CENTRE FOR HEALTH PROGRAM EVALUATION

WORKING PAPER 70

Difficulty with Life and Death: Methodological Issues and Results from the Utility Scaling of the Assessment of Quality of Life (AQoL) Instrument

Professor Jeff Richardson
Director, Health Economics Unit, Centre for Health Program Evaluation

Dr Graeme Hawthorne
Deputy Director, Program Evaluation Unit, Centre for Health Program Evaluation

Neil Day
Principal Research Fellow, Health Economics Unit, Centre for Health Program Evaluation

Dr Richard Osborne
NHMRC Post Doctoral Fellow, Department of Public Health and General Practice, The University of Melbourne

Helen McNeil
Research Fellow, Health Economics Unit, Centre for Health Program Evaluation

September, 1998
ISSN 1325-0663
ISBN 1 875677 77 1

CENTRE PROFILE

The Centre for Health Program Evaluation (CHPE) is a research and teaching organisation established in 1990 to:

- undertake academic and applied research into health programs, health systems and current policy issues;
- develop appropriate evaluation methodologies; and
- promote the teaching of health economics and health program evaluation, in order to increase the supply of trained specialists and to improve the level of understanding in the health community.

The Centre comprises two independent research units, the Health Economics Unit (HEU) which is part of the Faculty of Business and Economics at Monash University, and the Program Evaluation Unit (PEU) which is part of the Department of General Practice and Public Health at The University of Melbourne. The two units undertake their own individual work programs as well as collaborative research and teaching activities.

PUBLICATIONS

The views expressed in Centre publications are those of the author(s) and do not necessarily reflect the views of the Centre or its sponsors. Readers of publications are encouraged to contact the author(s) with comments, criticisms and suggestions.

A list of the Centre's papers is provided inside the back cover. Further information and copies of the papers may be obtained by contacting:

The Co-ordinator
Centre for Health Program Evaluation
PO Box 477
West Heidelberg Vic 3081, Australia
Telephone + 61 3 9496 4433/4434 **Facsimile** + 61 3 9496 4424
E-mail CHPE@BusEco.monash.edu.au

ACKNOWLEDGMENTS

The Health Economics Unit of the CHPE is supported by Monash University.

The Program Evaluation Unit of the CHPE is supported by The University of Melbourne.

Both units obtain supplementary funding through national competitive grants and contract research.

The research described in this paper is made possible through the support of these bodies.

AUTHOR(S) ACKNOWLEDGMENTS

The research outlined in this working paper was supported by a program grant from the Victorian Health Promotion Foundation whose assistance we would like to acknowledge.

Table of Contents

List of Tables

List of Figures

Postscript: May, 2000

Following the validation study in 1999 the utility scoring algorithm was modified and, dimension[1], illness removed from the utility instrument.

See Hawthorne G, Richardson J, Day N, Osborne R and McNeil H, Construction and Utility Scaling of the Assessment of Quality of Life (AQoL) Instrument, CHPE Working Paper 101, Centre for Health Prorgram Evaluation, Monash University.

Difficulty with Life and Death: Methodological Issues and Results from the Utility Scaling of the Assessment of Quality of Life (AQoL) Instrument

1 Introduction

The quantification of 'utility' in cost utility analysis (CUA) requires two broad tasks. First, the health state under investigation must be described; secondly, a scaling technique such as the time trade-off (TTO) or standard gamble (SG) must be used to attach a numerical value to the health state such that this value measures the strength of a person's preference (utility value) for the health state. Two broad approaches to this two stage procedure have normally been used[1], namely holistic (or 'composite') and multi-attribute utility (MAU) measurement (Torrance 1986). With the first of these, a scenario or vignette is constructed which describes the health state (Step 1). The entire scenario is then 'scaled' (Step 2), ie a survey is conducted specifically to elicit 'utility' values for the scenario. With the second approach a generic 'descriptive system' or ‘descriptive instrument’ is created which is capable of describing a wide range of health states and utility weights are attached to every possible state. This is normally done by measuring a limited number of health states and using these to calibrate a model which is then used to infer the utility values of every other health state in the 'descriptive system'.[2] The model may either be derived by econometric analysis of the observed utilities (as with the EuroQoL (Williams 1995)) or through the use of decision analytic techniques to fit the simple additive model (as used in the Quality of Wellbeing Instrument (QWB) (Kaplan et al 1996) and 15D (Sintonen and Pekurinen 1993)) or a multiplicative model (the Health Utilities Index (HUI1 and 2) (Feeny, Torrance et al 1996). The fully scaled MAU instrument may then be used to estimate the utility of health states.

Both approaches have strengths and weaknesses. Holistic measurement permits a description which is tailored to a particular health state. Unique aspects of the health state, its content, its consequences, the process of health delivery, risk and prognosis may all be included in the vignette. Validation of health state specific vignettes, however, is seldom, if ever, carried out. By contrast, the generic descriptive system of the MAU approach may be unable to capture many of the nuances of the health state and be incapable of capturing the importance of the process or

[1] In principle, these two steps can be collapsed by asking patients directly the value of the health state that they are currently experiencing. In practice this approach has seldom been used.

[2] In principle every health state may be individually measured. In practice, the number of health states in the 'descriptive system' is normally so large that this is infeasible. The only example of this approach is the original Rosser Kind Index which is now seldom used because of its limited sensitivity.

context of the health state or intervention. However, this approach should, in principle, be based upon a descriptive system, the reliability and validity of which can be investigated using standard procedures.[3] After construction, the use of an MAU instrument is cheap and easy and allows the rapid estimation of utilities in the context of a longitudinal trial. This means that it is feasible to construct a time profile of each of the dimensions of health included in the instrument. Because of these respective strengths and weaknesses both techniques have a role in CUA.

To date, only a handful of generic instruments have attempted to measure utility; *viz,* the UK Rosser-Kind Index (Rosser 1993), the US QWB (Kaplan, Ganiats et al 1996), the Canadian HUI instruments (Feeny, Torrance et al 1996), the Finnish 15D (Sintonen and Pekurinen 1993) and the European EuroQoL (Kind 1996). While each of these instruments has particular strengths, to our knowledge none were constructed using normal psychometric principles to ensure construct validity and structural independence. Consider, for example, this second issue. MAU theory postulates there should be no 'redundancy' amongst items in a descriptive system. That is, a single attribute should not be described in more than one way (von Winterfeldt and Edwards 1986). If redundancy occurs then the (dis)utility of the attribute will be double counted. A sufficient (but not necessary) condition for non-redundancy is that the different scales within the instrument are orthogonal.[4] However, the requirement of non redundancy appears to be in conflict with the need for 'sensitivity' and several instruments have reduced redundancy by the adoption of very simple descriptive systems; but this simplicity has been achieved at the expense of sensitivity. Other problems also exist. Some instruments have unsatisfactory models for inferring utility values; others have adopted scaling techniques which probably do not measure utility (Richardson 1994). Consequently there is a challenge to develop a generic instrument which overcomes these weaknesses.

The AQoL project was designed to assist with meeting this challenge. Specifically the project sought to create an instrument where the descriptive system is :

- derived using correct psychometric procedures for instrument construction and hence achieves construct validity;
- sensitive to as much of the full universe of HRQoL as is practical; and
- based upon structurally independent dimensions of health.

The achievement of these properties is described elsewhere (Hawthorne, Osborne et al 1996; Hawthorne, Richardson et al 1997). The procedures adopted in this part of the project resulted in an instrument which is unique in two respects: *viz,*

3 Essentially, HRQoL is a psychometric concept, as are utilities. They cannot be directly measured, but are uniquely individual. Although instruments can be developed from other measurement – traditions such as clinometrics, economics or decision-making – this property of HRQoL suggests that the application of psychometrics is particularly appropriate during instrument construction.

