% within +/-0.1	98.41			98.41	
% within +/-0.05	73.02			76.19	
MAE	0.0343			0.0349	
RMSE	0.0426			0.0431	
T test	-0.770			-0.336	

Figure 9.1: Distribution of utility values

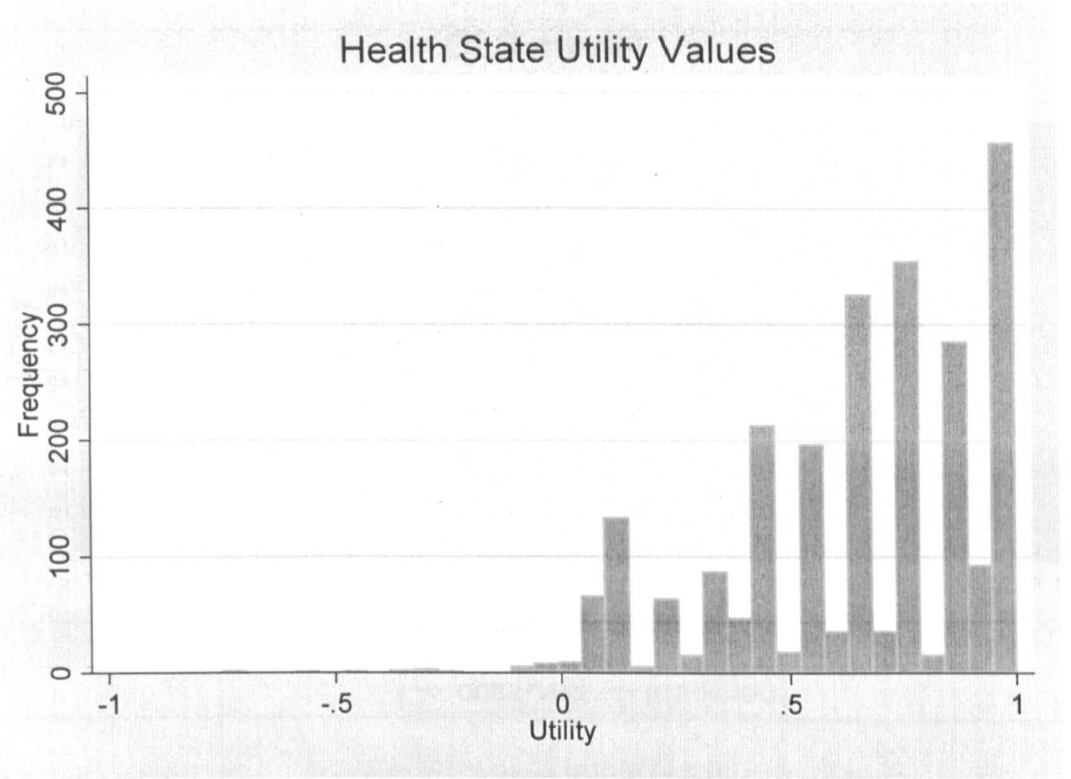


Figure 9.2 Observed and predicted values (OLS model)

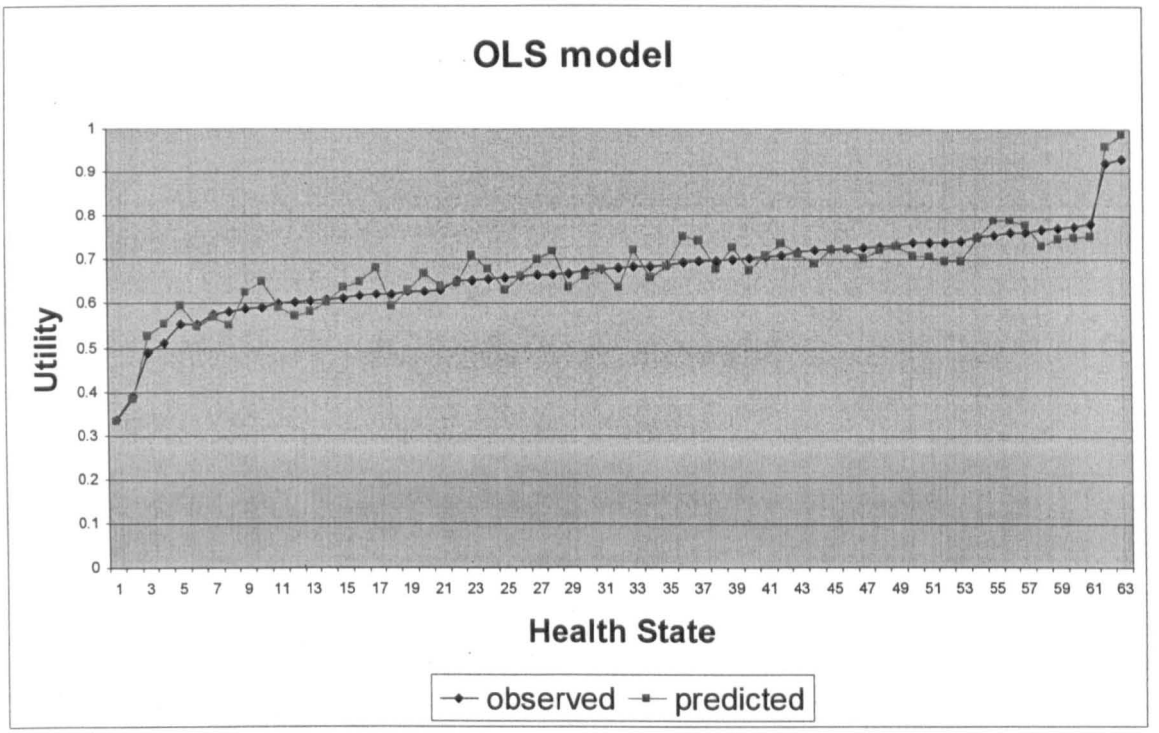


Figure 9.3 Observed and predicted values (RE model)

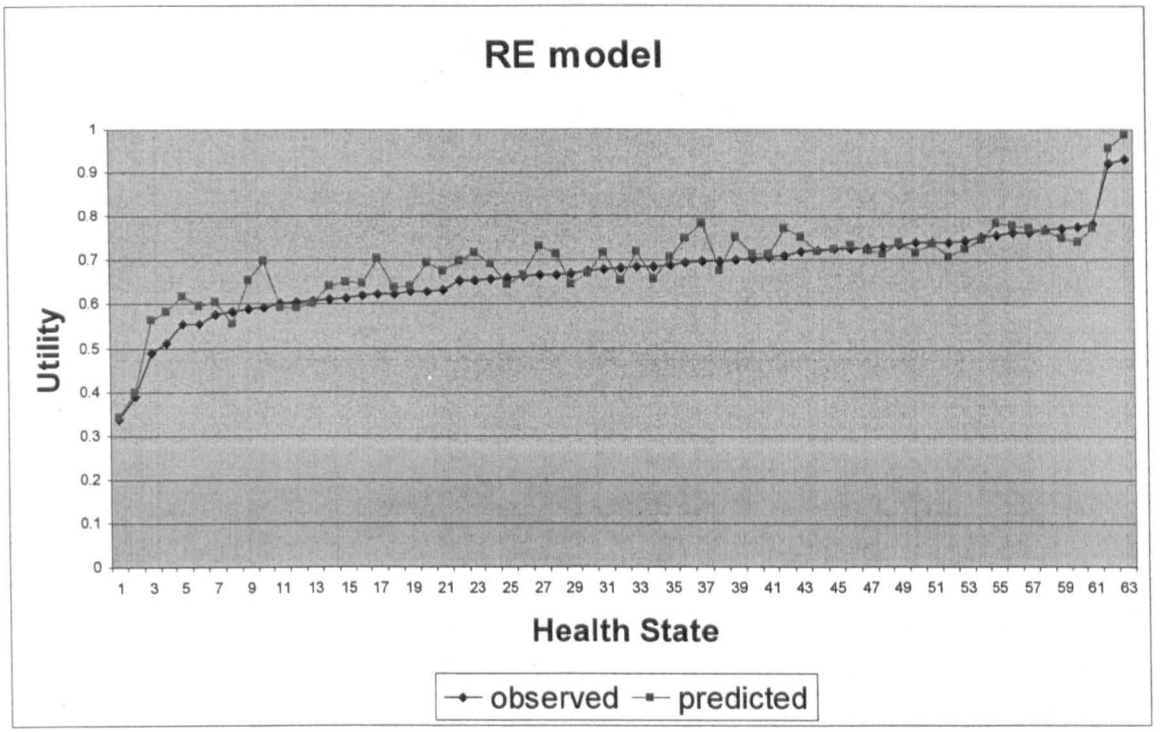


Figure 9.4 Observed and predicted values (mean model)

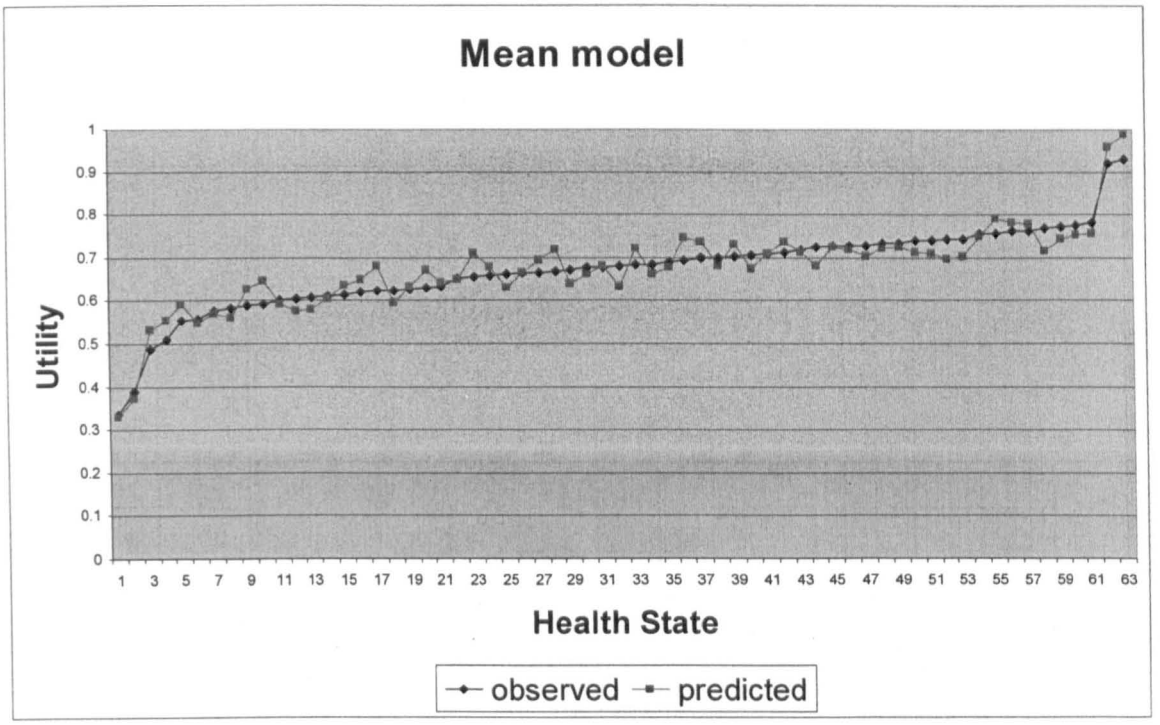


Figure 9.5 Observed and predicted values (rank model)

