Paper for HESG January 2003, Leeds

Multi-attribute utility functions for the revised HUI2 classification system: results from a UK valuation survey

Christopher McCabe,[+] Katherine Stevens[+], John Brazier[+]**

Sheffield Health Economics Group[+], and Trent Institute for Health Services Research[**], ScHARR, University of Sheffield, UK

Address for correspondence:

Christopher McCabe
Sheffield Health Economics Group
& Trent Institute for Health Services Research
School of Health and Related Research
University of Sheffield
Regent Court
30 Regent Street
Sheffield S1 4DA, UK

Telephone: 0114 222 0728
Fax: 0114 272 4095
Email: c.mccabe@sheffield.ac.uk

Multi-attribute utility functions for the revised HUI2 classification system: results from a UK valuation survey

Christopher McCabe, Katherine Stevens, John E Brazier

Abstract

This paper reports the results of a study to estimate a multi-attribute utility function (MAUF) for a revised version of the Health Utilities Index Mark 2, from Standard Gamble (SG) and Visual Analogue Scale data (VAS). Two hundred and one members of the UK general population undertook VAS valuations of all single attribute health states and 13 multi-attribute health states. Four of the 13 multi-attribute health states were also valued using SG. Power and Cubic VAS-SG transformation curves were estimated and used to map the single and multi-attribute VAS values to utilities. Two MAUF valuation models were constructed and their predictive performance compared in a validation sample of 65 health states valued using SG (n=252). The predictive performance of the two models was compared using the Root Mean Squared Error, (RMSE), the proportion of health states predicted to within 0.05 (absolute) of the observed mean value, and the Ljung Box test for systematic errors. The restricted Cubic MAUF has a lower RMSE than the Power MAUF (0.082 vs. 0.097) and it predicts 10% more health states to within 0.05 of the observed mean (44.6% vs 33.8%). There was no evidence of systematic errors for either MAUF. Both models had superior predictive performance compared to the original MAUF developed in Canada. The predictive performance of the two new models and the Canadian model does not compare well with that reported for statistical valuation models such as the SF-6D and the EQ-5D.

Acknowledgements

We wish to acknowledge tonstructive comments David Feeny, George Torrance and Bill Furlong, Jenny Roberts and Aki Tsuchiya on earlier versions of this paper. We would also like to acknowledge Tony O'Hagan's contribution to the work on mapping VAS to SG. Finally we would like to thank all the interviewers from SSRC, at Sheffield Hallam University, who carried out the survey and the respondents who gave up their time to complete the interviews. The usual disclaimer applies.

The work reported in this paper was supported by a grant from the UK Medical Research Council

Multi-attribute utility functions for the revised HUI2 classification system: results from a UK valuation survey

Introduction

The United Kingdom Paediatric Intensive Care Outcomes Study is using the Health Utilities Index Mark 2 to examine long term outcomes after paediatric intensive care in 23 centres in England, Scotland, Wales and Northern Ireland. Part of the study involves the construction of a UK valuation algorithm for the HUI 2 health state classification system. This paper reports some initial results of this work.

The Health Utilities Index Mark 2 (HUI2) is the only preference based multi-attribute health related quality of life instrument specifically developed for use with children.[1] [A] It consists of seven dimensions (sensation, mobility, emotion, cognition, self care pain and fertility), each of which has between three and five levels. The levels describe a range, from 'normal functioning for age' to 'extreme disability'. (Appendix One gives the dimensions and level descriptions.)

The first applications of the system were in paediatric oncology. The fertility dimension was added to the original six dimensions proposed by Cadman and colleagues,[2] in order to capture side effects of chemotherapy. The developers state that by assuming fertility to be normal, the HUI2 can be used as a generic health status instrument.[3 4]

Preference based quality of life weights can be calculated for all 24,000 health states in the descriptive system using a multiplicative multi-attribute utility function (MAUF) developed by Torrance and colleagues. This MAUF is based on health state valuation interviews with 194 parents of school age

[A] Apalossja has developed the 16D and 17D for application in adolescent and pre-adolescent children. However, the disaggregated valuation mechanism used in their development has yet to be shown to produce preference data.

children in Hamilton, Ontario, Canada. The valuation interview followed the standard McMaster Valuation Framework, whereby each health state is valued using a Visual Analogue Scale (VAS) called the Feeling Thermometer. The value data obtained using the VAS are converted to utilities by means of a power curve transformation. The power curve transformation is estimated using the person mean values and utilities of four health states, which are valued using VAS and Standard Gamble methods respectively. [4]

UK PICOS undertook three separate health state valuation surveys (see Figure 1). Survey 1, follows closely the health state valuation survey undertaken to estimate the Canadian Multi-attribute Utility Function (MAUF) for the HUI2. The second collected HUI2 health state valuation data for the estimation of a statistical inference valuation model for the HUI2 and the third collected a small HUI2 health state valuation dataset for use in comparing the predictive performance of the MAUF and statistical inference valuation models. .

UK PICOS uses a revised version of the HUI2, excluding the fertility dimension as a review of the literature established that this dimension has never been used successfully in practice and the developers of the HUI2 recommend that Fertility is fixed at full health to calculate the MAUF value for any given health state.

In this paper we report the estimation of the multi-attribute value and utility functions for the revised HUI2 health state classification system using the data collected in Survey 1, described above, and assess the new MAUFs' predictive performance using the direct health state valuation data from surveys 2 and 3.

The paper is structured as follows. In section one I report the results of the MAUF valuation survey. The second section reports the estimation of the multi-attribute value for the mobility corner state, and the adjusted (unconfounded) single attribute mobility values. Section three reports the estimation of the multi-attribute value function for the revised HUI2 health state classification system . Section four reports the estimation of two multi-attribute utility functions for the HUI2; the first using power curve transformation estimates of the single and multi-attribute utilities, the second using cubic transformation estimates of the single and multi-attribute utilities. The final section of this paper reports the predictive performance of each of the MAUF models in the 65 health states for which direct SG valuations were obtained in the surveys for estimating and validating the statistical inference valuation model.