4 It is not strictly necessary as scales may be 'environmentally correlated', which does not necessarily indicate double counting. Von Winterfeldt & Edwards illustrate this in the case of a manufacturing plant, the management of which is concerned with the cost of production and distribution (von Winterfeldt and Edwards 1986). These costs will correlate because each correlates with the scale of production. Despite this, there is no redundancy and each attribute is independently important. Even with this example, however, careful construction of the instrument can eliminate the correlation. There is no necessary reason why scale of production, *unit* production costs and *unit* distribution costs will correlate (if there are no economies of scale).

i it has a hierarchical descriptive structure in which structural independence is achieved between dimensions but not within dimensions. This permits greater sensitivity within dimensions. This is shown in Figure 1; and

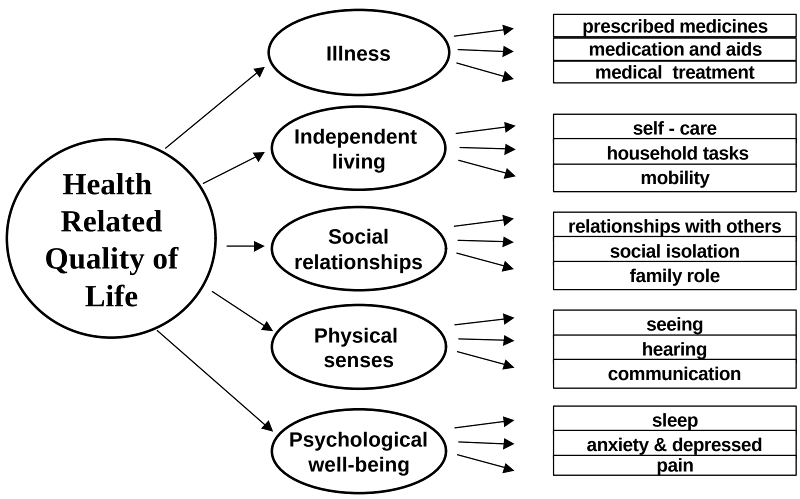
ii a descriptive system which can claim to have construct validity, which increases confidence in the validity of the health state descriptions.

Additional project objectives were:

- to scale the instrument using a flexible utility model and an accepted technique for preference measurement;
- to achieve preference independence between dimensions; and
- to achieve a valid trade-off between quality and length of life.

Preference independence was sought by the selection and content of items[5]. The achievement of this property was then assumed, as elsewhere (Feeny, Torrance et al 1996). In Section 2 of the present paper, we consider the alternative scaling techniques and models that are available for the multi-attribute estimation of utilities. In Section 3 we discuss and correct a possible error in the calculation of importance weights. The AQoL scaling survey[6] is outlined in Section 4 and the results presented in Section 5. These are used to derive the AQoL validation system in Section 6 and in Section 7 we outline the subsequent work that is needed to demonstrate the validity of an instrument such as the AQoL. The calibration of the utility model and its adjustment subsequent to initial validation is described in Section 6. Section 7 summarises future work.

Figure 1 Structural Equation Model of the AQoL



5 Preference independence indicates that the preference score for an item does not depend upon the level of another item, dimension or combination of items (see von Winterfeldt & Edwards 1993; Feeney, Torrance et al 1996).

6 The AQoL questionnaire is reproduced in Appendix 2.

2 Modeling utility

Possibly the most serious deficiency in the MAU literature is the failure to satisfactorily validate the trade-off between the quantity and quality of life implied by instruments' utility values. This omission is surprising. The *sine qua non* of the QALY (quality-adjusted life year) – its defining characteristic and claim to special status amongst QoL measurement instruments – is that it combines the quality and quantity of life according to people's preferences. It is, in effect, the exchange rate between quantity and quality. This implies a very stringent criterion for the validity of an MAU instrument; viz, that a percentage increase in the utility score (for example a 40 percent rise from 0.5 – 0.7) should be equally valuable as the same percentage increase in the number of life years (for example from 10 to 14). This was termed the 'strong interval' property by Richardson (1994). Despite this pivotal requirement, the MAU literature has totally ignored the issue and the need to validate this exchange rate. Thus, for example, the QWB may appear to have been exhaustively 'validated': a literature review reveals a wealth of 'validation' studies. The term 'validation' is, however, highly misleading.[7] The 'utility' values of the QWB imply that curing four people of pimples or a headache or a 'cough' is each as valuable as saving one life (Nord, Richardson et al 1993).

The reason for this omission may be that the validation of the QALY as an exchange rate between quantity and quality of life is particularly difficult. It is only possible to establish 'face validity', ie to use a scaling instrument which requires respondents to consider the exchange rate between quantity and quality. This is an exchange rate which explicitly involves a consideration of death. It is for this reason that most economists have argued for the use of either the Standard Gamble (SG), the TTO (Time Trade-off) or PTO (Person Trade-off). It is also the reason Rating Scale and Magnitude Estimation techniques have ceased to be serious contenders for the measurement of utility. We selected the TTO for the reasons outlined by Richardson (1994) and Dolan et al (1996). In brief, these reasons are: (i) that, as noted above, the TTO is one of three scaling techniques which requires a consideration of the value of life relative to the quality of life; (ii) each of these techniques is confounded by an irrelevant consideration (time – TTO; distributional considerations – PTO; risk that is unrelated to the medical intervention – SG). The TTO is unique to the extent that the degree of contamination may be estimated by the inclusion of tradeoffs over different time horizons. Preliminary analysis indicates that, consistent with the experience during the EuroQoL calibration exercise, the implied discount rate is 0; (iii) the TTO is simpler to apply than the SG; (iv) the PTO embodies the wrong perspective (societal not individual) and its psychometric properties have been less intensively investigated than those of the TTO.

The use of an acceptable scaling technique does not, however, ensure that the 'strong interval property' will be satisfied. The process of modeling the utility of health states inescapably involves the imposition of a particular structural relationship upon dimensions and the strong interval property may not necessarily survive the transformation of the initial utility scores.

[7] 'Validation' can be no more than an ongoing process of context specific validations which guarantee little about other contexts. However, the term 'validation' conveys the powerful impression that once an instrument has been 'validated' (ie in one context) then it is universally acceptable (in any context). The distinction also needs to be drawn between empirically demonstrated validity and reliability in the formal sense and revealed sensitivity, validity and reliability established through applied studies. Instruments which are routinely used in the medical field because of their usefulness, may or may not meet with accepted standards for validation. As noted, validity and reliability are not fixed properties, but are ongoing processes.

The Additive Model .[8] This problem is particularly obvious in the case of an additive model such as Equation 1.

$$DU* = \sum_{i=1}^{n} w_i DU*(x_i) \qquad \text{Equation (1)}$$

$$\sum_{i=1}^{n} w_i = 1 \qquad \text{Equation (2)}$$

In an *n* dimensional additive model like Equation 1, the average importance weight must be 1/n. For example, the average importance weight in the 15D is 1/15 or 0.067. Thus taking the disutility of an individual item from its lowest to its highest value can only reduce the overall utility by an average of 6.7 percent. This is grossly unrealistic. Psychological suffering can reduce a person's utility to a value close to zero. This is also true of pain. Taken alone other dimensions may also reduce utility by more than 6.7%. In the additive model a dimension such as illness can have its importance increased only if the importance of a second dimension, such as psychological distress, is correspondingly reduced. With these constraints it is impossible that several forms of suffering could each have a significant impact upon overall utility. Yet this is clearly a characteristic of HRQoL. Overall utility scores will only appear plausible in this simple model if sufficient dimensions are affected and, coincidentally, the overall disutility approximates the disutility of the illness. But this implies a 'Catch 22'. If the dimension has a significant impact upon HRQoL because it is correlated with other dimensions, then these dimensions are not structurally independent: the importance weights will be confounded by the overlap and the result will often be 'double counting'. Taken alone, coefficients will have relatively little meaning and the face validity of the strong interval property will be lost.

A related deficiency with the simple additive model shown in Equation 1 is that 'disutility' is subtracted from 'full health', where 'disutility' is equal to dimension disutility multiplied by the importance weight. This, in turn, depends upon the arbitrary number of dimensions that have been included in the instrument. This implies two other unsatisfactory characteristics of the model. First, the importance of a dimension such as illness is subject to arbitrary change as the number of dimensions change. Secondly, 'disutility' is calculated in a way that does not permit there to be any direct measurement of the exchange rate between quality and quantity of life which, as noted, is the *sine qua non* of the QALY. The model does not, at any point, require or permit the comparison of the 'model utility' the utility predicted by the model – and utility directly measured on a life death scale. The only relationship to the life/death exchange rate occurs if an appropriate scaling technique is used to calculate dimension utilities. But the properties of this scaling technique are destroyed by the rescaling carried out to satisfy Equation 2.