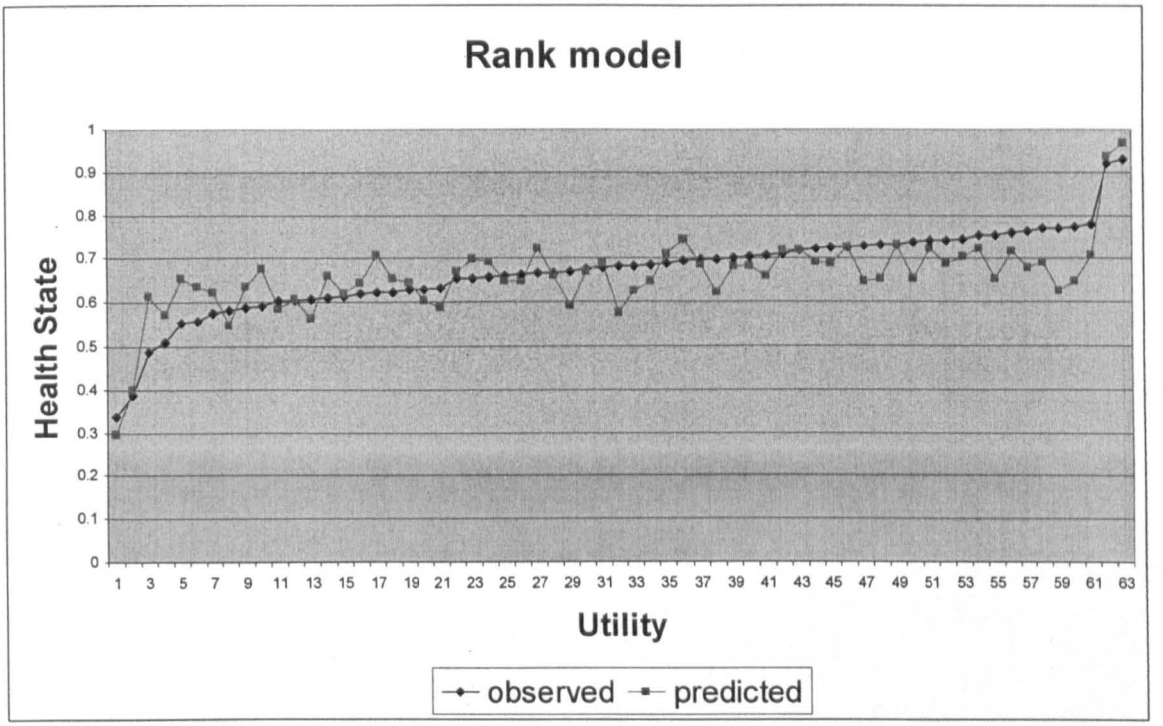


Figure 9.6 Observed and predicted values (OLS parsimonious model)

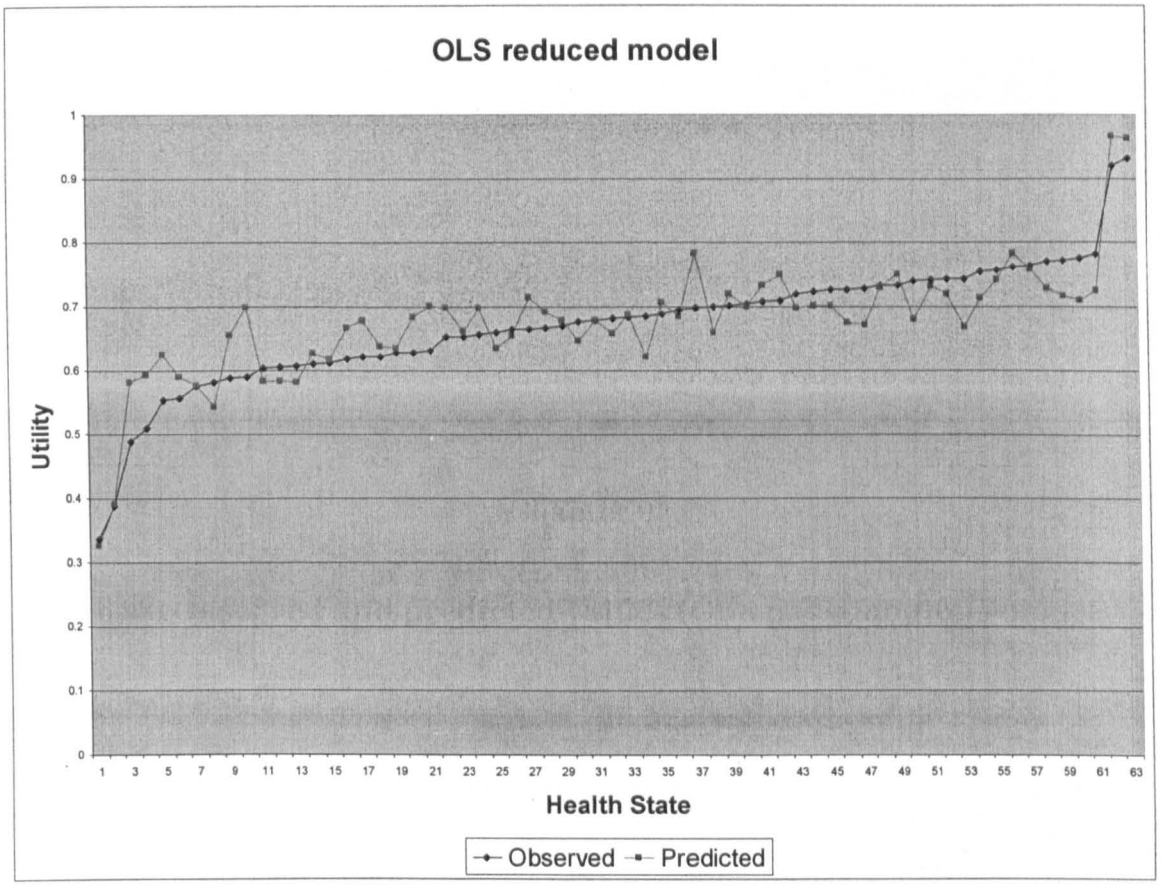
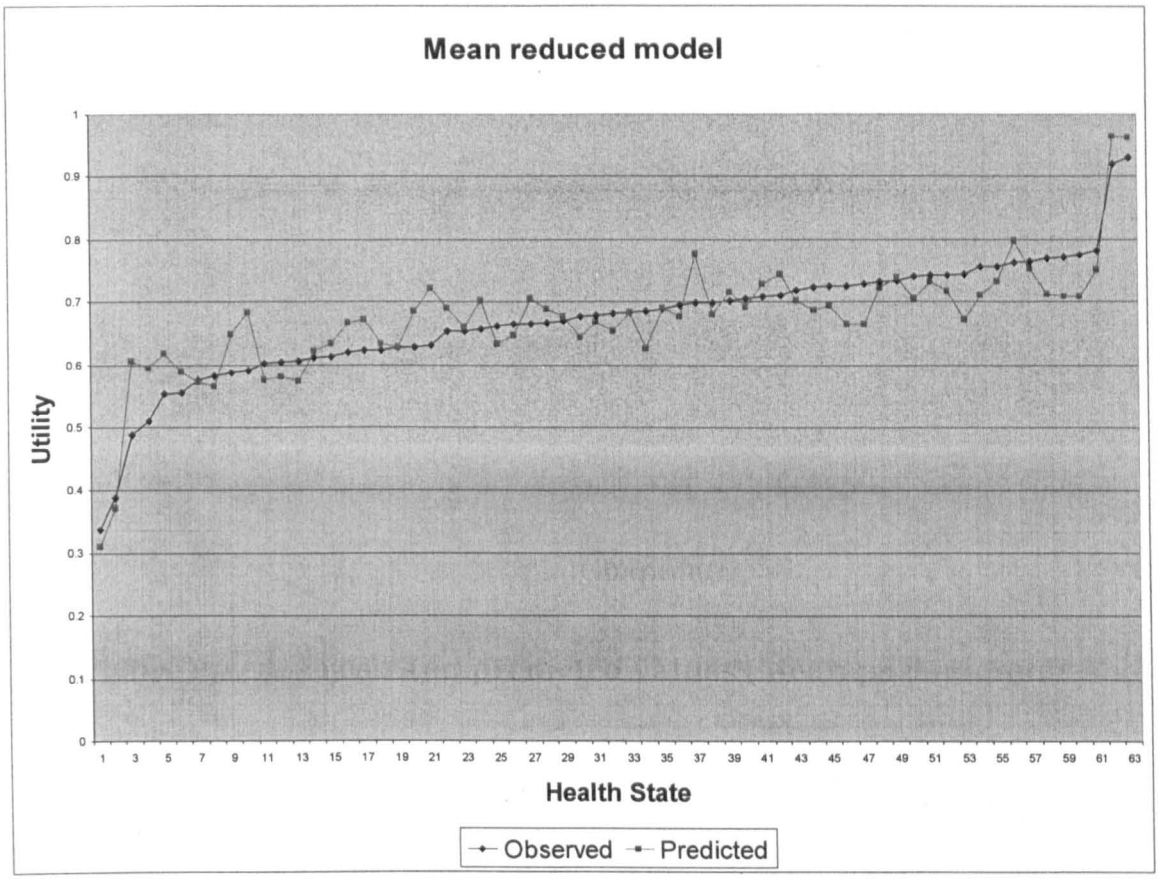


Figure 9.7 Observed and predicted values (Mean parsimonious model)



Appendix 9.1: Letter to Respondents

The
University
Of
Sheffield.

Sheffield Hallam University

SHARPENS YOUR THINKING

Centre for Research and Evaluation

Sheffield Hallam University
Howard Street
Sheffield S1 1WB
UK

Telephone +44 (0)114 225 5185
Fax +44 (0) 114 2255186
E-mail cre@shu.ac.uk
www.shu.ac.uk

Winter/Spring 2008

Dear Resident

I am writing to ask for your help with an important piece of research being carried out on behalf of the School of Health And Related Research (ScHARR) at the University of Sheffield. This study is being done to understand the different ways people value health and illness.

The information you provide is very important. It will inform future health policies. ScHARR have asked the Centre for Research and Evaluation at Sheffield Hallam University to undertake the interviewing for this research.

Addresses in your area have been randomly selected. An interviewer may call at your address. If you are at home when we call, we will ask for your help and provide further information on the research. You will be under no obligation to take part in this research. The interview will take about 30 minutes of your time and will take place in your home at your convenience. Any information you provide will be treated in the strictest confidence.

If you have any queries about the interview please contact either Anna Stevens on 0114 225 4656 or Dot Biggin on 0114 225 5185. If you would like any further information about the research please contact Katherine Stevens on 0114 222 0841.

I would like to thank you in advance for your co-operation and participation in this important project.

Yours sincerely

Anna Stevens
Project Manager

Appendix 9.2: Information Leaflet

The University Of Sheffield.

Sheffield Hallam University

SHARPENS YOUR THINKING

Information sheet for health preferences survey

You are being invited to take part in a research project. Before you decide whether to take part it is important for you to understand why the research is being done and what it will involve. Please take time to read the following information carefully and discuss it with others if you wish. Take time to decide whether or not you wish to take part. Thank you for reading this.