Section One: Multi-attribute Utility Function valuation survey

The valuation survey was designed to parameterise a multiplicative multi-attribute utility function (M-MAUF). To parameterise an individual M-MAUF requires that each respondent values all levels on each of the six dimensions of the classification system, plus the corner states.[3] Corner states are those states where one dimension is fixed at its lowest level of function whilst all other states are fixed at their highest level of functioning. One of the corner states in the HUI2 classification is recognised as being infeasible and therefore it was valued indirectly. The mobility corner states includes the descriptions ‘Unable to control or use arms and legs’ and ‘Eats,bathes,dresses and used the toilet normally for age’. It is generally accepted that this combination is not feasible within a single health state. Therefore, the corner state value for mobility had to be implied from the values given to feasible health states close to the mobility corner state. The methods for estimating the value of the mobility corner state are described below.

Survey

Following the recommendations of the Washington panel,[5] we aimed to achieve a representative sample of the general population. The sampling method aimed to achieve a sample that reflected the geographical distribution and socio-economic characteristics of the UK general population, in terms of age, sex, employment status and education.

The target sample size for the valuation survey was 200, approximately equal to the 194 used in the original HUI2 valuation exercise.[6]

Trained and experienced interviewers carried out all interviews. The interviews took place in the respondents' own home. SSRC, a professional survey and research group within the Sheffield Hallam University, employed the interviewers.

The interview consisted of five phases. In the first phase of the interview the purpose of the research was described to the respondent and consent to the interview was obtained. In the second phase of the interview the respondent was asked to value all the single attribute states in the revised HUI2 descriptive system, using the Visual Analogue Scale (VAS). The third phase of the interview used the VAS to obtain valuations for 12 multi-attribute health states in the revised HUI2 descriptive framework, plus immediate death. The multi-attribute states were the same as those used in the original HUI2 valuation survey, except that none of the health states included the fertility dimension and the fertility corner state was not valued.

The fourth phase of the interview revalued 4 of the multi-attribute health states using the standard gamble technique (SG) with the chance board prop. All respondents valued the same four health states, again based on the original HUI2 valuation study.[6] The final phase of the interview collected the same socio-economic and respondent performance data as Survey 1.[7]

The study employed the version of the SG developed by the HUI2 development team at McMaster University.[8] This version uses a prop called the Chance Board.

In the interview, the respondent is asked to choose between the certain prospect (A) of living in an intermediate health state defined by the HUI2 and the uncertain prospect (B) of two possible outcomes, the best state defined by the HUI2 (full health) and immediate death. The chance of the best outcome occurring is varied until the respondent is indifferent between the certain and the uncertain prospect. At all times the probabilities are displayed on a chance board, both numerically and in the form of a pie chart.

For health states valued as worse than immediate death on the VAS exercise, the reference states were the best state defined by theHUI2 (full health) and the impaired health state, with immediate death being the certain outcome.

The chance board is designed to lead the interviewer through a set of questions depending upon the interviewee's response to the previous question and thereby minimise the risk of interviewer variation. The lead investigator (CM) received training in the use of this method from the McMaster University team, who also produced the chance boards used in the survey. The lead investigator trained all interviewers.

The McMaster team did not produce the version of the VAS used for this valuation study, however, the dimensions were based upon the McMaster Feeling Thermometer.[8] Particular care was taken to provide clear differentiation between each point on the scale and ensure adequate space for the multiple health state description cards, which would have to be viewed simultaneously.

Two hundred and one interviews were successfully completed. Twenty five interviews had to be excluded because either all four SG/VAS health states were given the same value (on either the VAS or SG question) or two or more states were not valued on either the VAS or SG. A further 49 respondents were excluded because they exhibited one of two illogical valuations:

1. They valued multi-attribute health state 6 (1,5,1,1,4,1) more highly than multi-attribute health state 3 (1,3,1,1,1,1); or
2. They valued health state 8 (1,4,2,1,1,1) more highly than health state 3 (1,3,1,1,1,1).

Fourteen observations were excluded because of illogical ordering on one or more of the single attribute valuations. This resulted in a dataset of 113 observations for use in the estimation of the MAUF.

Tables 1 and 2 give the measured mean values for the 27 individual attribute levels and the 12 multi-attribute health states plus immediate death. The mobility single attribute values are the calculated person-mean values. The method of calculation is described below.

It should be noted that the MAUF is calculated in disutility space, as the corner states in disutility (and disvalue) space are considered easier to imagine.[6] The corner states in utility space set one dimension at full functioning and all other dimensions at the lowest level of functioning. The corner states is disutility space set one dimension at the lowest level of functioning and all other dimensions at the highest level of functioning. The utility of a health state is calculated as 1 minus the disutility.

Section 2: Estimating the mobility corner state value and unconfounded mobility single attribute scores.

The methods for estimating the mobility corner state and unconfounded mobility single attribute scores were originally described by Torrance et al. [9]

The mobility corner state was not valued directly because it is implausible ('Unable to control or use arms and legs', cannot be combined with 'Eats, bathes, dresses and uses the toilet normally for age') Instead, states (1,5,1,1,4,1,); the self-care mobility joint corner state and (1,3,1,1,1,1,) were valued. These data were then used to construct a multiplicative multi-attribute disvalue function in self-care mobility space:

$$\overline{v} = \alpha x + \beta y + (1 - \alpha - \beta)xy \qquad (1.1)$$

where:
v = disvalue of the health state; (disvalue = 1 minus VAS score).
x = the mobility single attribute disvalue;
y = the self care single attribute disvalue; and
α and β are the model parameters.