---

8 In this and subsequent sections the following terms are used:

U	=	Utility of a multi-dimensional health state
DU	=	Disutility of a multi-dimensional health state
DU_j	=	Disutility of 'dimension' or item j.
w_j	=	Importance weight of dimension j
k	=	Overall scaling constant used in the multiplicative model to increase the flexibility of the importance weights.
X_{ij}	=	ith response in dimension J
*	=	Indicates a value on a Life/Death Scale. Utilities on the Life/Death scale may be obtained from Equation 3.

The Multiplicative Model. Because of this flaw in the additive model, the AQoL adopted the more flexible multiplicative model which has also been used in the HUI instruments. The general form of the multiplicative model which combines dimension disutilities, DU_i, is shown in Equation 3 in which x_{ij} represents the j-th response in dimension i. The dimension weights, w_i, no longer have to satisfy Equation 2 but the more flexible constraint given by Equation 4 which must be solved for *k*, the overall scaling constant. Since $0 \leq w_i \leq 1$ it may be shown that $-1 \leq k \leq \infty$ (Keeney & Raiffa 1976; (von Winterfeldt & Edwards 1986).

$$DU = \frac{1}{k}\{\prod_{i=1}^{n}\{1 + kw_i DU(x_{ij})] - 1\} \qquad \text{Equation (3)}$$

$$k = \prod_{i=1}^{n}(1 + kw_i) - 1 \qquad \text{Equation (4)}$$

$$U^* = 1 - DU^* \qquad \text{Equation (5)}$$

where x_{ij} is the disutility of response j in dimension i.

Despite its complex appearance the multiplicative model is only a modest extension of the additive model and, if Σ w_i = 0, then the model collapses into simple additive model of Equation 1.[9]

A second characteristic of Equation 3 is that it constrains global DU values to the range 0.00-1.00 This may be seen by setting all $DU_i = 0$ from which DU = 0 and by setting all $DU_i = 1$. In the latter case Equation 3 collapses to the RHS of Equation 4 and, as this is equal to k then Equation 3 further collapses to $DU = \frac{1}{k} \cdot k = 1$. Since model values are constrained to the 0-1 range they must be converted to a disutility scale on which 1.0 and 0.0 represent death and best health respectively; i.e. 'the model utility' index numbers from the model must be converted from 'model space' to 'life death utility space'.

A further property of the multiplicative model is that the absolute values of the dimension importance weights are of significance and not simply the relative weights. This may be seen by setting all of the dimension disutility values equal to zero except for the first dimension which is set equal to unity [DU x_{ij} = 0 i $\neq$ 1,: DU (x_{1j})=1]. In this case Equation 3 reduces to Equation 6.

$$DU = \frac{1}{k}\Pi(1 + kw_1) - 1 = \frac{1}{k}(1 + kw_i - 1) = w_1 \qquad \text{Equation (6)}$$

The disutility is equal to the absolute magnitude of the importance weight. The procedure recommended for the calculation of importance weights is for each dimension to have the value

9 More generally:

If $\sum_i w_i > 1$ then $-1 \leq k < 0$;

If $\sum w_i = 1, \ k = 0$ and the multiplicative model collapses to the additive model.

If $\sum w_i \leq 1$, then $k > 0$

of the disutility empirically derived when the dimension is at its lowest values (DU_1 = 1) while all other dimensions are at their highest value (DU_i = 0, i ≠ 1) (for details see below). Thus, the numerical value of the disutility predicted by Equation 6 is the value of the importance weight, which, by construction, appears to be the correct disutility value for that state. Below we note an error which occurs when this recommended procedure is followed in the present problem context.

The significance of the absolute value, as distinct from the relative value, of the importance weights is also obvious from the fact that if only relative values were of importance, then any set of weights could be normalized so that their relative magnitudes were unchanged but the sum of the values was unity. This would mean that any model could be converted into an additive model. But clearly this imposes a different set of relationships on utility scores than the multiplicative model, which, in turn, implies that absolute as well as relative values are of importance.

While von Winterfeldt and Edwards (1986) assert that the multiplicative model represents only a modest extension of the simple additive model of Equation 1, the extension is of extreme importance for the modeling of health states and the achievement of the strong interval property. The increased flexibility of the model may be illustrated by setting k = -1. This is the lower limit which is approached when the sum of the importance weights, Σw_i , rises significantly above unity. This describes the present context in which the all-worst health states of several dimensions are sufficiently unpleasant that global utility would approach zero even when other dimensions have a disutility value of zero. When k= -1 is substituted into Equations 3 and 4 we obtain Equations 3a and 4a.

$$DU = 1 - \prod_{i=1}^{n} \left(1 - w_i DU_i\left(x_{ij}\right)\right) \qquad \text{Equation 3(a)}$$

$$0 = \prod_{i=1}^{n} \left[1 - w_i\right] \qquad \text{Equation 4(a)}$$

In the second equation one or more w_i values may (must) be close to unity (so that k approaches -1.0). It follows, that Equation 4a is consistent with any or all of the importance weights approaching unity. That is, a number of dimensions such as illness, psychological distress, etc may simultaneously have importance weights that are so large that the effect of any one of these could be sufficient to reduce utility to a very low score, or increase disutility to a correspondingly large value.

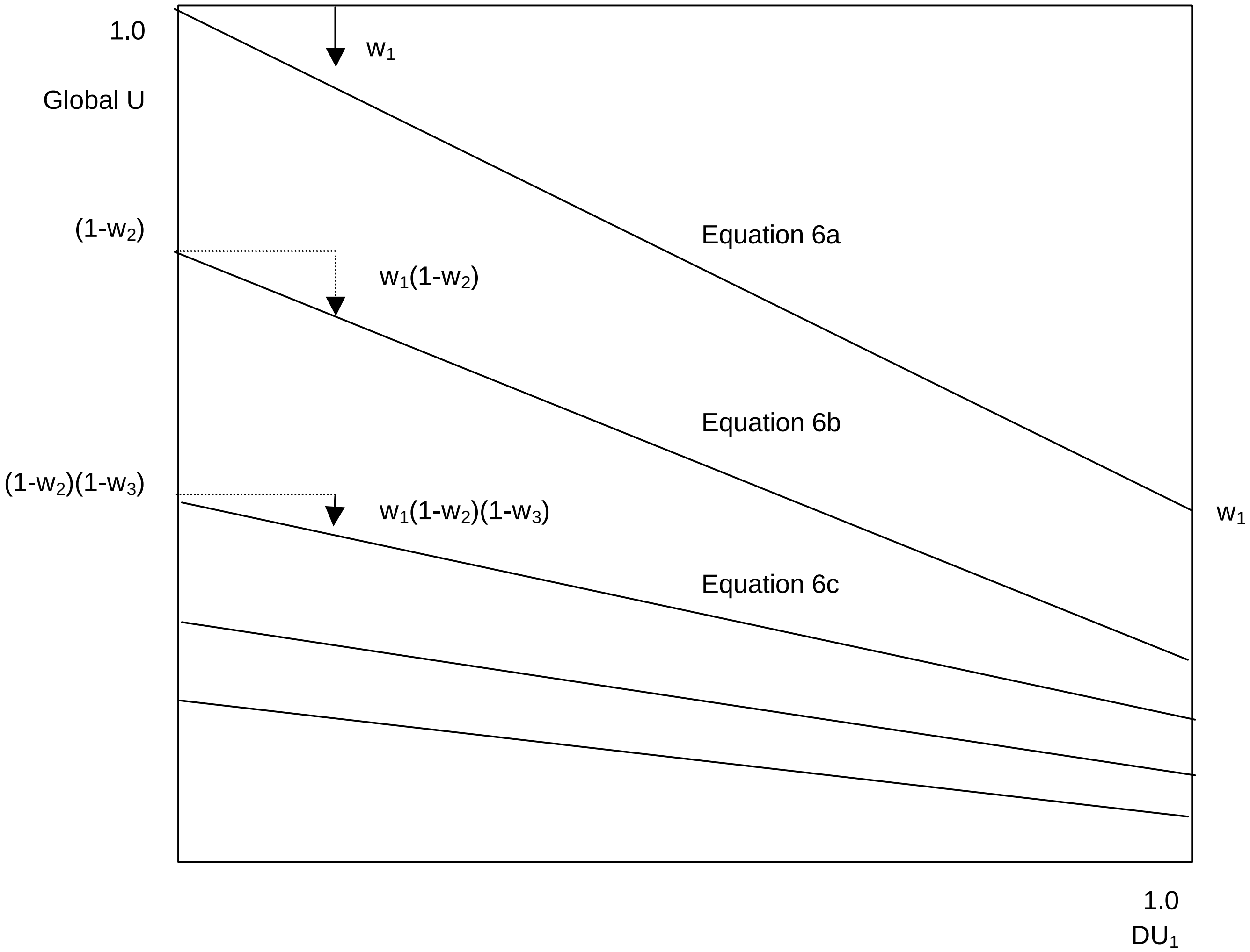
The interaction of dimensions in this case is illustrated in Figure 2. This shows the impact on global utility of an increase in the disutility score in dimension 1 when this dimension has a large importance weight w_1. Assuming first that all other dimensions have a disutility value of zero ($DU_i = 0; i \neq 1$) and substituting $U = 1 - DU$ we obtain the following relationship between global utility and DU_1.