Research project title:

Health preferences survey

What is the project's purpose?
This study is being done to understand the different ways people value health and illness. This work is important for understanding and comparing the benefit of different treatments. It may be used to help the NHS decide which treatments to fund.

Why have I been chosen?
Your address has been randomly selected using the Postcode Address File (PAF) register of addresses in the UK. Your address was as likely to be chosen as any other address. There will be about 300 people participating in this study.

Do I have to take part?
It is up to you to decide whether or not to take part. Refusal to take part will not affect you in any way.

What will happen to me if I take part?
If you agree to participate, one face-to-face interview will be arranged. The interview will take about 30 minutes and will take place in your home at your convenience. The interview will consist of three parts. In the first part the interviewer will ask you to complete a self-completion questionnaire about your general health. In the second part you will be asked to do a ranking and choice exercise about 9 health states and how bad you think they would be. The third part is a brief questionnaire about your background characteristics.

Will my taking part in this project be kept confidential?
All information which is collected about you during the course of the research will be kept strictly confidential. The information you give will not be used in any way that could identify you.

What will happen to the results of the research project?
The results of this study will be published in academic journals. You will not be able to be identified in any reports or publications. If you would like a copy of the results please contact the research team.

Who is organising and funding the research?
The Centre for Research and Evaluation at Sheffield Hallam University are undertaking the interviews on behalf of the School of Health and Related Research, (ScHARR) at the University of Sheffield. The project has been funded by the UK Medical Research Council.

Who has ethically reviewed the project?
This research project has been reviewed by external independent academic researchers at the MRC for scientific aspects, and by ScHARR at the University of Sheffield and by the University of Sheffield Research Ethics Committee for ethical aspects.

If you have any questions about the interview, please contact:
Anna Stevens or Dot Biggin
Centre for Research and Evaluation
Sheffield Hallam University
Telephone 0114 225 4656 (Anna) or 0114 225 5185 (Dot)

If you have any questions about the research, please contact:
Katherine Stevens
School of Health and Related Research
The University of Sheffield
Telephone 0114 222 0841

Thank you very much for reading this sheet.

You will be given a copy of this sheet and, if you agree to participate, a signed participant consent form to keep.

Appendix 9.3: Consent Form

The University Of Sheffield.

Sheffield Hallam University

SHARPENS YOUR THINKING

Participant Consent Form

Title of Project: Health preferences survey

Name of Researcher: Katherine Stevens, The University of Sheffield

Respondent ID for this project: ☐

Please initial box

1. I confirm that I have read and understood the information sheet dated 17/12/2007 for the above project and have had the opportunity to ask questions. ☐

2. I understand that my participation is voluntary and that I am free to withdraw at any time without giving any reason. ☐

3. I understand that my responses will be anonymised before analysis. I give permission for members of the research team to have access to my anonymised responses. ☐

4. I agree to take part in the above research project. ☐

Name of Participant	Date	Signature
Name of person taking consent	Date	Signature

If you would like any further information about the research please contact Katherine Stevens on 0114 222 0841.

Copies: One copy for the participant and one copy for the Principal Investigator / Researcher.

Appendix 9.4 Descriptive system

Dimension	Level	Description
Worried	1	I don't feel worried today
	2	I feel a little bit worried today
	3	I feel a bit worried today
	4	I feel quite worried today
	5	I feel very worried today
Sad	1	I don't feel sad today
	2	I feel a little bit sad today
	3	I feel a bit sad today
	4	I feel quite sad today
	5	I feel very sad today
Annoyed	1	I don't feel annoyed today
	2	I feel a little bit annoyed today
	3	I feel a bit annoyed today
	4	I feel quite annoyed today
	5	I feel very annoyed today
Tired	1	I don't feel tired today
	2	I feel a little bit tired today
	3	I feel a bit tired today
	4	I feel quite tired today
	5	I feel very tired today
Pain	1	I don't have any pain today
	2	I have a little bit of pain today
	3	I have a bit of pain today
	4	I have quite a lot of pain today
	5	I have a lot of pain today
Sleep	1	Last night I had no problems sleeping
	2	Last night I had a few problems sleeping
	3	Last night I had some problems sleeping
	4	Last night I had many problems sleeping
	5	Last night I couldn't sleep at all
Daily routine	1	I have no problems with my daily routine today
	2	I have a few problems with my daily routine today
	3	I have some problems with my daily routine today
	4	I have many problems with my daily routine today
	5	I can't do my daily routine today
Work	1	I have no problems with my work today
	2	I have a few problems with my work today
	3	I have some problems with my work today
	4	I have many problems with my work today
	5	I can't do my work today
Able to join in activities	1	I can join in with any activities today
	2	I can join in with most activities today
	3	I can join in with some activities today
	4	I can join in with a few activities today
	5	I can join in with no activities today

Appendix 9.5: Example Health State

Health State 153324122

I don't feel worried

I feel very sad

I feel a bit annoyed

I feel a bit tired

I have a little bit of pain

I have many problems sleeping

I have no problems with my daily routine

I have a few problems with my work

I can join in with most activities

The University Of Sheffield.

SCHOOL OF HEALTH AND RELATED RESEARCH

THE UNIVERSITY OF SHEFFIELD

Health Preferences Survey

INTERVIEWING SCRIPT

Thank you very much for agreeing to take part in this survey. As we explained in the letter, this is a survey for the University of Sheffield about the way people value health and illness.

All information you provide is confidential. The information you give will not be used in any way that could identify you.

We are interested in people's views, and there are no right or wrong answers. Please tell us what you think.

Before we start the first exercise, please could you fill in this short questionnaire about your health in general.

GIVE THE RESPONDENT THE HEALTH QUESTIONNAIRE AT THE START OF THE ANSWER BOOKLET. WHEN THEY HAVE COMPLETED IT, TAKE THE ANSWER BOOK BACK AND PUT THE START TIME ON THE FRONT COVER.

Appendix 9.6 Valuation Script

A: Ranking exercise

Thank you. We will now move on to the first exercise which asks you to think about some health states from the questionnaire you have just filled in.

HAND RESPONDENT ENVELOPE

In a moment I will ask you to place the cards in order of how good or bad you think they are. I would like you to imagine that you yourself are actually in each state of health and that it is going to last for the rest of your life without changing. Please read each card carefully to see exactly what the health state is and how it differs from the others. When you have finished reading through, please place the cards in order of how good or bad you think they are. Put the one you think is best at the top ***(POINT)*** *and the one that you think it worst at the bottom* ***(POINT).*** *If you think two health states are equal, put them side by side. You will notice that there is a card which says "Dead". Please also put this in the order in the place where you think it belongs.*

WHEN RESPONDENT HAS FINISHED, RECORD THEIR RANKING IN THE ANSWER BOOK (PAGE 4), REMOVE THE BLUE DEAD CARD (IMD) AND CARD Z AND PUT THEM BACK IN THE ENVELOPE, THEN SHUFFLE THE REMAINING CARDS AND PUT THEM IN A PILE.

Choice Exercise – Example

HAVE WHITE CARD X, PINK CARD PERFECT HEALTH AND BLUE CARD IMD READY FOR USE IN THE EXAMPLE EXERCISE.
I am now going to ask you to make choices between these same health states that you just placed in order. To make the task easier to understand we will use an aid similar to a game board.

PLACE CHANCE BOARD ON TABLE, SET WINDOWS TO 90/10

X1. *Using this board, you will be asked to pick either Choice A* **(POINT TO CHOICE A)** *or Choice B* **(POINT TO CHOICE B)**.

Choice B, at the bottom of the board, will describe a state of health. Here is an example:

PLACE WHITE CARD 'X' IN POCKET OF CHOICE B

If you choose to go for Choice B, you are 100% certain to be in the health state described on the card in this pocket **(POINT TO CARD X)** *for the rest of your life.*

But Choice A is less straight forward, as it is a treatment which doesn't always work. If the treatment does work, you will be in the health state shown on this pink card.

PLACE THE PINK 'PH' CARD IN THE LEFT POCKET OF CHOICE A

However, if the treatment does not work, you will be in the health state shown on this blue card.

PLACE THE BLUE 'IMD' CARD IN THE RIGHT POCKET OF CHOICE A

Therefore, if you choose to go for Choice A, there are two possible results.

The chances of each of these results occurring are shown by the numbers appearing in the windows above each pocket **(POINT TO THE WINDOWS)** *and by the amount of pink and blue inside the circle* **(POINT TO THE CIRCLE)**.

So, for example, at the moment it shows that there is a 90% chance of the treatment working and a 10% chance of the treatment not working.

Another way of explaining the chance aspect of Choice A is that for every 100 patients who choose Choice A, 90 will experience the health state on the left following treatment, **(POINT TO 90)** *but 10 will experience the health state on the right* **(POINT TO 10)**

No-one will know before choosing whether they will be one of the 90 or one of the 10. That is the chance they take.

During the interview, these chances will change and I will ask you to choose Choice A or Choice B each time I change the chances.

DEMONSTRATE BY TURNING THE WHEEL ON THE CHANCE BOARD TO 20/80

Before we start looking at other health states, would you like me to explain how the Chance Board works again?

Yes	REPEAT FROM X1
No	REMOVE CARDS FROM CHANCE BOARD AND PUT EXAMPLE CARD AWAY. KEEP PERFECT HEALTH AND IMD HANDY AND GO TO CHOICE INSTRUCTIONS

CHOICE INSTRUCTIONS

TAKE THE NEXT CARD ON TOP OF THE PILE.

IF IT WAS RANKED **BETTER** THAN DEAD GO TO **B** (PAGE 5)

IF IT WAS RANKED **WORSE** THAN DEAD, GO TO **C** (PAGE 9)

(WHEN ALL CARDS ARE DONE PUT THEM BACK IN THE ENVELOPE AND GO TO **D** IN THE ANSWER BOOKLET)

B: Choice Exercise – Better than Dead

WRITE THE CARD IDENTIFIER (FOUND ON THE BACK OF THE CARD) IN THE ANSWER BOOKLET IN PART B

When I asked you to place all the cards in order earlier, you judged that this health state **(SHOW RESPONDENT CARD)** *was better than being dead.*

Do you still think this is true?

Yes	GO TO B1	
No	GO TO BD	
Can't Decide	CODE AND GO TO CHOICE INSTRUCTIONS	A

BD *Do you think that this health state is worse than being dead?*

Yes	GO TO C	B
No	CODE AND GO TO CHOICE INSTRUCTIONS	C
Can't Decide	CODE AND GO TO CHOICE INSTRUCTIONS	D

B1. *Please read over the description of this card again.*
PLACE PINK PERFECT HEALTH IN LEFT HAND POCKET OF CHOICE A AND BLUE IMD IN RIGHT HAND POCKET OF CHOICE A.
SET THE WHEEL TO 100 ON THE LEFT AND 0 ON THE RIGHT.
WHEN RESPONDENT HAS FINISHED READING, PLACE HEALTH STATE CARD IN POCKET OF CHOICE B AND SAY:
As you can see Choice A is a 100% chance of being in the health state described on the pink card, with zero chance of being in the health state shown on the blue card.
Choice B is a 100% chance of being in the health state described on the yellow card.
Remember whichever choice you make you will be in the health state you end up in for the rest of your life. Would you prefer Choice A or Choice B now?