In equation 1.1, α is the corner state value for mobility (in the 2 dimension MAUF).

Figure 2 illustrates the 2 dimensional multi-attribute disvalue function used to calculate the mobility corner state value and adjust for confounding between self-care and mobility. State 1 has a disvalue of zero, and State 6 has a disvalue of 1. The disvalue of state 3 in this 2 dimensional mauf is calculated as the directly measured disvalue (0.282) divided by the difference between directly measured disvalue of state 6 (0.51) and the disvalue of state 1 (0.0).

The disvalue of state 3 is 0.554 in mobility self care space.

Equation 1.1, for state 3 becomes:

$$0.554 = \alpha\,(\text{mobility}_3) + \beta(\text{selfcare}_1) + (1\text{-}\alpha\text{-}\beta)\,(\text{mobility}_3 \text{ x selfcare}_1)$$

We know that selfcare_1 has a disulility value of zero, (this is imposed in the valuation method), therefore Equation 1.1 simplifies to:

$$0.554 = \alpha(\text{mobility}_3)$$

α is the mobility corner point in mobility self-care space. It is assumed that the minimum value for α sets α equal to mobility_3. In this two dimension model α is the value attached to health state mobility $_5$- selfcare_1, and the value for mobility_3 is the value for health state mobility_3- selfcare_1. Therefore, it seems reasonable to assume that the value for α will not be less than the value for mobility_3.

Torrance et al choose the mid-point of the range described by setting the α and mobility_3 equal, and assume that the mobility corner state is equal to the mobility-self care joint corner state, with a disvalue of 1. (See Figure 2)

$$\alpha = (1+ \bullet 0.554)/2 = 0.872$$

The estimated disvalue for α is then multiplied by the disvalue for the joint mobility-self care corner state, State 6 (0.51), to estimate the disvalue for the mobility corner state in the full HUI2. This is simply the reverse of the process used to map from the six dimension MAUF to the two dimension MAUF. Thus, the person mean mobility corner state in the full (six dimension MAUF) is estimated as:

$$0.872 * 0.51 = 0.445$$

It should be noted that the value ascribed to α reflects two assumptions. The first assumption sets the lower limit value for α, where α and x_3 are equal to each other. This is a logically reasonable assumption, given the meaning of α and x_3 (the corner state disvalue for mobility and the mobility single attribute level 3 disvalue respectively). The second assumption places α at the midpoint of the possible range.

The mobility single attribute value for level 3 is then calculated as the disvalue of state 3 (0.554), divided by the corner state disvalue, α, (0.872). The adjusted value for level 2 on the mobility single attribute is then interpolated from the new value for level 3 and the value for level 1, so as to maintain the same relative value as observed in the measured data.

The disvalue for level 4 on the mobility single attribute is estimated in two stages; the first stage interpolates a disvalue which maintains the same relative disvalue for level four compared to level 3 as in the measured data. The second stage divides this disvalue through by the corner state disvalue (0.872), to remove the confounding effect of self care on the single attribute valuation. This approach assumes that mobility levels 4 and 5 are confounded by an assumption that the individual would have to be at self care level 4. Thus the corner state weighting (α), has to be extracted from the (transformed) measured disvalue to obtain the unconfounded single attribute disvalue.

Tables 1 and 2 give the measured and estimated single and multi-attribute health state disvalues used to calculate the MAVF, and the two MAUFs.

Section 3: Estimation of the Multi-attribute value function

The multiplicative multi-attribute value function for the HUI2 has the following form:

$$V(x) = (1/k)\left[\prod_{j=1}^{6} (1 + k\,k_j\,v_j(x_j)) - 1\right] \qquad 1.3$$

Where:

$$(1+k) = \prod_{j=1}^{6} (1 + k\,k_j) \qquad 1.4$$

vj (xj) is the single attribute utility function for attribute j
V (x) is the value for health state x, represented by an x-element vector
k and kj are model parameters
Π is the product sign

For health state 6 (1,5,1,1,4,1), the disvalue formulation of equation 1.3 simplifies to:

$$0.51 = (1/d)\,[(1+d*d_m*u_{5m})*(1+d*d_{sc}*u_{4sc})-1] \qquad \mathbf{1.5}$$

where:

d = disvalue scalar parameter;

d_m = disvalue for mobility corner state (1,5,1,1,1,1);

u_{5m} = disvalue of level 5 on the mobility attribute;

d_{sc} = disvalue for selfcare corner state (1,1,1,1,4,1); and

u_{4sc} = disvalue for level 4 on the selfcare attribute.

The values for d and d_{sc} are unknown. The sum of the known corner state values is greater than 1 and therefore the value of d must lie between 0 and –1. [10] In addition d_{sc} must assume a value such that:

$$1+d = \prod_{j=1}^{6}(1+d*dj) \qquad \textbf{(1.6)}$$

where:
d = disvalue scalar parameter; and
d_j = disvalue for corner states 1 to 6 in the revised HUI2 classification.

Expanding Equation 1.5 gives:

$$0.51 =(1/d)*(((1+((d*0.445)*1))*(1+((d* d_{sc})*1)))-1) \qquad \textbf{(1.7)}$$

and Equation 1.6 becomes:

$$1+ d = ((1+d*0.499)*(1+d*0.445)*(1+d*0.665)*(1+d*0.248) *(1+d*d_{sc})*(1+d*0.459)) \qquad \textbf{(1.8)}$$

An iterative search process identifies that when d=-0.959 and d_{sc} = 0.113, equation 1.7 gives a value of 0.510, and equation 1.8 gives a value of 0.041. Thus the values for d and d_{sc} are –0.959 and 0.113 respectively. This is a unique solution.