$$U = 1 - w_1 DU_i \qquad \text{Equation (6a)}$$

As shown in Figure 2 Equation 6a has an importance weight, w_1, such that more than half of the global utility is lost when DU_1 is raised to its maximum value (1.0). When the second dimension is raised to its highest disutility score $\left(DU_i = 0 : i \neq 1,2 \text{ and } DU_2 = 1\right)$ we obtain Equation 6b.

$$U = (1 - w_2)(1 - w_1 DU_1)$$
$$U = (1 - w_2) - w_1(1 - w_2)DU_1 \qquad \text{Equation (6b)}$$

Figure 2 The effect of a declining dimension utility in a multiplicative model



As shown in Figure 2 Equation 6b indicates that the impact of an increasing DU_1 is now constrained to a much smaller range of utility; *viz,* (1 - w_2). Similarly, when DU_3 is raised to DU_3 = 1 then:

$$U = (1 - w_2)(1 - w_3) - w_1(1 - w_2)(1 - w_3)DU_1 \qquad \text{Equation (6c)}$$

Equation 6c in Figure 2 is similarly constrained.

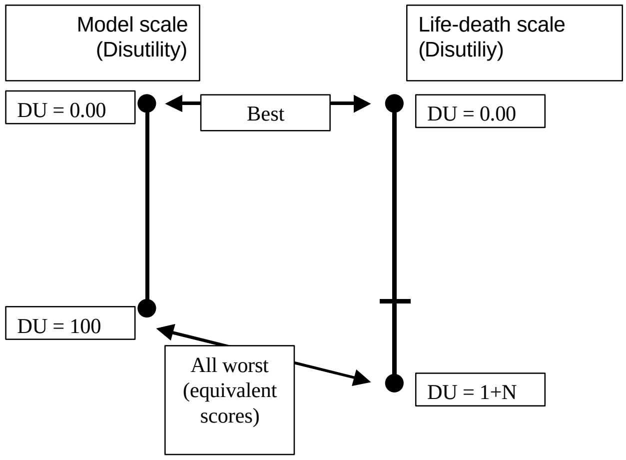
While Figure 2 illustrates the importance of the disutility arising from the first dimension with various interactions, exactly the same relationship would hold with any other dimension. In general, the multiplicative model allows any dimension, taken alone, to reduce global utility by any amount less than unity (in 'model space'). When other significant dimensions subtract from utility the impact of the dimension is scaled downwards proportionately; that is, in the multiplicative model the impact of a dimension is not fixed as in the additive model but proportional to the utility that remains after allowing for the disutility of other dimensions. (This is sometimes called second order or 'mutual' preference independence.) This proportionality property is easily shown. With the disutility of all other dimensions equal to zero, the maximum impact on disutility of dimension 1 is equal to 100 w_1 percent of total utility. With the disutility of dimension 2 at its maximum value, $(1-w_2)$ •100% of the utility remains. The full disutility from increasing DU_1 from 0 to 1.0 is now equal to $[w_1 (1-w_2) / 1-w_2]$•100 = w_1•100% of the remaining utility. In other words, the increasing disutility of the dimension has exactly the same percentage impact upon the residual utility after taking into account the disutility of other dimensions. Despite this restriction, the increased flexibility achieved by the multiplicative model is highly significant for the modeling of utility. It represents a plausible structure for approximating true utility without destroying the strong interval property. However, it is only a first approximation to the value of true utility and, as noted in the conclusion, it is our intention to carry out second order corrections based upon the econometric relationship between observed and predicted utility values.

Conversion to a Life/Death Scale As noted above Equations 3 and 4 constrain utility scores to the range 0 - 1 where these two extreme values correspond with no disutility and the disutility of the 'all worst' health state, ie the health state where all dimensions are set at their worst value. Except coincidentally, this score will not correspond with death on a disutility scale and for this reason it is necessary to convert the model score to the corresponding disutility value on a (0-1) Life-Death scale. The conversion, which is illustrated in Figure 3, in principle, is relatively straightforward when scoring is carried out with disutilities. From Figure 3 it may be seen that the full range of the model score (1-0) corresponds with a range of 1+ N in life/death 'utility space' where N is the amount by which the disutility of the all worst health state exceeds or falls short of 100, ie it is the maximum negative score which may be obtained from the instrument. Consequently it is possible to calculate disutility on a Life-Death scale with Equation 7.

$$DU^* = (1 + N)DU \qquad \text{Equation (7)}$$

where DU* indicates disutility on the Life-Death Scale.

Figure 3 Conversion to a Life Death Scale



3 Deriving importance weights: correcting an error

Although there are a variety of techniques for deriving importance weights (Keeney & Raiffa 1976; von Winterfeldt & Edwards 1986), the simplest procedure in the context of multi-attribute modeling is to commence with the all best health state – the state where all items of all dimensions are at their best level ($DU_i = 0$) and then to progressively create five new health states by replacing the best by the worst values in each dimension. As shown in Figure 4, in an instrument with five dimensions, the five health states will then have four dimensions at their best value and the fifth at its worst. The disutility scores for these five health states therefore reflect the relative importance of each dimension and the scores may be set equal to the disutility weights. These may then be used to calculate k, the scaling constant, from Equation 4. The k and w_j weights then fully parameterize the disutility function (Equation 3).

Figure 4 Evaluating importance weights

	Health states				
Dimensions	***1***	***2***	***3***	***4***	***5***
1	$\bar{X}_1$	$\dot{X}_1$	$\dot{X}_1$	$\dot{X}_1$	$\dot{X}_1$
2	$\dot{X}_2$	$\bar{X}_2$	$\dot{X}_2$	$\dot{X}_2$	$\dot{X}_2$
3	$\dot{X}_3$	$\dot{X}_3$	$\bar{X}_3$	$\dot{X}_3$	$\dot{X}_3$
4	$\dot{X}_4$	$\dot{X}_4$	$\dot{X}_4$	$\bar{X}_4$	$\dot{X}_4$
5	$\dot{X}_5$	$\dot{X}_5$	$\dot{X}_5$	$\dot{X}_5$	$\bar{X}_5$
DU-score	w_1	w_2	w_3	w_4	w_5

$\dot{X}$ All items in the dimension are set at their best values

$\bar{X}$ All items are set at their worst values

When this procedure was followed the final equation for combining dimensions predicted all worst dimension scores that were significantly below the observed disutilities that had been used for calibrating the model. As described by the authors elsewhere, the utility function obtained from the direct use of these procedures was adjusted to ensure the correct prediction of dimension all worst health states (Hawthorne, Richardson et al 1997). The *ad hoc* adjustment altered each of the dimension importance weights to within 2.5 percent of the value that would have been derived by fully consistent procedures as discussed below.