A	GO TO B2	
B	GO TO B13	V
Can't Decide	CODE AND GO TO CHOICE INSTRUCTIONS	100

B2. SET THE WHEEL TO 10 ON THE LEFT AND 90 ON THE RIGHT.
Choice A is now a 10% chance of the health state described on the pink card, with a 90% chance of being in the health state described on the blue card.
Choice B is still a 100% chance of the health state described on the yellow card.
Would you prefer Choice A or Choice B now?

A	GO TO B3	
B	GO TO B4	
Can't Decide	CODE AND GO TO CHOICE INSTRUCTIONS	10

B3. PLACE COVER 1 OVER CHOICE A OF THE CHANCE BOARD

Suppose now that Choice A was a zero chance of the health state described on the pink card, with a 100% chance of being in the health state described on the blue card. Choice B is still a 100% chance of the health state described on the yellow card.
Would you prefer Choice A or Choice B now?

A	GO TO B14	M
B	CODE AND GO TO CHOICE INSTRUCTIONS	5
Can't Decide	CODE AND GO TO CHOICE INSTRUCTIONS	0

B4. SET THE WHEEL TO 90 ON THE LEFT AND 10 ON THE RIGHT.

Choice A is now a 90% chance of the health state described on the pink card, with a 10% chance of being in the health state described on the blue card.
Choice B is still a 100% chance of the health state described on the yellow card.
Would you prefer Choice A or Choice B now?

A	GO TO B6	
B	GO TO B5	
Can't Decide	CODE AND GO TO CHOICE INSTRUCTIONS	90

B5. PLACE COVER 2 OVER CHOICE A OF THE CHANCE BOARD

Suppose now that Choice A was a 95% chance of the health state described on the pink card, with a 5% chance of being in the health state described on the blue card. Choice B is still a 100% chance of the health state described on the yellow card. Would you prefer Choice A or Choice B now?

A	CODE AND GO TO CHOICE INSTRUCTIONS	92.5
B	CODE AND GO TO CHOICE INSTRUCTIONS	97.5
Can't Decide	CODE AND GO TO CHOICE INSTRUCTIONS	95

B6. SET THE WHEEL TO 20 ON THE LEFT AND 80 ON THE RIGHT.

Choice A is now a 20% chance of the health state described on the pink card, with an 80% chance of being in the health state described on the blue card.
Choice B is still a 100% chance of the health state described on the yellow card.
Would you prefer Choice A or Choice B now?

A	CODE AND GO TO CHOICE INSTRUCTIONS	15
B	GO TO B7	
Can't Decide	CODE AND GO TO CHOICE INSTRUCTIONS	20

B7. SET THE WHEEL TO 80 ON THE LEFT AND 20 ON THE RIGHT.

Choice A is now an 80% chance of the health state described on the pink card, with a 20% chance of being in the health state described on the blue card.
Choice B is still a 100% chance of the health state described on the yellow card.
Would you prefer Choice A or Choice B now?

A	GO TO B8	
B	CODE AND GO TO CHOICE INSTRUCTIONS	85
Can't Decide	CODE AND GO TO CHOICE INSTRUCTIONS	80

B8. SET THE WHEEL TO 30 ON THE LEFT AND 70 ON THE RIGHT.

Choice A is now a 30% chance of the health state described on the pink card, with a 70% chance of being in the health state described on the blue card.
Choice B is still a 100% chance of the health state described on the yellow card.
Would you prefer Choice A or Choice B now?

A	CODE AND GO TO CHOICE INSTRUCTIONS	25
B	GO TO B9	
Can't Decide	CODE AND GO TO CHOICE INSTRUCTIONS	30

B9. SET THE WHEEL TO 70 ON THE LEFT AND 30 ON THE RIGHT.

Choice A is now a 70% chance of the health state described on the pink card, with a 30% chance of being in the health state described on the blue card.
Choice B is still a 100% chance of the health state described on the yellow card.
Would you prefer Choice A or Choice B now?

A	GO TO B10	
B	CODE AND GO TO CHOICE INSTRUCTIONS	75
Can't Decide	CODE AND GO TO CHOICE INSTRUCTIONS	70

B10. SET THE WHEEL TO 40 ON THE LEFT AND 60 ON THE RIGHT.

Choice A is now a 40% chance of the health state described on the pink card, with a 60% chance of being in the health state described on the blue card.
Choice B is still a 100% chance of the health state described on the yellow card.
Would you prefer Choice A or Choice B now?

A	CODE AND GO TO CHOICE INSTRUCTIONS	35
B	GO TO B11	
Can't Decide	CODE AND GO TO CHOICE INSTRUCTIONS	40

B11. SET THE WHEEL TO 60 ON THE LEFT AND 40 ON THE RIGHT.
Choice A is now a 60% chance of the health state described on the pink card, with a 40% chance of being in the health state described on the blue card.
Choice B is still a 100% chance of the health state described on the yellow card.
Would you prefer Choice A or Choice B now?

A	GO TO B1.12	
B	CODE AND GO TO CHOICE INSTRUCTIONS	65
Can't Decide	CODE AND GO TO CHOICE INSTRUCTIONS	60

B12. SET THE WHEEL TO 50 ON THE LEFT AND 50 ON THE RIGHT.
Choice A is now a 50% chance of the health state described on the pink card, with a 50% chance of being in the health state described on the blue card.
Choice B is still a 100% chance of the health state described on the yellow card.
Would you prefer Choice A or Choice B now?

A	CODE AND GO TO CHOICE INSTRUCTIONS	45
B	CODE AND GO TO CHOICE INSTRUCTIONS	55
Can't Decide	CODE AND GO TO CHOICE INSTRUCTIONS	50

B13. *Why did you choose a 100% chance of the health state on the yellow card rather than a 100% chance of the health state on the pink card?*
RECORD VERBATIM RESPONSE

_____

_____

_____

GO TO CHOICE INSTRUCTIONS

B14. *Why did you choose a 100% chance of the health state on the blue card rather than a 100% chance of the health state on the yellow card?*
RECORD VERBATIM RESPONSE

_____

_____

_____

GO TO CHOICE INSTRUCTIONS

C: Choice Exercise - Worse Than Dead

WRITE THE CARD IDENTIFIER (FOUND ON THE BACK OF THE CARD) IN THE ANSWER BOOKLET IN PART C

When I asked you to place all the cards in order earlier, you judged that this health state was worse than being dead. Do you still think this is true?

Yes	GO TO C3	
No	GO TO CD	
Can't Decide	CODE AND GO TO CHOICE INSTRUCTIONS	E

CD *Do you think that this health state is better than being dead?*

Yes	GO TO B	F
No	CODE AND GO TO CHOICE INSTRUCTIONS	G
Can't Decide	CODE AND GO TO CHOICE INSTRUCTIONS	H

C3. PLACE HEALTH STATE CARD IN RIGHT POCKET OF CHOICE A
PLACE PINK PERFECT HEALTH CARD IN LEFT POCKET OF CHOICE A
PLACE BLUE IMD CARD IN POCKET OF CHOICE B
MOVE CHANCE BOARD DIAL TO 100/0

As you can see Choice A is a 100% chance of being in the health state described on the pink card, with a zero chance of being in the health state shown on the yellow card.
Choice B is a 100% chance of dead.
Would you prefer Choice A or Choice B now?

A	GO TO C4	
B	GO TO C15	
Can't Decide	CODE AND GO TO CHOICE INSTRUCTIONS	K

C4. SET THE WHEEL TO 10 ON THE LEFT AND 90 ON THE RIGHT.
With the chances set to 10% (for the pink card) and 90% (for the yellow card), would you prefer Choice A or Choice B now?

A	GO TO C5	
B	GO TO C6	
Can't Decide	CODE AND GO TO CHOICE INSTRUCTIONS	-10

C5. PLACE COVER CARD 1 OVER CHOICE A OF THE CHANCE BOARD

With the chances set to zero (for the pink card) and 100% (for the yellow card), would you prefer Choice A or Choice B now?

A	GO TO C16	
B	CODE AND GO TO CHOICE INSTRUCTIONS	-5
Can't Decide	CODE AND GO TO CHOICE INSTRUCTIONS	0

C6. SET THE WHEEL TO 90 ON THE LEFT AND 10 ON THE RIGHT

With the chances set to 90% (for the pink card) and 10% (for the yellow card), would you prefer Choice A or Choice B now?

A	GO TO C8	
B	GO TO C7	
Can't Decide	CODE AND CODE AND GO TO CHOICE INSTRUCTIONS	-90

C7. PLACE COVER CARD 2 OVER CHOICE A OF THE CHANCE BOARD

With the chances set to 95% (for the pink card) and 5% (for the yellow card), would you prefer Choice A or Choice B now?

A	CODE AND GO TO CHOICE INSTRUCTIONS	-92.5
B	CODE AND GO TO CHOICE INSTRUCTIONS	-97.5
Can't Decide	CODE AND GO TO CHOICE INSTRUCTIONS	-95

C8. SET THE WHEEL TO 20 ON THE LEFT AND 80 ON THE RIGHT

With the chances set to 20% (for the pink card) and 80% (for the yellow card), would you prefer Choice A or Choice B now?

A	CODE AND GO TO CHOICE INSTRUCTIONS	-15
B	GO TO C9	
Can't Decide	CODE AND GO TO CHOICE INSTRUCTIONS	-20

C9. SET THE WHEEL TO 80 ON THE LEFT AND 20 ON THE RIGHT

With the chances set to 80% (for the pink card) and 20% (for the yellow card), would you prefer Choice A or Choice B now?

A	GO TO C10	
B	CODE AND GO TO CHOICE INSTRUCTIONS	-85
Can't Decide	CODE AND GO TO CHOICE INSTRUCTIONS	-80

C10. SET THE WHEEL TO 30 ON THE LEFT AND 70 ON THE RIGHT.
With the chances set to 30% (for the pink card) and 70% (for the yellow card), would you prefer Choice A or Choice B now?

A	CODE AND GO TO CHOICE INSTRUCTIONS	-25
B	GO TO C11	
Can't Decide	CODE AND GO TO CHOICE INSTRUCTIONS	-30

C11. SET THE WHEEL TO 70 ON THE LEFT AND 30 ON THE RIGHT.
With the chances set to 70% (for the pink card) and 30% (for the yellow card), would you prefer Choice A or Choice B now?

A	GO TO C12	
B	CODE AND GO TO CHOICE INSTRUCTIONS	-75
Can't Decide	CODE AND GO TO CHOICE INSTRUCTIONS	-70

C12. SET THE WHEEL TO 40 ON THE LEFT AND 60 ON THE RIGHT.
With the chances set to 40% (for the pink card) and 60% (for the yellow card), would you prefer Choice A or Choice B now?

A	CODE AND GO TO CHOICE INSTRUCTIONS	-35
B	GO TO C13	
Can't Decide	CODE AND GO TO CHOICE INSTRUCTIONS	-40

C13. SET THE WHEEL TO 60 ON THE LEFT AND 40 ON THE RIGHT.
With the chances set to 60% (for the pink card) and 40% (for the yellow card), would you prefer Choice A or Choice B now?

A	GO TO C14	
B	CODE AND GO TO CHOICE INSTRUCTIONS	-65
Can't Decide	CODE AND GO TO CHOICE INSTRUCTIONS	-60

C14. SET THE WHEEL TO 50 ON THE LEFT AND 50 ON THE RIGHT
With the chances set to 50% (for the pink card) and 50% (for the yellow card), would you prefer Choice A or Choice B now?