Table 3 gives all the parameter values for the MAVF.

Section 4: Estimating the multi-attribute utility functions

The process of estimating the MAUF is analogous to the process for estimating the MAVF, however the measured value parameters have to be converted to utilities prior to solving the equations to identify the values for the two scalar parameters d and d_{sc}.

The standard VAS to SG transformation is a power curve,[3] however, a constrained cubic transformation was found to perform better than the power curve in these data. The person mean power curve and cubic functions are reported in Tables 4 and 6. The predictive performance of the mapping functions estimated on the 113 respondents used to estimate the MAUFs is reported in Appendix 1. The constrained cubic transformation is superior to the power transformation in both samples, in terms of explanatory power and predictive performance.

In this section I present a person mean MAUF for both the power curve utility estimates and the constrained cubic utility estimates.

An MAUF based on the power curve transformation.

Table 4 reports the power curve mapping function. Table 5 reports the multi-attribute and single attribute utility values produced by the power curve mapping function.

The disutility formulation of the multiplicative MAUF (equation 1.3 above) for health state 6 is:

$$0.381= (1/d)*(((1+((d*0.313)*1))*(1+((d* d_{sc})*1)))-1) \qquad \textbf{(1.9)}$$

The disutility formulation for the scalar constraint (equation 1.4 above) is:

$$1+d = ((1+d*0.369)*(1+d*0.313)*(1+d*0.557)*(1+d*0.136)*(1+d*d_{sc})*(1+d*0.328)) \qquad \textbf{(1.10)}$$

If d=-0.839 and dsc = 0.088, equation 1.9 gives a value of 0.381, and equation 1.10 gives a value of 0.161. Thus the values for d and d_{sc} are –0.839 and 0.088 respectively.

An MAUF based on the constrained cubic transformation.

Table 6 reports the constrained cubic VAS-SG mapping function. Table 7 reports the multi-attribute and single attribute utility values produced by the constrained cubic VAS-SG mapping function.

The disutility formulation of the multiplicative MAUF (equation 1.3 above) for health state 6 is:

$$0.297 = (1/d)*(((1+((d*0.278)*1))*(1+((d*\ d_{sc})*1)))-1) \qquad \textbf{(1.11)}$$

The disutility formulation for the scalar constraint (equation 1.4 above) is:

$$1+d\ = ((1+d*0.294)*(1+d*0.278)*(1+d*0.382)*(1+d*0.222)$$
$$*(1+d*d_{sc})*(1+d*0.282)) \qquad \textbf{(1.12)}$$

If d=-0.668 and dsc = 0.023, equation 1.11 gives a value of 0.297, and equation 1.12 gives a value of 0.332. Thus the values for d and d_{sc} are –0.668 and 0.023 respectively.

Section 5: Predictive performance

A separate health state valuation survey was undertaken to value 15 health states using the Standard Gamble technique. The interviews used the same version of the SG questionnaire as was used in the MAUF valuation survey. The interviews differed in two distinct ways. The direct valuation survey consisted of a ranking exercise, followed by 8 standard gamble questions, whilst the MAUF interview consisted of 6 single attribute VAS exercises, followed by 12 multi-attribute VAS exercises, prior to undertaking 4 standard gamble questions.

Trained and experienced interviewers carried out all interviews. The interviews took place in the respondents' own home. SSRC, a professional survey and research group within the Sheffield Hallam University, employed the interviewers. All the interviewers received training from the lead investigator (CM).

The interview consisted of four phases. In the first phase of the interview the purpose of the research was described to the respondent and consent to the interview was obtained. In the second phase of the interview the respondent was asked to rank 9 health states from the HUI2 classification system, plus immediate death. This allowed the individual to familiarise themselves with the descriptive system and become familiar with the task of comparing health states. In addition, the data could be used to check the consistency of the valuation data obtained in the third phase of the interview. The ranking data was also used to identify which version of the Standard Gamble question should be used in the second phase of the interview; i.e for health states better or worse than death. The third phase of the interview consisted of 8 standard gamble exercises. All respondents valued the PITS state in the HUI2 classification system, i.e. the state where each dimension is at the lowest level of functioning. The seven remaining states were a sample from the 14 states to be valued in the survey.

The 14 states valued in the survey were chosen to reflect the range of health states within the HUI2 classification system subject to the following constraints;

- They must not be valued in the valuation survey for the Statistical Inference HUI2 health state valuation model; and
- They must not be valued (using either VAS or SG) in the valuation survey for the multi-attribute utility function HUI2 health state valuation model.

In order to ensure that all health states were valued a similar number of times, the health states were split in to 2 groups. Each interviewer was issued with two envelopes, each envelope containing 7 cards from the sample of 14 plus the PITS state (the lowest level of functioning on all six dimensions of the HUI2). The interviewers were instructed to alternate between envelopes 1 and 2 until the sample was reached. The survey company who carried out the interviews were instructed to monitor the return scripts, to ensure that any differences in recruitment rates between interviewers did not lead to significant in-balance in the number of valuations per state.

The fourth phase of the interview consisted of a series of questions about the respondents' socio-economic circumstances. Finally the respondent was asked to rate how easy or difficult they had found the each set of questions (ranking and standard gamble), on a five point scale, ranging from very difficult to very easy. After the interview had been completed, the interviewer completed a brief assessment of the respondents understanding and effort on the ranking and standard gamble tasks.

For the purposes of comparing the predictive performance of the MAUF models, the direct health state valuations used to estimate the statistical valuation model could also be used.[7] The interview processes for eliciting these values were identical to that described above.

Combining the two datasets (the statistical inference valuation survey and the validation survey) produces a sample of direct valuations on 65 health states.