The reasons for the error in the unadjusted results arises from the fact that, from Equation 7, any DU index number in the multiplicative model (DU) corresponds with a different numerical value on the life-death scale (DU*) but, from Equation 6, the predicted DU from the model is equal to the actual DU on the life-death scale when there is only one dimension with less than full health. In essence, the error in the simple procedure is the adoption as the importance weights of values that are measured in 'true' 'life-death utility space', and to use these to estimate dimension weights in 'model space'. After transformation from 'model space' to 'life-death utility space', all disutility index numbers are multiplied by the factor (1+N); hence the disutility in 'life-death utility space' of the dimension all worst scores (which already have the correct numerical values for (life-death) utility) are erroneously inflated by the factor (1+N).

The solution to this problem is to discount the dimension importance weights, w_i, by the scale factor (1+N) and to employ these adjusted values as the importance weights. Subsequent transformation back to utility space (Equation 7) will then result in these weights correctly predicting the utility of the corresponding health states. In sum, weights are set equal to $w_i/(1+N)$ where w_i is the observed disutility of dimension i. From Equation 6 the model will then predict a value (in model space) of $w_i/(1+N)$ for the health state where dimension i is at its all worst and dimensions $j \neq i$ are at their all best. From Equation 7 this translates into a value of $(1+N) \cdot w_i/(1+N) = w_i$ in life-death utility space, which is the observed value.

This adjustment is incorporated in the models discussed below.

4 Survey summary and results

For the reasons discussed earlier, scaling was carried out using the TTO technique. Interviews were conducted with a stratified sample of 437 Victorians, selected from 15 electoral divisions representative of the Australian population. There were 249 females and 188 males in the sample. During piloting of interviews it became apparent participants found the task very tiring because of unfamiliarity with and the cognitive difficulty of the TTO task.[10] To reduce this cognitive burden, a stratified procedure was adopted by which each participant responded only to selected items. Six different interview schedules were assembled with the number of TTO-items varying from 16 to 30 depending on item complexity. The average length of interviews was 68 minutes.

Participants were asked to evaluate items selected from:

- each intermediate item response on an ‘item best-worst’ response scale;
- each ‘item worst’ response on its dimension ‘best-worst’ scale;
- each dimension’s all-worst health state on a ‘good health - death’ health scale; and
- the AQoL instrument all-worst health state on a ‘good health - death scale’.

In each case TTO scores were derived using standard TTO procedures as described by Torrance (1986). Where utilities were positive, respondents were asked to compare ten years in the health state being evaluated (utility, U_S) with T years of good health and (10-T) years of poor health as described by the end point of the scale (item worst, dimension worst, or death). T was varied until the respondent considered the two options to be equivalent. At this point the index of U was defined as U = T/10.[11] All item utilities were then converted to disutilities:

$$DU_i = 1 - U^* \qquad \text{Equation (8)}$$

Item disutilities on the item best-item worst (0, 1)scale are presented in Table 1.

10 Both the literature and practical experience suggest that, in general, people do not hold well formed opinions about the values of health states and that they are not familiar with the cognitive tasks of the TTO. Under these circumstances the spontaneous values elicited are likely to be unstable. This gives rise to the possibility that health state responses will suffer from response bias of one form or another. This issue is critical, particularly for evaluating the life-death trade when there is a very heavy cognitive burden on respondents. This is discussed in Section 6: ‘Validating the Life-Death Trade-Off’. The problem is a common one, although one that is not often discussed. However, Ware et al (1983) report that between 40-60% of respondents will exhibit some response bias and that between 2-10% will exhibit severe response bias (Ware, Snyder et al 1983). In the present study response bias was detected in about 28% of cases. The precise effects of this on the study findings are being currently investigated, as are methods of minimising response bias during data collection, and modelling it during data analysis.

11 The exceptions were the intermediate health state disutilities for each of the items. These were obtained by time trade-off using a 10-year timeframe, where the scale endpoints were each item's best health state response, (assigned a value of ‘1’), and each item's worst health state response (assigned a value of ‘0’).

Table 1 Individual item response: TTO disutility weights

	Items	Disutility values for each response level[1] Response level 1	2	3	4
Illness					
1.	I use five or more medicinal drugs regularly.	0.000	0.328	0.534	1.000
2.	I have to constantly take medicines or use a medical aid.	0.000	0.269	0.467	1.000
3.	My life is dependent upon regular medical treatment.	0.000	0.166	0.440	1.000
Independent living					
4.	I need daily help with most or all personal care tasks.	0.000	0.154	0.403	1.000
5.	I need daily help with most or all household tasks.	0.000	0.244	0.343	1.000
6.	I cannot get around either the community or my home by myself.	0.000	0.326	0.415	1.000
Social relationships					
7.	I have no close and warm relationships.	0.000	0.169	0.396	1.000
8.	I am socially isolated and feel lonely.	0.000	0.095	0.191	1.000
9.	I cannot carry out any part of my family role.	0.000	0.147	0.297	1.000
Physical senses					
10.	I only see general shapes or am blind.	0.000	0.145	0.288	1.000
11.	I hear very little indeed.	0.000	0.253	0.478	1.000
12.	I cannot adequately communicate with others.	0.000	0.219	0.343	1.000
Psychological wellbeing					
13.	I sleep in short bursts only. I am awake most of the night.	0.000	0.107	0.109	1.000
14.	I am extremely anxious worried or depressed.	0.000	0.141	0.199	1.000
15.	I suffer unbearable pain.	0.000	0.104	0.312	1.000

Notes:
[1] Scores are constrained to the range 0.00-1.00.

When a respondent indicated that the health state was worst than death, he/she was asked to choose between death and a two-part scenario; *n* years in the health state followed by (10-n) years in full health. An equivalence point was obtained by varying *n* until the respondent considered the two options to be equivalent. At that point:

$$0 = nU^*_s + (10-n)(1.00)$$

or $$U^*_s = -(10-n)/n\,. \qquad \text{Equation (9)}$$

where U^*_S is the 'apparent utility' which would be calculated from a literal interpretation of the answer given during the interview.

As *n* varies from 10.0 to 0.0 the value of U^*_S varies from zero to minus infinity. As this latter value has little meaning in terms of the usual concept of utility, it was initially constrained to the range (0, -1) using Equation 10.

$$DU = 2 + 1/(U^*_s - 1) \qquad \text{Equation (10)}$$

where DU is disutility. Thus, when $U_s^* = 0$, DU = 1. When $U_s^* = -\infty$, DU = 2. This is the constraint recommended by Torrance (1986) and used in the EuroQoL (Kind 1996). For reasons discussed in Richardson and Hawthorne (1999) Equation 10 was replaced by Equation 11 which initially approximates the unconstrained value of U* but subsequently constrains negative utility to the range (0, -0.25).

$$DU = 1.25 + 1/28.57U* - 4 \qquad \text{Equation (11)}$$

The two sets of DU-values (ie positive and negative utilities) were then combined into a single variable with the utility range of -0.25-1.00 where -0-25 is the worst observed value, 0.00 represents death and 1.00 is good health. The results from these procedures are reported in Tables 2 and 3.

Table 2 Item worst disutility values

Dimension	Item	DU	Dimension	Item	DU
1 Illness	1	0.387	4 Physical Wellbeing	10	0.414
	2	0.686		11	0.338
	3	0.566		12	0.557
2 Independent Living	4	0.666	5 Psychological Wellbeing	13	0.219
	5	0.509		14	0.332
	6	0.630		15	0.819
3 Social Relationships	7	0.730			
	8	0.646			
	9	0.691			

Table 3 Disutility Values on a Life-Death Scale

Dimension	Definition	DU
1 Illness	(444, 111, 111, 111, 111)	0.638
2 Independent Living	(111, 444, 111, 111, 111)	0.875
3 Social Relationships	(111, 111, 444, 111, 111)	0.889
4 Physical Wellbeing	(111, 111, 111, 444, 111)	0.968
5 Psychological Wellbeing	(111, 111, 111, 111, 444)	1.037
AQoL all worst	(444, 444, 444, 444, 444)	1.040

Source:
Richardson and Hawthorne, 1999.