A	CODE AND GO TO CHOICE INSTRUCTIONS	-45
B	CODE AND GO TO CHOICE INSTRUCTIONS	-55
Can't Decide	CODE AND GO TO CHOICE INSTRUCTIONS	-50

C15. *Why did you choose a 100% chance of dead rather than a 100% chance of the health state on the pink card?*
RECORD VERBATIM RESPONSE

_____

_____

_____

GO TO CHOICE INSTRUCTIONS

C16. *Why did you choose a 100% chance of the health state on the yellow card rather than a 100% chance of dead?*
RECORD VERBATIM RESPONSE

_____

_____

_____

GO TO CHOICE INSTRUCTIONS

Appendix 9.7 Models with interactions

OLS main and interactions model with a constant

utility	Coef.	Std. Err.	t	P>t	[95% Conf.	Interval]
worry2	0.005872	0.014732	0.4	0.69	-0.02302	0.03476
worry3	-0.02493	0.014719	-1.69	0.09	-0.05379	0.003937
worry4	-0.01509	0.018011	-0.84	0.402	-0.05041	0.020223
worry5	-0.01864	0.018223	-1.02	0.306	-0.05438	0.017091
sad2	-0.02881	0.014746	-1.95	0.051	-0.05773	0.000107
sad3	-0.0322	0.014711	-2.19	0.029	-0.06104	-0.00335
sad4	-0.06692	0.017875	-3.74	0	-0.10197	-0.03186
sad5	-0.05587	0.01815	-3.08	0.002	-0.09146	-0.02028
annoy2	-0.02635	0.014752	-1.79	0.074	-0.05528	0.002574
annoy3	-0.02021	0.014782	-1.37	0.172	-0.04919	0.008779
annoy4	-0.01296	0.017794	-0.73	0.466	-0.04786	0.021929
annoy5	-0.0117	0.018241	-0.64	0.521	-0.04747	0.024075
tired2	-0.05527	0.014709	-3.76	0	-0.08411	-0.02642
tired3	-0.02808	0.014692	-1.91	0.056	-0.05689	0.000728
tired4	-0.02405	0.017898	-1.34	0.179	-0.05915	0.011043
tired5	-0.02507	0.018256	-1.37	0.17	-0.06087	0.01073
pain2	-0.02197	0.015306	-1.44	0.151	-0.05199	0.008044
pain3	-0.00967	0.015146	-0.64	0.523	-0.03937	0.020035
pain4	-0.10926	0.018256	-5.99	0	-0.14506	-0.07346
pain5	-0.13125	0.018468	-7.11	0	-0.16746	-0.09504
sleep2	-0.02028	0.014732	-1.38	0.169	-0.04917	0.008607
sleep3	0.002247	0.01472	0.15	0.879	-0.02662	0.031113
sleep4	-0.03776	0.018064	-2.09	0.037	-0.07318	-0.00234
sleep5	-0.08233	0.018192	-4.53	0	-0.118	-0.04666
daily2	-0.02341	0.014748	-1.59	0.113	-0.05233	0.005507
daily3	-0.0481	0.014633	-3.29	0.001	-0.0768	-0.01941
daily4	-0.05693	0.01809	-3.15	0.002	-0.09241	-0.02146
daily5	-0.08416	0.018112	-4.65	0	-0.11968	-0.04864
work2	-0.02081	0.017811	-1.17	0.243	-0.05574	0.014119
work3	-0.0177	0.017873	-0.99	0.322	-0.05274	0.017352
work4	-0.05655	0.020792	-2.72	0.007	-0.09732	-0.01578
work5	-0.02639	0.020495	-1.29	0.198	-0.06658	0.013797
activ2	0.00701	0.015417	0.45	0.649	-0.02322	0.037241
activ3	-0.04852	0.015097	-3.21	0.001	-0.07813	-0.01892
activ4	-0.02765	0.018425	-1.5	0.134	-0.06378	0.008479
activ5	-0.09882	0.018558	-5.32	0	-0.13521	-0.06243
most	-0.03737	0.023706	-1.58	0.115	-0.08386	0.009114
least	0.024656	0.048917	0.5	0.614	-0.07127	0.12058
_cons	0.908408	0.060031	15.13	0	0.790692	1.026124

Source	SS	df	MS	Number of obs	=	2478
				F(38, 2439)	=	17.88
Model	40.8304	38	1.074484	Prob > F	=	0
Residual	146.5818	2439	0.060099	R-squared	=	0.2179
				Adj R-squared	=	0.2057

Total	187.4122	2477	0.075661	Root MSE	=	0.24515

Fixed effects model, main effects with interactions

utility	Coef.	Std. Err.	t	P>t	[95% Conf.	Interval]
worry2	-0.00854	0.012673	-0.67	0.501	-0.03339	0.016313
worry3	-0.0181	0.011512	-1.57	0.116	-0.04068	0.004475
worry4	-0.02587	0.016832	-1.54	0.124	-0.05888	0.007135
worry5	-0.02027	0.014818	-1.37	0.171	-0.04933	0.00879
sad2	-0.01819	0.011489	-1.58	0.113	-0.04072	0.004338
sad3	-0.02445	0.010997	-2.22	0.026	-0.04602	-0.00289
sad4	-0.04804	0.014938	-3.22	0.001	-0.07734	-0.01875
sad5	-0.04872	0.014441	-3.37	0.001	-0.07704	-0.0204
annoy2	-0.01212	0.011044	-1.1	0.272	-0.03378	0.009534
annoy3	-0.01363	0.010643	-1.28	0.201	-0.0345	0.007244
annoy4	-0.01029	0.013455	-0.76	0.445	-0.03667	0.0161
annoy5	-0.01879	0.013567	-1.38	0.166	-0.0454	0.007817
tired2	-0.01756	0.013548	-1.3	0.195	-0.04412	0.009012
tired3	-0.01206	0.014114	-0.85	0.393	-0.03974	0.015617
tired4	-0.01607	0.01449	-1.11	0.267	-0.04449	0.012344
tired5	0.002936	0.015263	0.19	0.847	-0.027	0.032869
pain2	-0.02759	0.010969	-2.52	0.012	-0.04911	-0.00608
pain3	-0.01992	0.01191	-1.67	0.095	-0.04327	0.003439
pain4	-0.12009	0.014029	-8.56	0	-0.1476	-0.09258
pain5	-0.12945	0.013633	-9.5	0	-0.15619	-0.10272
sleep2	-0.01336	0.010892	-1.23	0.22	-0.03472	0.007999
sleep3	-0.00973	0.010908	-0.89	0.373	-0.03112	0.011664
sleep4	-0.04363	0.013412	-3.25	0.001	-0.06993	-0.01733
sleep5	-0.07073	0.013365	-5.29	0	-0.09694	-0.04452
daily2	-0.02712	0.01282	-2.12	0.035	-0.05226	-0.00198
daily3	-0.04757	0.011803	-4.03	0	-0.07072	-0.02443
daily4	-0.07206	0.013827	-5.21	0	-0.09918	-0.04495
daily5	-0.08601	0.014315	-6.01	0	-0.11408	-0.05794

work2	-0.023	0.014636	-1.57	0.116	-0.05171	0.005698
work3	-0.01424	0.014922	-0.95	0.34	-0.0435	0.015027
work4	-0.04551	0.017352	-2.62	0.009	-0.07954	-0.01149
work5	-0.03517	0.014702	-2.39	0.017	-0.064	-0.00634
activ2	0.004424	0.012619	0.35	0.726	-0.02032	0.02917
activ3	-0.03384	0.011814	-2.86	0.004	-0.05701	-0.01067
activ4	-0.01799	0.014382	-1.25	0.211	-0.04619	0.010215
activ5	-0.07422	0.015868	-4.68	0	-0.10534	-0.0431
most	-0.04335	0.019383	-2.24	0.025	-0.08136	-0.00534
least	0.066142	0.036651	1.8	0.071	-0.00573	0.138016
_cons	0.847576	0.045424	18.66	0	0.758498	0.936655

sigma_u .18492976
sigma_e .1725545
rho .53457622 (fraction of variance due to u_i)

F test that all u_i=0: F(281, 2158) = 9.84 Prob > F =

Fixed-effects (within) regression
Group variable: id

Number of obs = 2478
Number of groups = 282

R-sq: within = 0.3728
between = 0.0151
overall = 0.2106

Obs per group: min = 6
avg = 8.8
max = 9

F(38,2158) = 33.76
Prob > F = 0

corr(u_i, Xb) = -0.0036

Random effects model, main effects with interactions

utility	Coef.	Std. Err.	z	P>z	[95% Conf.	Interval]
worry2	0.025616	0.012702	2.02	0.044	0.000721	0.050511
worry3	0.028288	0.011613	2.44	0.015	0.005527	0.051049
worry4	-0.00094	0.016744	-0.06	0.955	-0.03376	0.031877
worry5	0.077414	0.01418	5.46	0	0.049622	0.105207
sad2	0.002725	0.011888	0.23	0.819	-0.02057	0.026024
sad3	-0.00498	0.01144	-0.44	0.663	-0.0274	0.017441
sad4	-0.03375	0.015247	-2.21	0.027	-0.06364	-0.00387
sad5	0.014748	0.014436	1.02	0.307	-0.01355	0.043041
annoy2	0.022509	0.011361	1.98	0.048	0.000242	0.044776
annoy3	0.017913	0.011084	1.62	0.106	-0.00381	0.039636
annoy4	0.030473	0.013899	2.19	0.028	0.003231	0.057716
annoy5	0.058528	0.013549	4.32	0	0.031974	0.085083
tired2	0.02412	0.013346	1.81	0.071	-0.00204	0.050277
tired3	0.035369	0.013664	2.59	0.01	0.008589	0.062149
tired4	0.019815	0.01473	1.35	0.179	-0.00906	0.048685
tired5	0.094801	0.014534	6.52	0	0.066314	0.123287
pain2	0.024051	0.011243	2.14	0.032	0.002015	0.046087
pain3	0.027911	0.012022	2.32	0.02	0.004347	0.051474
pain4	-0.0778	0.014469	-5.38	0	-0.10616	-0.04944
pain5	-0.03176	0.013213	-2.4	0.016	-0.05765	-0.00586
sleep2	0.018505	0.011254	1.64	0.1	-0.00355	0.040563
sleep3	0.017846	0.011372	1.57	0.117	-0.00444	0.040135
sleep4	-0.01831	0.01399	-1.31	0.191	-0.04573	0.009113
sleep5	0.012019	0.013201	0.91	0.363	-0.01385	0.037893
daily2	0.004928	0.012937	0.38	0.703	-0.02043	0.030283
daily3	-0.01071	0.012032	-0.89	0.373	-0.03429	0.012873
daily4	-0.04149	0.014347	-2.89	0.004	-0.06961	-0.01337
daily5	0.008225	0.013889	0.59	0.554	-0.019	0.035448
work2	-0.003	0.014941	-0.2	0.841	-0.03229	0.026279
work3	0.006665	0.015148	0.44	0.66	-0.02302	0.036354
work4	-0.02998	0.017647	-1.7	0.089	-0.06457	0.004605
work5	0.05437	0.014725	3.69	0	0.025509	0.08323
activ2	0.052545	0.012702	4.14	0	0.02765	0.07744
activ3	0.006714	0.011995	0.56	0.576	-0.0168	0.030224
activ4	0.009929	0.014791	0.67	0.502	-0.01906	0.038919
activ5	0.00148	0.015412	0.1	0.924	-0.02873	0.031686
most	-0.06193	0.019812	-3.13	0.002	-0.10077	-0.0231
least	0.595691	0.023512	25.34	0	0.549608	0.641773

sigma_u	.167671	.2040911				
.1849869						
.0092761						
/sigma_e	.1776171	.1885414				
.1829977						
.0027865						
rho	.4530136	.557704				
.5054054						
.0267843						