These valuations were all obtained using the same script, props and interviewers. They can therefore be used to compare the predictive performance of the MAUF models. It should be noted that *only* the data on the 14 health states not used in estimating the statistical inference model can be used to compare the performance of the MAUF and statistical inference models.

For each MAUF we report the Root Mean Squared Error (RMSE) for each of the models. We also use the Ljung-box test to examine whether there is a systematic pattern in the prediction errors. In addition, we report the predicted values and associated RMSE for these health states using the original McMaster valuation algorithm, with the assumption that Fertility is at level 1.[13]

Table 8 reports the mean observed health state value for the 65 health states in the validation dataset. It also shows the predicted value for each state from the Cubic MAUF, the Power MAUF and the original McMaster MAUF; and the RMSE for each model.

The RMSE for the Power MAUF is 0.112. The RMSE for the Cubic MAUF is slightly better, at 0.104. The RMSE for the McMaster MAUF (assuming that Fertility is fixed at level one) is 0.13.

The Ljung Box test scores were 6.928 for the power MAUF and 7.256 for cubic MAUF, indicating that the prediction error is not related to the severity of the health state for either MAUF.

Discussion

In this paper I have reported two MAUFs for the revised HUI2 health state classification system. As with the original MAUF developed by Torrance and colleagues, both models are estimated using transformed VAS data. For one model I have used the power curve mapping function recommended by Torrance and colleagues.[11] For the second model, the data were transformed using a constrained cubic transformation. This function was chosen on the basis of superior explanatory power and superior predictive performance to power, linear and quadratic functions. As with the original MAUF, both the transformation functions are estimated from the person mean values, not from the individual level data.

Both models predict the mean values in the validation set comparatively well, outperforming the original McMaster algorithm by a considerable margin. The model based on the cubic formulation performs slightly better. This said, the cubic model has a slightly higher RMSE than the recently reported SF-6D instrument, 0.104 compared to 0.096. (personal communication Dr. Jennifer Roberts)..

A number of issues have arisen in the process of constructing the models. The proportion of the respondents whose data could be used in the estimation of the models is relatively small (56%). This is considerably less than for the statistical inference model (89%).[7] Respondents were largely excluded because of illogicalities in multi-attribute and single attribute valuation exercises. This suggests that the VAS may not be an efficient means of collecting health state valuation data for these purposes (setting aside the debate around whether the VAS does capture preferences). Torrance and colleagues themselves state that the VAS is not a simple method for health state valuation data.[12] When one considers that the multi-attribute health state exercise requires that the respondent consider 85 pieces of information (the levels of each attribute in each of 12 health states, plus immediate death, plus the distance between these 13 states, reflecting the individuals preferences), it is perhaps unsurprising that apparently illogical valuations are produced. The psychological literature suggests that individuals can consider between 7 and 9 pieces of data in making a decision. One would have expected the single attribute valuations to be completed more successfully. Even the attributes with five levels do not exceed upper limit of 9 pieces of data. Possible explanations for illogical valuations on the single attributes are less obvious. For some attributes, it might be argued that reversals in the rankings are justifiable. For example, some individuals may view health states with significant disability but not support as being worse than more health disabled health states with support.

Against the background of poor performance of the VAS data collection in this survey and in the original HUI2 valuation survey, the use of VAS should be reconsidered. The Standard Gamble questions work well in the surveys reported in this paper and the paper on the statistical inference valuation model.[7] A sample size of 200 would be sufficient to estimate a person mean MAUF for the revised HUI2, using Standard Gamble data only, with over 70 observations per parameter.[14]

The utilisation of a restricted cubic mapping function for estimating utility values from VAS data represents a significant departure from the standard method. Torrance and colleagues recently reported that all existing VAS-SG mapping studies has used the power curve function, even those that criticised its use. In the recently published HUI3 valuation study, Feeny et al report that the power curve was the best performing of the alternative functional forms considered.[13] In this data, the power curve is not the best functional form and the utilisation of the cubic mapping function leads to a small but significant improvement in the predictive performance of the MAUF, compared to the power curve.

The use of VAS data is now recognised as introducing bias.[11] It has been suggested that the application of the PW adjustment, to remove the context bias from the VAS data should be standard practice. However, VAS data that have been transformed using the PW adjustment, are no longer on the 0-1 scale required for the estimation of the MAUF. It appears that the PW adjustment is not a solution to the problem of bias in this particular situation.[14] The essential problem with the PW adjustment is that it relies on the estimation of a non-observable parameter, *w*, which captures the relative weigh that respondents give to the frequency and range effects in valuing states using VAS. Whilst it is possible to establish a feasible range of values for *w*, it is not possible to demonstrate that any one value is correct.

Both of the MAUFs reported in this paper are complex functional forms compared to the statistical valuation model that has been developed for the revised HUI2. The MAUF models do not perform as well as the simpler statistical valuation model.[7] However, given the complexity of the valuation and estimation process, the difference in the performance of these valuation models may be attributable to a number of factors:

- A misspecification of the functional form;
- A misspecification of the relationship between VAS and SG data;
- Bias in the VAS data;
- Unknown differences in the response samples used to estimate the models; and
- A combination of the above factors.

Further work is required to disentangle these separate issues. When the objective of a valuation study is to predict person mean values there seems little reason to continue using VAS data to collect health state valuation data and many reasons to stop.

For those who believe that the MAUF is the correct formulation for utility functions over health, the cubic MAUF appears to be superior to either the power MAUF or the original McMaster MAUF, for reflecting the health state values of this UK sample. However, it should be noted that the relationship implied by the cubic mapping function is not consistent with a stable attitude to risk, and therefore the assumed relationship between VAS and SG data may not hold.