5 Modeling quality of life

Each of the dimensions was modeled. The observed item disutilities presented in Table 2 were used to compute the constant (k) for each dimension, i, using a modified version of Equation 4:

$$k+1=(1+k\cdot DU(x_{14}))(1+k\cdot DU(x_{24}))(1+k\cdot DU(x_{34}))$$ Equation (12)

where DU_{j4} is the disutility of the worst outcome (response 4) for item j, when disutility is measured on a (0 - dimension all worst) scale.

Once the constants were computed, computation of each dimension's disutility was calculated using a modified version of Equation 3, viz:

$$DU\ (\text{model})=\frac{1}{k}\left[(1+kw_1\cdot DU(x_{ip}))(1+kw_2\cdot DU(x_{2p}))(1+kw_3\cdot DU(x_{3p}))-1\right]$$

Equation (13)

where DU_{jp} is the disutility of response, p, for item j.

The dimension utility values Ud_i were obtained by substituting DU into Equation 8 for each dimension. Substituting the values for k and w_i for each of the five dimensions produces the following five equations which predict 'model utility' scores (ie scores on a 1-0 scale) for the five dimensions.

$$U_{d1}=1.164(1-0.335DU_1)(1-0.593DU_2)(1-0.490DU_3)-0.164$$ Equation (14)

$$U_{d2}=1.099(1-0.610DU_1)(1-0.464DU_2)(1-0.573DU_3)-0.0989$$ Equation (15)

$$U_{d3}=1.04(1-0.702DU_1)(1-0.625DU_2)(1-0.664DU_3)-0.0340$$ Equation (16)

$$U_{d4}=1.656(1-0.248DU_1)(1-0.205DU_2)(1-0.338DU_3)-0.656$$ Equation (17)

$$U_{d5}=1.292(1-0.170DU_1)(1-0.255DU_2)(1.-0.635DU_3)-0.292$$ Equation (18)

where DU_m is the disutility of item m in the relevant dimension, as shown in Table 1 and the numerical coefficients are the item weights shown in Table 2 times the scaling constant k obtained by solving Equation 12 for each dimension. The 5 sets of disutility scores are reproduced in Appendix 1.

The five dimension all-worst scores in Table 3 were similarly used as dimension weights and the corresponding scale constant calculated to derive the global utility model using the following three formulae:

$$DU(model)=1-(1-.613DU_1)(1-.841DU_2)(1-.855DU_3)(1-.931DU_4)(1-.997DU_5)$$

Equation (19)

where DU (model) = disutility on the (01) 'model' scale and DU_i is the dimension disutility score for dimension i on a (0-1) scale predicted from Equations 14 - 18 or given in the 'look up' tables in Appendix 1.

$$DU(LD) = 1.04 \cdot DU(\text{model}) \qquad \text{Equation (20)}$$

where 1.04 is the empirically derived conversion factor, between 'model' and 'life death utility

$$U = 1 - DU(LD) \qquad \text{Equation (21)}$$

Substituting (19) and (20) in Equation 21 gives Equation 22 which is the equation for the health state utility, U, predicted by the AQoL on a scale from 1.0 (best) to – 0.04 (worst).

$$U = 1.04(1-.613DU_1)(1-.841DU_2)(1-.855DU_3)(1-.931DU_4)(1-.977DU_5) - 0.04 \qquad \text{Equation (22)}$$

6 Validating the life death trade off

As in the case of the HUI, the results reported here convert the utility index number measured in 'model space' to disutility measured in 'utility space'. The validity of this process depends upon the validity of the 'bridge' which is established between utility scores in the model ('model utility') and utility scores measured on the life-death scale. In Part 2 we argued that this vital exchange rate has received very little attention. Torrance, Zhang et al (1992) indirectly address the issue by predicting the empirical value of several independently measured multi-attribute health states. This is the only reference we have found where this exchange rate has been even obliquely discussed. Our concern is two-fold. First, the all-worst health state used for the transformation to life death utility space involves a scenario with 15 pieces of information. Even with some economising in our presentation it is likely that respondents would find such a state difficult to appreciate. Second, and more importantly, we are concerned that, as in any QALY measurement (holistic or multi-attribute), persons who had not experienced the health state would rate it lower than the equivalent rating provided by patients who had some experience of the health state or similar states. In particular, it is likely that the response to an unpleasant health state would evoke an immediate 'shock-horror' reaction, the magnitude of the resulting distortion being proportional to the unpleasantness of the scenario.

This issue and the appropriate treatment of negative scores are considered in detail in a companion article (Richardson and Hawthorne 1999). In this, the bridge between model and life-death utility space and the corresponding value of the AQoL all-worst health state are based upon 13 estimates of the relationship which employ seven independent observations. The outcome of this analysis was an estimated AQoL all-worst of 1.04 and this value is incorporated in Equations 19 – 22.

In the present study we obtained a number of multi-attribute health states which were designed to test and validate the exchange rate based upon the empirical value of the model all-worst health state. These values and the corresponding values predicted by Equation 19 are also presented in Richardson and Hawthorne (1999).

7 Future work and conclusions

This paper has explored several issues arising from the application of multi-attribute utility theory to the measurement of HRQoL. While the model estimated and reported here represents, in many respects, an advance on the MAU models in current use, confidence in the model and the validity of its utility estimates requires ongoing research. The AQoL is not, of course, unique in this respect. Our current research effort is focussed upon the following issues:

- the effect of response bias;
- the discount rate which should be applied; and
- the life-death relationship.

In addition, we have been funded to:

- undertake a validation study directly comparing the leading utility instruments;
- undertake a validation of the utility values reported here using TTO scores directly elicited from patients;
- assess the application of the AQoL and WHOQoL-AQoL BREF to those suffering psychoses; and
- extend the AQoL to health promotion, and particular population sub-groups (the aged, those from non-English-speaking backgrounds and adolescents).

We are also planning to explore:

- the relationship between 'spontaneous weights' (obtained immediately upon presentation of the question) deliberative weights (values elicited after discussion and contemplation); and to investigate the relationship between the AQoL and the DALY.

To date, the descriptive system has been subject to sufficient validation for it to be recommended for use with some confidence. Results from the scaling exercise reported here are less secure. The bridge between 'model' and 'life death' utilities has been based upon seven (7) independent sets of observations. This does not, however, represent a fully satisfactory validation of the relationship. We would emphasise, however, that no other instrument, to our knowledge, has achieved greater validation with respect to the utility score. The HUI II predicts the utility value of a limited number of 'observation'. However, as discussed in Richardson and Hawthorne (1999) there are grounds for concern about the validity of the 'observations' arising from the further treatment of negative utility scores. There is a similar concern with respect to the TTO values embodied in the EuroQoL scoring system. Other instruments have not attempted to validate utility scores.

In sum, results reported here can be recommended with as great or greater confidence as is warranted by other instruments. However we anticipate that as we receive additional information there will be further revisions made to the scoring system.

References

Dolan P, Gudex C et al 1996, 'Valuing health states: A comparison on methods', *Journal of Health Economics*, vol 15, pp209–231.

Feeny D, Torrance G et al 1996, Health utilities index. *Quality of Life and Pharmacoeconomics in clinical trials*, B Spilker, Philadelphia, Lippincott-Raven Publishers.

Hawthorne G, Osborne G et al 1996, The Australian Multi-attribute Utility (AMAU): Construction and Initial Evaluation, Melbourne, Centre for Health Program Evaluation.

Hawthorne G, Richardson J et al 1988, Utility Weights for the Assessment of Quality of Life (AQoL) Instrument, Melbourne, Centre for Health Program Evaluation.

Hawthorne G, Richardson J et al 1997, The Assessment of Quality of Life (AQoL) Instrument: Construction, Initial Validation and Utility Scaling, Melbourne, Centre for Health Program Evaluation.