Likelihood-ratio	test	of	sigma_u=0:	chibar2(01)=	983.85	Prob>=chibar2 =

```
Iteration 0:  =          340.656
log
likelihood
Iteration 1:  =          367.3613
log
likelihood
Iteration 2:  =          368.0325
log
likelihood
Iteration 3:  =          368.0345
log
likelihood

Random-effects ML regression          Number of obs      =      2478
Group variable: id                    Number of groups   =       282

Random effects u_i ~ Gaussian         Obs per group: min =         6
                                                     avg =       8.8
                                                     max =         9

                                      Wald chi2(38)      =   3576.24
Log likelihood  = 368.03446           Prob > chi2        =         0
```

Appendix 9.8 Utility models

OLS main effects model with constant

utility	Coef.	Std. Err.	t	P>t	[95% Conf.	Interval]
worry2	0.006026	0.014733	0.41	0.683	-0.02286	0.034917
worry3	-0.02456	0.014719	-1.67	0.095	-0.05343	0.0043
worry4	-0.01462	0.01801	-0.81	0.417	-0.04994	0.020698
worry5	-0.02209	0.017359	-1.27	0.203	-0.05613	0.011954
sad2	-0.02865	0.014747	-1.94	0.052	-0.05757	0.000266
sad3	-0.03188	0.014711	-2.17	0.03	-0.06073	-0.00303
sad4	-0.06647	0.017874	-3.72	0	-0.10152	-0.03142
sad5	-0.05953	0.017246	-3.45	0.001	-0.09335	-0.02572
annoy2	-0.02613	0.014753	-1.77	0.077	-0.05505	0.002803
annoy3	-0.01986	0.014781	-1.34	0.179	-0.04884	0.009129
annoy4	-0.01271	0.017795	-0.71	0.475	-0.04761	0.022184
annoy5	-0.01519	0.017353	-0.88	0.382	-0.04922	0.01884
tired2	-0.05496	0.014709	-3.74	0	-0.0838	-0.02611
tired3	-0.0278	0.014692	-1.89	0.059	-0.05661	0.001013
tired4	-0.02379	0.017899	-1.33	0.184	-0.05889	0.011311
tired5	-0.02872	0.017346	-1.66	0.098	-0.06273	0.005295
pain2	-0.01929	0.015205	-1.27	0.205	-0.04911	0.010527
pain3	-0.00811	0.015113	-0.54	0.592	-0.03774	0.021529
pain4	-0.10776	0.018231	-5.91	0	-0.14351	-0.07201
pain5	-0.13368	0.017575	-7.61	0	-0.16815	-0.09922
sleep2	-0.02008	0.014733	-1.36	0.173	-0.04897	0.008812
sleep3	0.002577	0.01472	0.18	0.861	-0.02629	0.031441
sleep4	-0.03728	0.018063	-2.06	0.039	-0.0727	-0.00186
sleep5	-0.08609	0.017272	-4.98	0	-0.11996	-0.05222
daily2	-0.02324	0.014749	-1.58	0.115	-0.05216	0.005684
daily3	-0.04779	0.014633	-3.27	0.001	-0.07648	-0.01909
daily4	-0.0566	0.01809	-3.13	0.002	-0.09207	-0.02113
daily5	-0.08763	0.01722	-5.09	0	-0.1214	-0.05387
work2	-0.03681	0.014635	-2.51	0.012	-0.06551	-0.00811
work3	-0.03368	0.014715	-2.29	0.022	-0.06254	-0.00483
work4	-0.07253	0.018152	-4	0	-0.10812	-0.03693
work5	-0.04159	0.01715	-2.43	0.015	-0.07522	-0.00796
activ2	0.009727	0.015313	0.64	0.525	-0.0203	0.039755
activ3	-0.04696	0.015064	-3.12	0.002	-0.0765	-0.01742
activ4	-0.02587	0.018391	-1.41	0.16	-0.06193	0.010192
activ5	-0.10098	0.017691	-5.71	0	-0.13567	-0.06629
_cons	0.911321	0.031718	28.73	0	0.849123	0.973519

Source	SS	df	MS	Number of obs	=	2478
				F(36, 2441)	=	18.8
Model	40.67924	36	1.129979	Prob > F	=	0
Residual	146.7329	2441	0.060112	R-squared	=	0.2171
				Adj R-squared	=	0.2055
Total	187.4122	2477	0.075661	Root MSE	=	0.24518

Mean model, main effects with constant

Mean	Coef.	Std. Err.	t	P>t	[95% Conf.	Interval]
worry2	0.004082	0.016179	0.25	0.803	-0.029176	0.037339
worry3	-0.026656	0.016033	-1.66	0.108	-0.059612	0.0063
worry4	-0.014749	0.019077	-0.77	0.446	-0.053962	0.024464
worry5	-0.023012	0.018684	-1.23	0.229	-0.061419	0.015394
sad2	-0.030737	0.016179	-1.9	0.069	-0.063995	0.00252
sad3	-0.034464	0.016033	-2.15	0.041	-0.06742	-0.001509
sad4	-0.066861	0.019077	-3.5	0.002	-0.106074	-0.027649
sad5	-0.06045	0.018684	-3.24	0.003	-0.098856	-0.022044
annoy2	-0.027471	0.016179	-1.7	0.101	-0.060728	0.005787
annoy3	-0.022075	0.016033	-1.38	0.18	-0.05503	0.010881
annoy4	-0.013016	0.019077	-0.68	0.501	-0.052228	0.026197
annoy5	-0.016346	0.018684	-0.87	0.39	-0.054752	0.02206
tired2	-0.055599	0.016179	-3.44	0.002	-0.088856	-0.022342
tired3	-0.028848	0.016033	-1.8	0.084	-0.061804	0.004108
tired4	-0.025002	0.019077	-1.31	0.201	-0.064215	0.014211
tired5	-0.028196	0.018684	-1.51	0.143	-0.066602	0.01021
pain2	-0.021697	0.016844	-1.29	0.209	-0.05632	0.012927
pain3	-0.010721	0.016498	-0.65	0.521	-0.044633	0.023191
pain4	-0.107369	0.01942	-5.53	0	-0.147288	-0.06745
pain5	-0.134255	0.01899	-7.07	0	-0.173291	-0.09522
sleep2	-0.022096	0.016179	-1.37	0.184	-0.055353	0.011162
sleep3	0.000652	0.016033	0.04	0.968	-0.032304	0.033608
sleep4	-0.037263	0.019077	-1.95	0.062	-0.076475	0.00195
sleep5	-0.087732	0.018684	-4.7	0	-0.126138	-0.049326
daily2	-0.024851	0.016179	-1.54	0.137	-0.058108	0.008406
daily3	-0.049708	0.016033	-3.1	0.005	-0.082663	-0.016752
daily4	-0.057369	0.019077	-3.01	0.006	-0.096582	-0.018156
daily5	-0.08957	0.018684	-4.79	0	-0.127976	-0.051164
work2	-0.034018	0.016227	-2.1	0.046	-0.067373	-0.000663
work3	-0.030581	0.016227	-1.88	0.071	-0.063937	0.002774
work4	-0.06974	0.019588	-3.56	0.001	-0.110005	-0.029476
work5	-0.039446	0.019062	-2.07	0.049	-0.078629	-0.000262
activ2	0.007207	0.016844	0.43	0.672	-0.027417	0.041831
activ3	-0.049427	0.016498	-3	0.006	-0.083339	-0.015514
activ4	-0.027346	0.01942	-1.41	0.171	-0.067265	0.012573
activ5	-0.103064	0.01899	-5.43	0	-0.1421	-0.064029
_cons	0.917957	0.035572	25.81	0	0.844837	0.991076

Source	SS	df	MS		
				Number of obs =	63
				F(36, 26 =	7.89
Model	0.547256	36	0.015202	Prob > F =	0
Residual	0.050081	26	0.001926	R-squared =	0.9162
				Adj R-squa =	0.8001
Total	0.597337	62	0.009634	Root MSE =	0.04389

Mean model, main effects, no constant

Mean	Coef.	Std. Err.	t	P>t	[95% Conf.	Interval]
worry2	0.141775	0.077322	1.83	0.078	-0.016877	0.300427
worry3	0.1002	0.077253	1.3	0.206	-0.05831	0.258711
worry4	0.100633	0.093882	1.07	0.293	-0.091997	0.293263
worry5	0.08217	0.092307	0.89	0.381	-0.107229	0.271569
sad2	0.106956	0.077322	1.38	0.178	-0.051696	0.265609
sad3	0.092392	0.077253	1.2	0.242	-0.066118	0.250903
sad4	0.048521	0.093882	0.52	0.609	-0.144109	0.24115
sad5	0.044732	0.092307	0.48	0.632	-0.144667	0.234131
annoy2	0.110223	0.077322	1.43	0.165	-0.048429	0.268875
annoy3	0.104782	0.077253	1.36	0.186	-0.053728	0.263292
annoy4	0.102366	0.093882	1.09	0.285	-0.090263	0.294996
annoy5	0.088837	0.092307	0.96	0.344	-0.100562	0.278236
tired2	0.082095	0.077322	1.06	0.298	-0.076558	0.240747
tired3	0.098008	0.077253	1.27	0.215	-0.060502	0.256519
tired4	0.09038	0.093882	0.96	0.344	-0.10225	0.28301
tired5	0.076987	0.092307	0.83	0.412	-0.112413	0.266386
pain2	0.202693	0.07303	2.78	0.01	0.052848	0.352537
pain3	0.159483	0.076554	2.08	0.047	0.002408	0.316559
pain4	0.051361	0.09325	0.55	0.586	-0.139972	0.242695
pain5	0.014275	0.091613	0.16	0.877	-0.1737	0.20225
sleep2	0.115598	0.077322	1.5	0.147	-0.043055	0.27425
sleep3	0.127509	0.077253	1.65	0.11	-0.031002	0.286019
sleep4	0.078119	0.093882	0.83	0.413	-0.11451	0.270749
sleep5	0.017451	0.092307	0.19	0.851	-0.171949	0.20685
daily2	0.112843	0.077322	1.46	0.156	-0.04581	0.271495
daily3	0.077149	0.077253	1	0.327	-0.081361	0.235659
daily4	0.058013	0.093882	0.62	0.542	-0.134616	0.250643
daily5	0.015612	0.092307	0.17	0.867	-0.173787	0.205011
work2	0.047578	0.080571	0.59	0.56	-0.117741	0.212897
work3	0.051015	0.080571	0.63	0.532	-0.114304	0.216334
work4	0.011856	0.097862	0.12	0.904	-0.188939	0.212651
work5	0.031951	0.095477	0.33	0.74	-0.163952	0.227854
activ2	0.231597	0.07303	3.17	0.004	0.081752	0.381441
activ3	0.120778	0.076554	1.58	0.126	-0.036297	0.277853
activ4	0.131384	0.09325	1.41	0.17	-0.05995	0.322717
activ5	0.045466	0.091613	0.5	0.624	-0.142509	0.233441