Figure1: Design of Valuation Studies in UK PICOS

Survey 1: Multi-Attribute Utility Function Valuation Survey Single attribute VAS Multi attribute VAS Multi attribute SG Socioeconomic characteristics n respondents =201	Survey 2: Statistical Inference Valuation Survey Ranking exercise Multi-attribute SG questions Socioeconomic characteristics n respondents =198 n health states = 51	Survey 3: Validation Survey Ranking exercise Multi-attribute SG questions Socioeconomic characteristics n respondents =51 n health states = 15

Analyses

1. Mapping Visual Analogue Scale data to Standard Gamble data; Survey 1
2. Estimating Multi-attribute utility functions; Survey 1
3. Testing external predictive performance of Multi-attribute utility functions; Survey 2 plus Survey 3.
4. Estimating Statistical Inference Valuation Model; Survey 2
5. Testing internal predictive performance of statistical inference model; Survey 2
6. Testing external predictive performance of statistical inference model; Survey 3
7. Comparing external predictive performance of Statistical Inference and UK and Canadian MAUF models; Survey 3

Figure 2: Estimating the Mobility and Self Care Disvalues in HUI2 MAUF

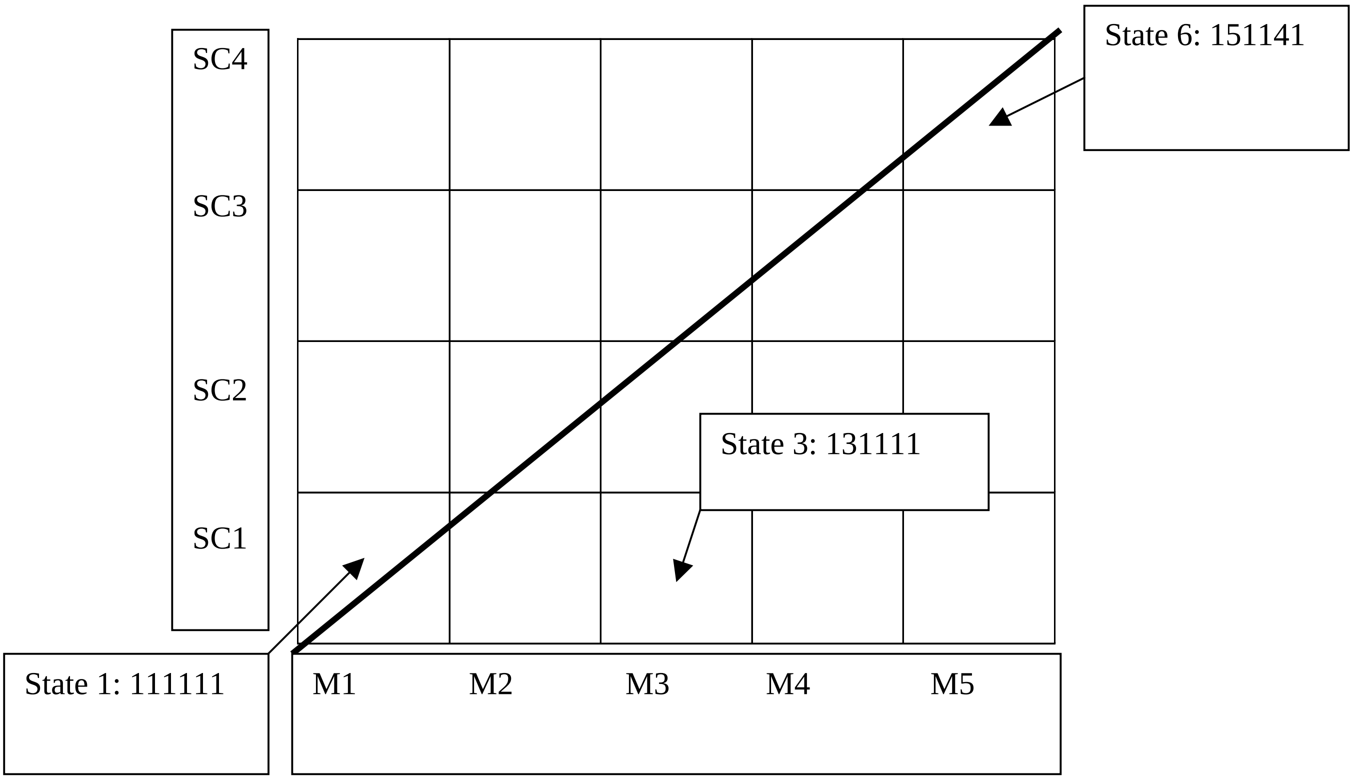

Table 1: Mean Single Attribute Disvalues by Health State (n=113)

Level	Sensation	Mobility	Emotion	Cognition	Self Care	Pain
1	0	0	0	0	0	0
2	0.368	0.311	0.351	0.387	0.39	0.244
3	0.66	0.635	0.61	0.664	0.71	0.543
4	1	0.866	0.814	1	1	0.778
5		1	1			1

Mean disvalues for Mobility attribute have been adjusted for confounding with Self care.