Kaplan R, Ganiats T et al 1996, The Quality of Well-being Scale, *Medical Outcomes Trust Bulletin*, vol 4, pp2–3.

Keeney R & Raiffa H 1976, Decisions with multiple objectives: preferences and value trade-offs, New York, Wesley pp224–241.

Kind P 1996, The EuroQoL instrument: an index of health-related quality of life, *Quality of Life and Pharmacoeconomics in Clinical Trials*, B. Spilker, Philadelphia, Lippincott-Raven Publishers.

Nord E, Richardson J et al 1993, 'Social evaluation of health care versus personal evaluation of health states, Evidence on the validity of four health-state scaling instruments using Norwegian and Australian surveys', *Int. J. Technol. Assess Health Care*, vol 9 (4), pp463–78.

Richardson J 1994, 'Cost utility analysis: what should be measured?', *Social Science and Medicine*, vol 39(1), pp7–21.

Richardson J & Hawthorne G 1999, Negative Utility Scores and Evaluating the AQoL All Worst Health States, Working Paper No 73, Melbourne, Centre for Health Program Evaluation.

Rosser R 1993, A health index and output measure, *Quality of Life Assessment: Key Issues in the 1990s*, S Walker & R Rosser, Dordrecht, Kluwer Academic Publishers.

Sintonen H & Pekurinen M 1993, A fifteen-dimensional measure of health-related quality of life (15D) and its applications, *Quality of Life Assessment*, S Walker & R Rosser, Dordrecht, Kluwer Academic Publishers.

Torrance G 1986, 'Measurement of health state utilities for economic appraisal: a review', *Journal of Health Economics*, vol 5, pp1–30.

Torrance G & Zhang Y et al, 1992, Multi-attribute preference functions for a comprehensive health status classification system, Hamilton, Ontario, CHEPA.

Ware J & Snyder M et al 1983, 'Defining and measuring patient satisfaction with medical care', *Evaluation and Program Planning*, vol 6, pp247–263.

Williams A 1995, The measurement and valuation of health: final report on the modelling of valuation tariffs, York, MVH Group, Centre for Health Economics, University of York.

von Winterfeldt DV & Edwards W 1986, *Decision Analysis and Behavioural Research*, Cambridge, Cambridge University Press.

Appendix 1
Look up Tables

The following five sets of numbers give the 'disutility' score for each of the 5 dimensions on a scale of 0 (best) to 100 (worst). They may be inserted in Equation 22 (page 16) to obtain the overall AQoL utility score.

Table A 1 Illness items, health state levels and disabilities

Items & health state levels				Items & health state levels			
1	**2**	**3**	**Disutilities [1]**	**1**	**2**	**3**	**Disutilities [1]**
1	1	1	0.0000	3	1	1	0.2083
1	1	2	0.0946	3	1	2	0.2860
1	1	3	0.2508	3	1	3	0.4142
1	1	4	0.5700	3	1	4	0.6763
1	2	1	0.1856	3	2	1	0.3607
1	2	2	0.2651	3	2	2	0.4260
1	2	3	0.3964	3	2	3	0.5338
1	2	4	0.6647	3	2	4	0.7541
1	3	1	0.3222	3	3	1	0.4728
1	3	2	0.3907	3	3	2	0.5290
1	3	3	0.5036	3	3	3	0.6218
1	3	4	0.7345	3	3	4	0.8113
1	4	1	0.6900	3	4	1	0.7748
1	4	2	0.7285	3	4	2	0.8065
1	4	3	0.7921	3	4	3	0.8587
1	4	4	0.9222	3	4	4	0.9654
2	1	1	0.1279	4	1	1	0.3900
2	1	2	0.2121	4	1	2	0.4529
2	1	3	0.3512	4	1	3	0.5568
2	1	4	0.6353	4	1	4	0.7690
2	2	1	0.2931	4	2	1	0.5134
2	2	2	0.3639	4	2	2	0.5663
2	2	3	0.4808	4	2	3	0.6536
2	2	4	0.7196	4	2	4	0.8320
2	3	1	0.4147	4	3	1	0.6043
2	3	2	0.4757	4	3	2	0.6498
2	3	3	0.5762	4	3	3	0.7249
2	3	4	0.7817	4	3	4	0.8784
2	4	1	0.7421	4	4	1	0.8488
2	4	2	0.7764	4	4	2	0.8745
2	4	3	0.8330	4	4	3	0.9168
2	4	4	0.9487	4	4	4	1.0032

1 = Based on means

Table A 2 Independent living items, health state levels and disabilities

Items & health state levels				Items & health state levels			
4	5	6	Disutilities [1]	4	5	6	Disutilities [1]
1	1	1	0.0000	3	1	1	0.2700
1	1	2	0.2054	3	1	2	0.4249
1	1	3	0.2615	3	1	3	0.4672
1	1	4	0.6300	3	1	4	0.7452
1	2	1	0.1244	3	2	1	0.3639
1	2	2	0.3066	3	2	2	0.5012
1	2	3	0.3563	3	2	3	0.5388
1	2	4	0.6831	3	2	4	0.7853
1	3	1	0.1749	3	3	1	0.4020
1	3	2	0.3476	3	3	2	0.5322
1	3	3	0.3948	3	3	3	0.5678
1	3	4	0.7046	3	3	4	0.8015
1	4	1	0.5100	3	4	1	0.6547
1	4	2	0.6201	3	4	2	0.7377
1	4	3	0.6501	3	4	3	0.7604
1	4	4	0.8476	3	4	4	0.9094
2	1	1	0.1032	4	1	1	0.6700
2	1	2	0.2893	4	1	2	0.7502
2	1	3	0.3401	4	1	3	0.7720
2	1	4	0.6740	4	1	4	0.9159
2	2	1	0.2159	4	2	1	0.7186
2	2	2	0.3810	4	2	2	0.7897
2	2	3	0.4260	4	2	3	0.8091
2	2	4	0.7221	4	2	4	0.9366
2	3	1	0.2617	4	3	1	0.7383
2	3	2	0.4182	4	3	2	0.8057
2	3	3	0.4609	4	3	3	0.8241
2	3	4	0.7417	4	3	4	0.9450
2	4	1	0.5653	4	4	1	0.8691
2	4	2	0.6650	4	4	2	0.9120
2	4	3	0.6922	4	4	3	0.9237
2	4	4	0.8712	4	4	4	1.0008

1 = Based on means

Table A 3 Social relationships items, health state levels and disabilities

Items & health state levels				Items & health state levels			
7	8	9	Disutilities [1]	7	8	9	Disutilities [1]
1	1	1	0.0000	3	1	1	0.2891
1	1	2	0.1014	3	1	2	0.3623
1	1	3	0.2049	3	1	3	0.4370
1	1	4	0.6900	3	1	4	0.7872
1	2	1	0.0617	3	2	1	0.3337
1	2	2	0.1572	3	2	2	0.4025
1	2	3	0.2545	3	2	3	0.4728
1	2	4	0.7108	3	2	4	0.8022
1	3	1	0.1241	3	3	1	0.3787
1	3	2	0.2135	3	3	2	0.4423
1	3	3	0.3046	3	3	3	0.5090
1	3	4	0.7317	3	3	4	0.8173
1	4	1	0.6500	3	4	1	0.7583
1	4	2	0.6880	3	4	2	0.7858
1	4	3	0.7268	3	4	3	0.8138
1	4	4	0.9085	3	4	4	0.9450
2	1	1	0.1234	4	1	1	0.7300
2	1	2	0.2128	4	1	2	0.7602
2	1	3	0.3040	4	1	3	0.7910
2	1	4	0.7315	4	1	4	0.9354
2	2	1	0.1778	4	2	1	0.7484
2	2	2	0.2619	4	2	2	0.7768
2	2	3	0.3477	4	2	3	0.8058
2	2	4	0.7498	4	2	4	0.9416
2	3	1	0.2328	4	3	1	0.7670
2	3	2	0.3115	4	3	2	0.7936
2	3	3	0.3918	4	3	3	0.8207
2	3	4	0.7683	4	3	4	0.9479
2	4	1	0.6962	4	4	1	0.9235
2	4	2	0.7297	4	4	2	0.9348
2	4	3	0.7639	4	4	3	0.9464
2	4	4	0.9241	4	4	4	1.0005