Source	SS	df	MS		
				Number of =	63
				F(36, 27 =	15.58
Model	27.67768	36	0.768824	Prob > F =	0
Residual	1.332772	27	0.049362	R-squared =	0.9541
				Adj R-squa =	0.8928
Total	29.01045	63	0.460483	Root MSE =	0.22218

Median model, main effects with constant

Median	Coef.	Std. Err.	t	P>t	[95% Conf.	Interval]
worry2	0.02463	0.025402	0.97	0.341	-0.027585	0.076845
worry3	-0.019267	0.025172	-0.77	0.451	-0.071009	0.032475
worry4	-0.003983	0.029951	-0.13	0.895	-0.065549	0.057583
worry5	-0.010449	0.029335	-0.36	0.725	-0.070749	0.04985
sad2	-0.033476	0.025402	-1.32	0.199	-0.085691	0.018739
sad3	-0.02454	0.025172	-0.97	0.339	-0.076282	0.027202
sad4	-0.051151	0.029951	-1.71	0.1	-0.112717	0.010415
sad5	-0.038867	0.029335	-1.32	0.197	-0.099167	0.021432
annoy2	-0.029765	0.025402	-1.17	0.252	-0.08198	0.02245
annoy3	-0.017997	0.025172	-0.71	0.481	-0.069739	0.033745
annoy4	0.014767	0.029951	0.49	0.626	-0.046799	0.076333
annoy5	0.002051	0.029335	0.07	0.945	-0.058249	0.06235
tired2	-0.055058	0.025402	-2.17	0.04	-0.107273	-0.002843
tired3	-0.047392	0.025172	-1.88	0.071	-0.099134	0.00435
tired4	-0.047733	0.029951	-1.59	0.123	-0.109299	0.013833
tired5	-0.026074	0.029335	-0.89	0.382	-0.086374	0.034225
pain2	-0.002133	0.026446	-0.08	0.936	-0.056493	0.052228
pain3	0.000653	0.025903	0.03	0.98	-0.052591	0.053897
pain4	-0.08377	0.030491	-2.75	0.011	-0.146445	-0.021096
pain5	-0.130276	0.029816	-4.37	0	-0.191563	-0.068989
sleep2	-0.003788	0.025402	-0.15	0.883	-0.056003	0.048427
sleep3	0.01296	0.025172	0.51	0.611	-0.038782	0.064702
sleep4	-0.040213	0.029951	-1.34	0.191	-0.101779	0.021352
sleep5	-0.09668	0.029335	-3.3	0.003	-0.156979	-0.036381
daily2	-0.034745	0.025402	-1.37	0.183	-0.08696	0.01747
daily3	-0.067997	0.025172	-2.7	0.012	-0.119739	-0.016255
daily4	-0.064335	0.029951	-2.15	0.041	-0.1259	-0.002769
daily5	-0.096387	0.029335	-3.29	0.003	-0.156686	-0.036088
work2	-0.037331	0.025477	-1.47	0.155	-0.0897	0.015038
work3	-0.040456	0.025477	-1.59	0.124	-0.092825	0.011913
work4	-0.067995	0.030755	-2.21	0.036	-0.131212	-0.004779
work5	-0.021923	0.029929	-0.73	0.47	-0.083442	0.039597
activ2	0.013406	0.026446	0.51	0.616	-0.040955	0.067766
activ3	-0.064381	0.025903	-2.49	0.02	-0.117625	-0.011137
activ4	-0.026929	0.030491	-0.88	0.385	-0.089604	0.035746
activ5	-0.083396	0.029816	-2.8	0.01	-0.144683	-0.022108
_cons	0.928732	0.05585	16.63	0	0.813932	1.043533

Source	SS	df	MS		
				Number of =	63
				F(36, 26 =	3.27
Model	0.558602	36	0.015517	Prob > F =	0.0012
Residual	0.123451	26	0.004748	R-squared =	0.819
				Adj R-squa =	0.5684
Total	0.682053	62	0.011001	Root MSE =	0.06891

Median model, main effects no constant

Median	Coef.	Std. Err.	t	P>t	[95% Conf.	Interval]
worry2	0.16394	0.080273	2.04	0.051	-0.000767	0.328646
worry3	0.109079	0.080201	1.36	0.185	-0.05548	0.273638
worry4	0.112754	0.097465	1.16	0.257	-0.087227	0.312734
worry5	0.095968	0.09583	1	0.326	-0.100659	0.292595
sad2	0.105834	0.080273	1.32	0.198	-0.058873	0.270541
sad3	0.103805	0.080201	1.29	0.207	-0.060754	0.268365
sad4	0.065586	0.097465	0.67	0.507	-0.134395	0.265566
sad5	0.06755	0.09583	0.7	0.487	-0.129077	0.264177
annoy2	0.109545	0.080273	1.36	0.184	-0.055162	0.274252
annoy3	0.110348	0.080201	1.38	0.18	-0.054211	0.274908
annoy4	0.131504	0.097465	1.35	0.188	-0.068477	0.331484
annoy5	0.108468	0.09583	1.13	0.268	-0.088159	0.305095
tired2	0.084252	0.080273	1.05	0.303	-0.080455	0.248959
tired3	0.080954	0.080201	1.01	0.322	-0.083605	0.245513
tired4	0.069004	0.097465	0.71	0.485	-0.130977	0.268984
tired5	0.080343	0.09583	0.84	0.409	-0.116284	0.27697
pain2	0.224891	0.075817	2.97	0.006	0.069328	0.380454
pain3	0.172855	0.079475	2.17	0.039	0.009786	0.335925
pain4	0.076823	0.096809	0.79	0.434	-0.121812	0.275458
pain5	0.019998	0.095109	0.21	0.835	-0.17515	0.215146
sleep2	0.135522	0.080273	1.69	0.103	-0.029185	0.300229
sleep3	0.141305	0.080201	1.76	0.089	-0.023254	0.305865
sleep4	0.076523	0.097465	0.79	0.439	-0.123458	0.276504
sleep5	0.009737	0.09583	0.1	0.92	-0.186889	0.206364
daily2	0.104565	0.080273	1.3	0.204	-0.060142	0.269271
daily3	0.060348	0.080201	0.75	0.458	-0.104211	0.224908
daily4	0.052402	0.097465	0.54	0.595	-0.147579	0.252383
daily5	0.01003	0.09583	0.1	0.917	-0.186596	0.206657
work2	0.045223	0.083646	0.54	0.593	-0.126405	0.21685
work3	0.042098	0.083646	0.5	0.619	-0.12953	0.213725
work4	0.014559	0.101596	0.14	0.887	-0.193899	0.223017
work5	0.050312	0.099121	0.51	0.616	-0.153067	0.253691
activ2	0.240429	0.075817	3.17	0.004	0.084866	0.395992
activ3	0.107822	0.079475	1.36	0.186	-0.055248	0.270891
activ4	0.133664	0.096809	1.38	0.179	-0.064971	0.332299
activ5	0.066879	0.095109	0.7	0.488	-0.12827	0.262027

Source	SS	df	MS		
				Number of obs =	63
				F(36, 27 =	16.06
Model	30.75546	36	0.854318	Prob > F =	0
Residual	1.436434	27	0.053201	R-squared =	0.9554
				Adj R-squa =	0.8959
Total	32.1919	63	0.510982	Root MSE =	0.23065

References

Brazier, J.E., Deverill, M., Harper, R., et al. 1999. A review of the use of Health Status measures in economic evaluation. Health Technology Assessment. 3(9)

Brazier, J.E., Roberts, J. & Deverill, M. 2002. The estimation of a preference based measure of health from the SF-36. Journal of Health Economics, 21 (2), 271-292.

Brazier, J. E. & Roberts, J. R. 2004. The Estimation of a Preference-Based Measure of Health from the SF-12. Medical Care. 42(9).

Brazier, J.E., Yang, Y. & Tsuchiya A. 2006. Using DCE to value (asthma) health states for calculating quality adjusted life years. Health Economics and Decision Science Discussion Paper 07/06, ScHARR, The University of Sheffield. http://www.sheffield.ac.uk/scharr/sections/heds/discussion.html Accessed 15/07/2008

Brazier, J.E., Ratcliffe, J., Salomon, J. & Tsuchiya A. 2007. Measuring and Valuing Health Benefits for Economic Evaluation. Oxford University Press.

Dolan, P. 1997. Modelling valuations for EuroQol Health States. Medical Care, 35(11), 1095-1108.

Gold, M. R., Siegel, J. E., Russell, L.B. et al. 1996. Cost Effectiveness in Health and Medicine. Oxford University Press. Oxford

Gujarati, D.N. 1995. Basic Econometrics. Chapter 3. McGraw-Hill, Inc. Third Edition.

Kharroubi SA, Brazier JE, Roberts J, O'Hagan A. 2007. Modelling SF-6D health state preference data using a nonparametric Bayesian method. Journal of Health Economics 26: 597-612.

Kharroubi S, McCabe C. (Forthcoming). Modelling HUI 2 health state preference data using a nonparametric Bayesian method. Medical Decision Making.

Ljung, G., Box, G. 1979. On a measure of lack of fit in time series models. Biometrika. 66, 265-270.

Lloyd, A. 2003. Threats to the estimation of benefit: are preference estimation methods accurate? Health Economics. 12, 393-402.

McCabe, C., Stevens, K., Roberts, J. & Brazier, J.E. 2005. Health State Values for the Health Utilities Index Mark 2 descriptive system: results from a UK valuation survey. Health Economics 14 (3) 231-44.

McCabe, C., Brazier, J.E., Gilks, P., Tsuchiya, A., Roberts, J., O'Hagan, A. & Stevens, K. 2006. Using rank data to estimate health state utility models. Journal of Health Economics. 25, 418-431.

Names and numbers manual.
http://www.afd.co.uk/manuals/namesandnumbers/contents.htm
Accessed 03/07/2008

National Institute for Clinical Excellence. April 2004. Guide to the Methods of Technology Appraisal. NICE.

O'Brien, B. J., Drummond, M. F. 1994. Statistical versus quantitative significance in the socioeconomic evaluation of medicines. Pharmacoeconomics. 5, 389.

Patrick, D.L., Starks, H.E., Cain, K.C. et al. 1994. Measuring preferences for health states worse than death. Medical Decision Making. 14, 9-18.

Ratcliffe, J., Brazier, J.E. 2006. Estimation of a preference based single index from the sexual quality of life questionnaire (SQOL) using ordinal data. Health Economics and Decision Science Discussion Paper 06/06, ScHARR, The University of Sheffield.
http://www.sheffield.ac.uk/scharr/sections/heds/discussion.html
Accessed 15/07/2008

Stevens, K. J., Brazier, J. E., McKenna, S. P., Doward, L. C. & Cork, M. J. 2005. The development of a preference-based measure of health in children with atopic dermatitis. British Journal of Dermatology. 153, 372-377

The MVH Group, Dolan, P., Gudex, C., Kind, P. & Williams, A. May 2004. The Measurement and Valuation of Health. First Report on the Main Survey. Centre for Health Economics, University of York.

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

Walters S.J., & Brazier J.E. 2005. Comparison of the minimally important difference for two health state utility measures: EQ-5D and SF-6D. Quality of Life Research. 14, 1523-1532.

Overall Discussion, Conclusions and Future Research

10.1 Aim

The aims of this chapter are to summarize the thesis, assess the contribution of this research to the literature and also to highlight issues for future research.