Table 2: Mean Disvalues of Multi-attribute Health States: PITS Full Health Scale (n=113)

State	Mean	Parameter	Modelled Value
1,1,1,1,1,1	0.000		
4,1,1,1,1,1	0.499	C1	
1,3,1,1,1,1	0.282		
1,1,5,1,1,1	0.665	C3	
1,1,1,4,1,1	0.248	C4	
1,5,1,1,4,1	0.510		
1,1,1,1,1,5	0.459	C6	
1,1,1,1,4,1		C5	0.113
1,5,1,1,1,1		C2	0.445
1,4,2,1,1,1	0.5593		
1,1,3,2,1,1	0.526359		
3,3,2,3,3,2	0.782921		
3,3,4,4,4,4	0.916649		
4,5,5,4,4,5	1		
Scalar parameter		C	-0.959

Table 3: Parameter values for the person mean multi-attribute disvalue function: PITS-Full Health Scale

Level	Sensation	Mobility	Emotion	Cognition	Self Care	Pain
1	0	0	0	0	0	0
2	0.368	0.311	0.351	0.387	0.39	0.244
3	0.66	0.635	0.61	0.664	0.71	0.543
4	1	0.866	0.814	1	1	0.778
5		1	1			1
Parameter Estimates						
d= -0.959	d_1=0.499	d_2=0.445	d_3=0.665	d_4=0.248	d_5=0.113	d_6=0.459
Mean disvalue of immediate death		0.971				

Table 4: 'Power-curve' VAS-SG mapping function

Variables	Coefficients				
Ln disutilty					
	B	**Sig.**	**R-squared**	**F**	**Sig. F.**
Ln Value	1.433	0.008	0.93	40.102	0.008

Ln disutilty = Natural Log of (1 minus Utility), Ln value = Natural Log of (1 minus Value)
Disutility of any given VAS score is calculated as: $U = V^{1.433}$

Table 5: Parameters for the person mean multi-attribute disutility function: PITS-Full Health Scale; power curve mapping function

Level	**Sensation**	**Mobility**	**Emotion**	**Cognition**	**Self Care**	**Pain**
1.0	0.000	0.000	0.000	0.000	0.000	0.000
2.0	0.239	0.188	0.223	0.257	0.259	0.132
3.0	0.551	0.522	0.492	0.556	0.612	0.417
4.0	1.000	0.814	0.745	1.000	1.000	0.698
5.0		1.000	1.000			1.000
Parameter Estimates						
D	d1	d2	d3	d4	d5	d6
-0.839	0.369	0.316	0.557	0.136	0.088	0.328
Person Mean Disutility of Immediate Death		0.959				

Table 6: Cubic VAS-SG Mapping Function

Variables	Coefficients	Sig.			
CUBUT					
	B		R-squared	F	Sig. F.
CUBVAL1	3.183	0.006	0.998	500.584	0.002
CUBVAL2	-4.901	0.022			

Where: CUBUT = (Utility – VAS^3); CUBVAL1 = (VAS-VAS^3); and CUBVAL2 = (VAS^2- VAS^3). Utility of any given VAS score is calculated as:
U= (3.183 x (VAS – VAS^3)) – (4.901 x (VAS^2 - VAS^3)) + VAS^3 Disutility= (1-Utility)

Table 7: Parameters for the person mean multi-attribute disutility function: PITS-Full Health Scale; cubic mapping function

Level	Sensation	Mobility	Emotion	Cognition	Self Care	Pain
1.0	0.000	0.000	0.000	0.000	0.000	0.000
2.0	0.260	0.245	0.256	0.264	0.265	0.22
3.0	0.378	0.359	0.343	0.381	0.423	0.310
4.0	1.000	0.655	0.560	1.000	1.000	0.505
5.0		1.000	1.000			1.000
Parameter Estimates						
d	d1	d2	d3	d4	d5	d6
-0.668	0.294	0.278	0.382	0.222	0.023	0.282
Person-mean disutility of death		0.912				

This is obtained by applying the cubic mapping function to all the directly valued parameters and solving for the d5 and d, as per the disvalue model.

Table 8: Observed and Predicted Health State Values

State	Observed	Power	Cubic	McMaster		State	Observed	Power	Cubic	McMaster
1,1,2,2,2,2	0.770	0.779	0.766	0.821		3,4,3,1,1,2	0.501	0.358	0.526	0.453
2,2,1,2,2,1	0.768	0.799	0.781	0.835		3,2,3,3,3,1	0.495	0.432	0.599	0.507
1,1,2,1,2,3	0.732	0.724	0.797	0.747		3,4,2,2,2,3	0.488	0.369	0.483	0.427
1,4,1,3,4,1	0.709	0.605	0.698	0.471		4,2,1,1,1,4	0.481	0.416	0.498	0.331
3,1,3,3,3,1	0.703	0.467	0.658	0.527		1,2,4,1,3,4	0.470	0.351	0.566	0.348
2,2,1,3,2,1	0.685	0.762	0.756	0.773		2,1,3,3,2,1	0.466	0.571	0.694	0.631
1,2,2,2,2,2	0.649	0.728	0.702	0.796		3,2,2,4,1,2	0.461	0.507	0.465	0.453
1,1,1,2,3,2	0.639	0.867	0.860	0.825		3,4,1,3,1,4	0.458	0.370	0.498	0.319
1,2,1,1,3,2	0.638	0.844	0.851	0.840		3,1,5,1,4,3	0.458	0.207	0.408	0.252
3,2,2,2,2,1	0.635	0.593	0.656	0.698		2,2,2,3,3,3	0.456	0.533	0.585	0.548
1,3,3,2,1,3	0.634	0.467	0.623	0.510		4,1,2,4,3,1	0.439	0.407	0.402	0.287
1,2,2,2,2,2	0.623	0.728	0.702	0.792		4,2,3,1,2,2	0.439	0.350	0.455	0.408
2,1,1,2,2,3	0.612	0.728	0.762	0.726		3,1,5,3,1,2	0.432	0.252	0.382	0.341
1,4,2,3,1,1	0.611	0.577	0.630	0.566		3,3,1,2,3,3	0.429	0.487	0.634	0.492
1,2,5,2,1,1	0.610	0.368	0.480	0.443		4,3,1,3,2,2	0.429	0.412	0.467	0.381
2,2,1,2,1,4	0.610	0.616	0.651	0.525		1,3,2,3,3,2	0.422	0.585	0.645	0.576
2,2,4,1,1,2	0.609	0.444	0.580	0.603		2,1,4,2,4,2	0.410	0.407	0.569	0.449
3,3,1,1,3,1	0.596	0.606	0.768	0.631		3,2,1,3,4,5	0.408	0.373	0.452	0.163
3,1,1,1,2,4	0.584	0.574	0.728	0.512		2,5,5,3,3,2	0.406	0.146	0.207	0.166
4,2,1,3,1,3	0.578	0.433	0.477	0.411		3,1,4,4,3,1	0.397	0.326	0.461	0.309
2,3,4,1,1,1	0.572	0.404	0.605	0.524		3,4,4,2,2,2	0.395	0.242	0.408	0.346
2,3,2,1,4,1	0.555	0.581	0.696	0.570		4,2,4,3,1,3	0.391	0.198	0.317	0.256
3,2,2,2,2,2	0.553	0.562	0.602	0.681		2,2,3,2,3,5	0.381	0.323	0.410	0.186
2,3,5,1,2,1	0.544	0.276	0.437	0.359		1,3,3,4,4,4	0.341	0.305	0.419	0.166
2,3,1,4,1,2	0.533	0.605	0.545	0.464		1,2,5,4,2,5	0.340	0.154	0.162	0.054
3,4,2,1,2,4	0.521	0.332	0.483	0.314		4,5,2,2,4,1	0.337	0.266	0.308	0.195
3,1,1,3,2,4	0.518	0.522	0.651	0.433		2,4,1,4,2,3	0.330	0.453	0.451	0.319
1,2,1,4,3,4	0.515	0.565	0.558	0.332		3,3,2,2,2,5	0.275	0.307	0.388	0.178
2,1,2,3,1,4	0.508	0.538	0.601	0.462		1,4,4,3,2,5	0.249	0.181	0.301	0.100
2,2,2,1,4,2	0.506	0.632	0.670	0.636		2,4,2,1,3,5	0.246	0.318	0.390	0.162
3,4,2,4,1,1	0.506	0.400	0.421	0.348		4,4,5,2,3,4	0.186	0.026	0.104	0.062
3,1,3,3,3,1	0.503	0.467	0.658	0.537		4,4,4,3,3,5	0.126	0.042	0.114	0.045
						4,5,5,4,4,5	-0.064	-0.043	-0.097	-0.038