1 = Based on means

Table A 4 Physical senses items, health state levels and disabilities

Items & health state levels				Items & health state levels			
10	**11**	**12**	**Disutilities [1]**	**10**	**11**	**12**	**Disutilities [1]**
1	1	1	0.0000	3	1	1	0.1181
1	1	2	0.1226	3	1	2	0.2320
1	1	3	0.1921	3	1	3	0.2965
1	1	4	0.5600	3	1	4	0.6381
1	2	1	0.0860	3	2	1	0.1980
1	2	2	0.2023	3	2	2	0.3059
1	2	3	0.2681	3	2	3	0.3671
1	2	4	0.6169	3	2	4	0.6910
1	3	1	0.1625	3	3	1	0.2690
1	3	2	0.2731	3	3	2	0.3717
1	3	3	0.3357	3	3	3	0.4299
1	3	4	0.6675	3	3	4	0.7380
1	4	1	0.3400	3	4	1	0.4388
1	4	2	0.4375	3	4	2	0.5243
1	4	3	0.4926	3	4	3	0.5756
1	4	4	0.7850	3	4	4	0.8471
2	1	1	0.0595	4	1	1	0.4100
2	1	2	0.1777	4	1	2	0.5023
2	1	3	0.2446	4	1	3	0.5545
2	1	4	0.5993	4	1	4	0.8313
2	2	1	0.1424	4	2	1	0.4747
2	2	2	0.2545	4	2	2	0.5622
2	2	3	0.3179	4	2	3	0.6117
2	2	4	0.6542	4	2	4	0.8741
2	3	1	0.2161	4	3	1	0.5323
2	3	2	0.3228	4	3	2	0.6155
2	3	3	0.3831	4	3	3	0.6626
2	3	4	0.7031	4	3	4	0.9122
2	4	1	0.3872	4	4	1	0.6658
2	4	2	0.4812	4	4	2	0.7391
2	4	3	0.5344	4	4	3	0.7806
2	4	4	0.8163	4	4	4	1.0006

1 = Based on means

Table A 5 Psychological wellbeing items, health state levels and disabilities

Items & health state levels				Items & health state levels			
13	14	15	Disutilities [1]	13	14	15	Disutilities [1]
1	1	1	0.0000	3	1	1	0.0240
1	1	2	0.0853	3	1	2	0.1077
1	1	3	0.2558	3	1	3	0.2751
1	1	4	0.8200	3	1	4	0.8288
1	2	1	0.0465	3	2	1	0.0696
1	2	2	0.1287	3	2	2	0.1503
1	2	3	0.2932	3	2	3	0.3117
1	2	4	0.8370	3	2	4	0.8454
1	3	1	0.0657	3	3	1	0.0884
1	3	2	0.1466	3	3	2	0.1679
1	3	3	0.3085	3	3	3	0.3268
1	3	4	0.8440	3	3	4	0.8523
1	4	1	0.3300	3	4	1	0.3479
1	4	2	0.3935	3	4	2	0.4102
1	4	3	0.5205	3	4	3	0.5348
1	4	4	0.9406	3	4	4	0.9471
2	1	1	0.0235	4	1	1	0.2200
2	1	2	0.1073	4	1	2	0.2908
2	1	3	0.2747	4	1	3	0.4323
2	1	4	0.8286	4	1	4	0.9004
2	2	1	0.0692	4	2	1	0.2586
2	2	2	0.1499	4	2	2	0.3268
2	2	3	0.3114	4	2	3	0.4632
2	2	4	0.8453	4	2	4	0.9145
2	3	1	0.0880	4	3	1	0.2745
2	3	2	0.1675	4	3	2	0.3416
2	3	3	0.3264	4	3	3	0.4760
2	3	4	0.8522	4	3	4	0.9203
2	4	1	0.3475	4	4	1	0.4938
2	4	2	0.4099	4	4	2	0.5465
2	4	3	0.5345	4	4	3	0.6519
2	4	4	0.9470	4	4	4	1.0004

1 = Based on means

Appendix 2
Assessment of Quality of Life (AQoL) Questionnaire

TICK ☑ ONE BOX for each question to show which statement best describes you during the last week.

1. Concerning my use of prescribed medicines:

   I do not or rarely use any medicines at all

   I use one or two medicinal drugs regularly

   I need to use three or four medicinal drugs regularly

   I use five or more medicinal drugs regularly.

2. To what extent do I rely on medicines or a medical aid? (NOT glasses or a hearing aid.) (*For example: walking frame, wheelchair, prosthesis etc.*)

   I do not use any medicines and/or medical aids

   I occasionally use medicines and/or medical aids

   I regularly use medicines and/or medical aids

   I have to constantly take medicines or use a medical aid.

3. Do I need regular medical treatment from a doctor or other health professional?

   I do not need regular medical treatment

   although I have some regular medical treatment, I am not dependent on this

   I am dependent on having regular medical treatment

   my life is dependent upon regular medical treatment.

4. Do I need any help looking after myself?

   I need no help at all

   occasionally I need some help with personal care tasks

   I need help with the more difficult personal care tasks

   I need daily help with most or all personal care tasks.

5. When doing household tasks: (*For example: preparing food, gardening, using the video recorder, radio, telephone or washing the car.*)

   I need no help at all

   occasionally I need some help with household tasks

   I need help with the more difficult household tasks

   I need daily help with most or all household tasks.

6. Thinking about how easily I can get around my home and community:

I get around my home and community by myself without any difficulty

I find it difficult to get around my home and community by myself

I cannot get around the community by myself, but I can get around my home with some difficulty

I cannot get around either the community or my home by myself.

7. Because of my health, my relationships (*for example: with my friends, partner or parents*) generally:

   are very close and warm

   are sometimes close and warm

   are seldom close and warm

   I have no close and warm relationships.

8. Thinking about my relationship with other people:

   I have plenty of friends, and am never lonely

   although I have friends, I am occasionally lonely

   I have some friends, but am often lonely for company

   I am socially isolated and feel lonely.

9. Thinking about my health and my relationship with my family:

   my role in the family is unaffected by my health

   there are some parts of my family role I cannot carry out

   there are many parts of my family role I cannot carry out

   I cannot carry out any part of my family role.

10. Thinking about my vision, including when using my glasses or contact lenses if needed:

    I see normally

    I have some difficulty focusing on things, or I do not see them sharply. For example: small print, a newspaper or seeing objects in the distance.

    I have a lot of difficulty seeing things. My vision is blurred. For example: I can see just enough to get by with.

    I only see general shapes, or am blind. For example: I need a guide to move around.

11. Thinking about my hearing, including using my hearing aid if needed:

I hear normally

I have some difficulty hearing or I do not hear clearly. For example: I ask people to speak up, or turn up the TV or radio volume.

I have difficulty hearing things clearly. For example: Often I do not understand what is said. I usually do not take part in conversations because I cannot hear what is said.

I hear very little indeed. For example: I cannot fully understand loud voices speaking directly to me.

12. When I communicate with others: (*For example: by talking, listening, writing or signing.*)

I have no trouble speaking to them or understanding what they are saying

I have some difficulty being understood by people who do not know me. I have no trouble understanding what others are saying to me.

I am only understood by people who know me well. I have great trouble understanding what others are saying to me.

I cannot adequately communicate with others.

13. Thinking about how I sleep:

I am able to sleep without difficulty most of the time

My sleep is interrupted some of the time, but I am usually able to go back to sleep without difficulty

My sleep is interrupted most nights, but I am usually able to go back to sleep without difficulty

I sleep in short bursts only. I am awake most of the night.

14. Thinking about how I generally feel:

I do not feel anxious, worried or depressed

I am slightly anxious, worried or depressed

I feel moderately anxious, worried or depressed

I am extremely anxious, worried or depressed.

15. How much pain or discomfort do I experience:

none at all

I have moderate pain

I suffer from severe pain

I suffer unbearable pain.