10.2 Summary of thesis

Chapter 2 began by summarizing the need for preference based measures (PBM) of health related quality of life and their usefulness in resource allocation decisions. The paucity of research into paediatric PBM was highlighted and the need for further research in this area was argued for.

Chapter 3 reported on a review of the literature for generic paediatric quality of life (QoL) measures and assessed their performance, their purpose and their suitability for use in economic evaluation or adaptation to become a preference based measure. Additional search questions included whether there was any evidence on common health related quality of life frameworks across age and whether children were able to provide information about their health for the purposes of constructing a descriptive system. The conclusion of this review was that there are a range of generic paediatric quality of life measures with different purposes and different definitions of QoL or HRQoL. Children had been involved in their development, but this was generally at a later stage. Only two measures were found in the review that had been developed explicitly for use in economic evaluation but children were not used to develop the dimensions of HRQoL. There was also very limited guidance on key methodological issues facing the development of new measures. Some, albeit limited, evidence was found that children as young as 6 years old can provide information about their health. There was no evidence on whether children share similar health related quality of life frameworks.

Chapter 4 covered key decisions that had to be made in order to develop the descriptive system. This included which population to use to develop the content, which techniques to use, the concept of HRQoL/QoL used, the age range and any developmental issues, constraints imposed by a PBM and finally issues in working and researching with children. It was argued that due to the nature of working with children and ensuring good content validity, that bottom up methods were the most appropriate for this research and children age 7-11 were the only population included. The concept used was HRQoL, which was defined as the impact your health has on your life.

Chapter 5 reported on the qualitative interview work undertaken with children age 7-11 and demonstrated that children of this age were able to talk about and provide information about their HRQoL. It also demonstrated that research of this type is possible with children and was successful. Another finding was that this type of interview work worked best in a 1 to 1 situation and this was what children (with the exception of 1 case) always chose to do. This part of the research also provided evidence that children age 7-11 have a very similar framework of HRQoL and there was no substantial difference in terms of the dimensions of HRQoL developed between the 2 age groups (7-9 years and 9-11 years).

Chapters 6a and 6b reported on developing the dimensions into a descriptive system. The descriptive system had to meet the constraints of a PBM. Instead of using standard scales from the literature, a novel approach was taken, in that the original qualitative data was used to develop scales so that they were based on the children's wording and everything about the scale, whether it was frequency or severity based and the terminology used came from the data. Subsequent testing of the scales by asking children to rank the items was successful and children were able to do this task fairly easily. The overall approach to the scale development was successful and a draft descriptive system was developed.

Chapters 7 and 8a reported on the testing of the measure in 2 different paediatric populations; school and hospital. The testing was extensive in that the researcher was present whilst each child completed their questionnaire and so was in a good position to be able to observe how the measure worked in practice. The draft measure demonstrated good psychometric performance and the data generated from these studies was used to refine the descriptive system.

Final refinement of the descriptive system to make it more amenable to valuation was reported in chapter 8b and this took into account evidence from the psychometric testing as well as using the original qualitative data to inform the reduction of the number of dimensions. A final descriptive system with 9 dimensions, each with 5 levels was developed and this was felt suitable for the pilot preference based work.

Finally, chapter 9 reported on the feasibility valuation study which was undertaken to obtain preference weights for each health state defined by the descriptive system. This study used valuation interviews with 300 members of the UK general population to obtain preference weights for a sample of the health states and then subsequent modelling work was undertaken to estimate a model to predict values for all health states defined by the descriptive system. Whilst intended to be a feasibility study, the study was still reasonably large with 300 respondents. The study was successful although several levels on some of the dimensions had to be collapsed due to inconsistencies and some dimensions were not significant.

10.2.1 Contribution of this research to the literature

The development of a new preference based measure of HRQoL addresses 4 gaps in the literature:

1. The measure was explicitly designed for use in economic evaluation as a paediatric PBM
2. The measure involved children in its development from the beginning

3. The development used qualitative interviews with children only to develop the content of the descriptive system.
4. The development tested whether there is a common HRQoL framework across age.

This generic preference based measure for children has enormous potential to expand the use of cost utility analysis in economic evaluation in the paediatric population. One of its key strengths and differences from the HUI2 and the EQ5D for children is that it is more child focused as children have been involved in the development at every stage and have been the only population involved.

This research has taken a genuine mixed methods approach in that it has combined the use of qualitative (interviews for generating dimensions and determining wording) and quantitative (ranking work, psychometric testing and valuation) methods to develop the measure. The use of this mixed methods approach has worked well and is particularly suited to developing a PBM.

This research has also demonstrated the use of qualitative data to generate a health state classification. The use of these techniques is likely to increase in the future as the demand for PBMs increases and this research has demonstrated that qualitative methods can be used successfully in this context and importantly, successfully used with children.

This research has also shown the extent of differences between 2 age groups; 7-9 year old and 9-11 year olds. The research demonstrated that in fact there is very little difference between these 2 age populations in terms of how health affects their lives.

This research has also demonstrated that child health states for this descriptive system can be valued by adults.

There are several key advantages to this new measure:

- It is preference based and can be used to calculate QALYs for children for use in CUA
- It is only based on the views of children and has been developed using bottom up methods
- It has been developed with children with a wide range of health problems and diverse backgrounds and there is evidence for its use as a generic measure
- The measure can be self completed by children
- The recall period is appropriate for children
- The measure performs well in both a school and hospital setting
- The breadth of coverage of the dimensions is still good for a PBM as the reduction of dimensions that had to be done to meet valuation constrains was small (jealous and embarrassed)
- The age appropriateness is excellent as the vocabulary, instructions, sentence structure, content and response options have all been developed by children for children. This is an issue highlighted by Matza (2004) who notes that before implementing a child report HRQoL measure, the age appropriateness needs to be evaluated in terms of vocabulary, instructions, sentence structure, content and response options.

10.3. Issues for future research

There are two main areas where further research is required, the first is around testing the measure and the second is around the valuation.

10.3.1 Further testing of the measure

Whilst the measure developed has been tested on a paediatric population there still remain issues for further research. Firstly, whilst the measure was developed with 7-11 year old children, there is no reason that it may not also be suitable for other age groups of children, for example 5-7 year olds or 11-13 year olds. However, it may be the case that there are dimensions of HRQoL that are missing from the measure for these other age groups. Equally, there may be

dimensions that are redundant. This is something that can be tested empirically in the future.

Another important area of future research is to test the performance of the measure on specific patient populations, including both acute, chronic, minor and more serious conditions, for example asthma, diabetes and cancer. Plans for this are already underway as the measure is already included in studies with children with diabetes, obesity, tooth decay and the measure is planned for inclusion in a study of childhood appendicitis in the UK. The measure is also being used as an outcome in measuring the effectiveness of child protection measures in Australia. Further details of the studies are given in Table 9.1 below.

Table 9.1: Details of studies using the CHU9D

Population	Study type	Where	Research Group	Funder
Diabetes Type 1, self management	Trial	UK	The University of Sheffield and Sheffield Children's Hospital Trust	Diabetes UK
Investigation of the effects of an after school physical activity and lifestyle modification programme		UK	Leeds Metropolitan University	
Fitness and Health in South Australians	Survey	Australia	University of South Australia	
Tooth Decay	Trial	Scotland	University of Glasgow	

Diabetes Type 1 Near patient sensing, wireless data transmission and analysis	Trial	UK	The University of Sheffield and Sheffield Children's Hospital Trust	In submission
Appendicitis	Observational Study	England	The University of Sheffield and Sheffield Children's Hospital Trust	In submission

In addition, the performance of the measure compared to other paediatric measures is desirable as this allow further testing on the comparable practicality, reliability and validity of this measure compared to others. Finally, another area for future research is to assess the psychometric properties of the measure including the preference weights, using the checklist by Brazier (1999) discussed in chapter 3.

10.3.2 Valuation

There are many important questions to be addressed around the issue of valuation, perhaps the most important of which is whose values should be used to obtain preference weights. The research in this thesis contains a feasibility study where the health states have been valued using a choice based method on the UK general population. This is what is currently recommended by The National Institute for Health and Clinical Excellence, NICE does not explicitly state that adult valuations are required and it may be that children's valuations are more appropriate. There are two separate issues, firstly should we use children's values and secondly how should we obtain them given that the

methods in which we obtain preferences are deemed to be too cognitively demanding for children. The first issue is essentially a normative question and perhaps can be seen in the same manner as the issue over whether to use the values of the patients or the general population. The second issue is a key area for future research. As mentioned in chapter 9, obtaining preferences from children has not really been done before due to the difficult methods of health state valuation. However, given the recent interest and development in ordinal techniques, and the use of discrete choice experiments to generate cardinal preferences, there may be an opportunity here to access children's preferences through these methods. It is a very important issue as whilst the measure developed here describes how children's lives are affected by their health, it does not tell us the (relative) importance of the dimensions for children.

It may be that valuation by children may lead to further refinement of the descriptive system, for example if there were dimensions that were not significant to children, however this is not known at this stage. There is some evidence that there are already too many dimensions from the adult valuation work. Refinement of the descriptive system may also arise after the measure has been used in large studies and trials and larger datasets gathered on the performance of the measure.

Another option for future valuation work would be to use informed adult preferences. The adults in this valuation study did not know that these preference weights were for child health states, they were asked to imagine themselves in the health state as an adult. It may be that if they knew the states were childhood states, they would value them differently. Again, this is something that can be tested empirically.

Whilst the feasibility study was reasonably large, at 300 respondents, it may be that a valuation study with a larger number of respondents may overcome some of the problems encountered with non significant coefficients and inconsistencies in the model.

Finally, the issue of valuation survey design is one that has already been highlighted as a priority for future research (Brazier 2007). There is very little guidance in the literature about health state selection for valuation work of this type and it may be that the orthogonal design is unsatisfactory and different designs are needed.

10.4 Conclusion

The research reported in this thesis has developed a generic paediatric preference based measure of health related quality of life that can now be used in practice. The descriptive system was developed using the population it is intended for (children), using bottom up methods and there was no introduction of any pre existing ideas or concepts that may have influenced what children said/thought. The final measure contains 9 dimensions; *worried, sad, annoyed, tired, pain, sleep, daily routine, school work* and *joining in activities*. Each of the response scales has five levels which were developed empirically from the data generated in this study, therefore they are in the children's language. The measure has been successfully tested in both the hospital and school setting and has good psychometric performance. The measure has been developed for a specific purpose right from the beginning of the research and this has driven the key decisions made and the analysis undertaken. This purpose was to be a PBM and the measure has not had to be adapted like most other condition specific PBMs to be suitable for preference work. Preference weights which conform to the guidelines by The National Institute for Health and Clinical Excellence (i.e. adult general population using a choice based method) have been developed and the final measure, the Child Health Utility 9D (CHU9D) is now ready to be used in practice and can be used to calculate quality adjusted life years (QALYs).

References

Brazier, J.E., Deverill, M., Green, C., Harper R & Booth A. 1999. A review of the use of health status measures in economic evaluation. 3[9]

Brazier, J.E., Ratcliffe, J., Salomon, J. & Tsuchiya A. 2007. Measuring and Valuing Health Benefits for Economic Evaluation. Oxford University Press.

Matza, L. S., Swensen, A. R., Flood, E. M., Secnik, K. & Leidy, N. K. 2004. Assessment of Health Related Quality of Life in Children: A Review of Conceptual, Methodological, and Regulatory Issues. Value in Health. 7(1), 79-92.