Appendix 1: Predictive performance of Cubic and Power Curve mapping formulations in sample of used for estimating MAUF models

Table A1: Predicted and observed values for the Cubic Mapping Function

State	Observed Mean Utility	Observed Mean Value	Predicted Utility
3	0.77	0.71	0.762
7	0.68	0.52	0.712
8	0.71	0.43	0.679
10	0.47	0.21	0.477

Table A2: Predicted and observed values for the Power Curve Mapping Function

State	Observed Mean Disvalue	Predicted Utility	Observed Utility
3	0.29	0.830	0.77
7	0.48	0.651	0.68
8	0.57	0.553	0.71
10	0.79	0.287	0.47

References

[1] Feeny D Furlong W Barr RD Torrance GW Rosenbaum P Weitzman S A comprehensive multi-attribute system for classifying the health status of survivors of childhood cancer Journal of Clinical Oncology 1992; 10(6):923-928
[2] Cadman D. Goldsmith C. Torrance G. Boyle M. Furlong W. Development of a health status index for Ontario Children: Final Report to the Ontario Ministry of Health Grant Research DM648 (00633) McMaster University Hamilton Ontario December 1986
[3] Torrance GW Furlong W Feeny D et al. Multi-attribute preference functions : health utilities index. Pharmacoeconomics 1995;7(6):503-520
[4] Torrance GW Feeny DH Furlong WJ Barr RD Zhang Y Wang Q A multi-attribute utility function for a comprehensive health status classification system: Health Utilities Mark 2 Medical Care 1996;34(7):702-722
[5] Gold MR Siegel JE Russell LB Weinstein MC Cost effectiveness in Health Care and Medicine Oxford University Press 1996 Oxford
[6] Torrance GW Feeny DH Furlong WJ Barr RD Zhang Y Wang Q Multi-attribute utility function for a comprehensive health status classification system: Health Utilities Index Mark 2 Medical Care 1996;34(7):702-722
[7] McCabe C Stevens K Roberts J Brazier JE Health state values for the HUI2: initial results from a UK survey DEEM2 University of York October 2002
[8] Furlong W Feeny D Torrance GW Barr R Horsman J Guide to design and development of health state utility instrumentation CHEPA Working Paper #90-9 McMaster University Hamilton, Ontario 1990
[9] Torrance GW Zhang Y Feeny D Furlong W Barr R Multi-attribute preference functions for a comprehensive health status classification system CHEPA Working Paper 92-18 McMaster University 1992 Ontario Canada
[10] Torrance GW Furlong W Feeny D et al Multi-attribute preference functions : Health Utilities Index. Pharmacoeconomics 1995;7(6):503-520
[11] Torrance GW Feeny D Furlong W Visual Analogue Scales: do they have a role in the measurement of preferences for health states? Medical Decision Making 2001;329-334
[12] Furlong W Feeny D Torrance GW Barr R Horsman Guide to Design and Development of Health State Utility Instrumentation CHEPA Working paper Series 90-9 McMaster University 1990 Ontario Canada
[13] Feeny D Furlong W Torrance GW Goldsmith CH Zhu Z DePauw S et al Multiattribute and single-attribute utility functions for the Health Utilities Index Mark 3 System Medical Care 2002; 40(2):113-128

14 Stevens K McCabe C Brazier JE Mapping between Visual Analogue Scale and Standard Gamble data; Results from a UK study using the Health Utilities Index II Framework. DEEM 2 University of York October 2002