Accounting for Initial Disease Severity When Measuring Health Gains

Matthew Taylor [a], Sue Chilton [b], Sarah Whitehead [a]

a. York Health Economics Consortium, University of York, Heslington, York, UK.

b. University of Newcastle upon Tyne, UK.

Main author contact information:

Matthew Taylor

Senior Consultant

York Health Economics Consortium Ltd

Level 2 Market Square

University of York

York YO10 5NH

Telephone: +44 (0) 1904 433631

Fax: +44 (0) 1904 433628

E-Mail: mt25@york.ac.uk

1. Background

Until recent years, it has been the objective of healthcare policy makers to maximise the production of health within budget constraints. In simple terms, this has meant maximising the total health-related utility of a population. In traditional cost-utility analysis, an increase in health-related quality of life (HRQL) utility from 0 to 0.5 is considered to be of equal value to an increase from 0.5 to 1 [1; 2]. Recent evidence however, suggests that the initial state of health may have some impact upon the individual's valuation of any given health improvement [3]. Moreover, it is likely that a given health improvement will be valued more if the initial state is low, and this will be particularly evident for initial states close or equal to zero utility [4].

In this case, under normal cost-utility analysis simply stating that a treatment costs say, £10,000 per quality-adjusted life year (QALY) gained is not enough. Information would need to be made available on the initial and final states of the treatment. This paper proposes a method that would provide a measurement of any health change, adjusted to account for the initial health state, as well as experimental tests of this method. This adjusted measurement could then be used in normal cost-utility analysis.

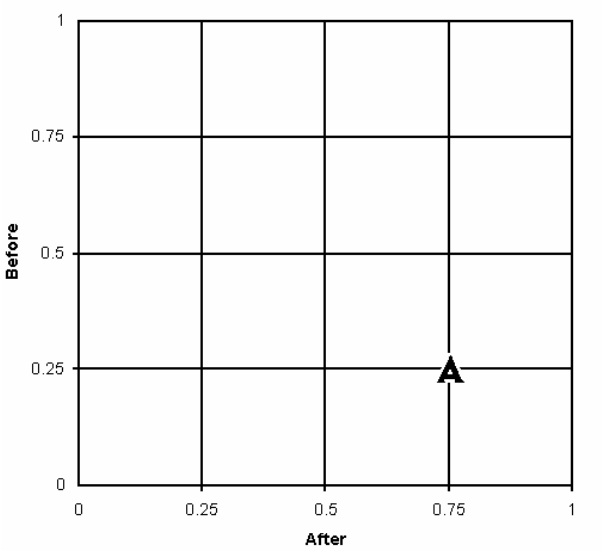
Whilst much research has been made into society's preferences for the allocation of health and social welfare [5-7], this paper aims to consider a method to place a valuation on any given health change at a purely *individual* level. The method is compatible with existing methods of health state utility measurement such as the standard gamble, time trade-off, person trade-off or visual analogue scale, and simply utilises these measurements to determine a valuation of the health *change*.

2. Methods

The Health Change Matrix

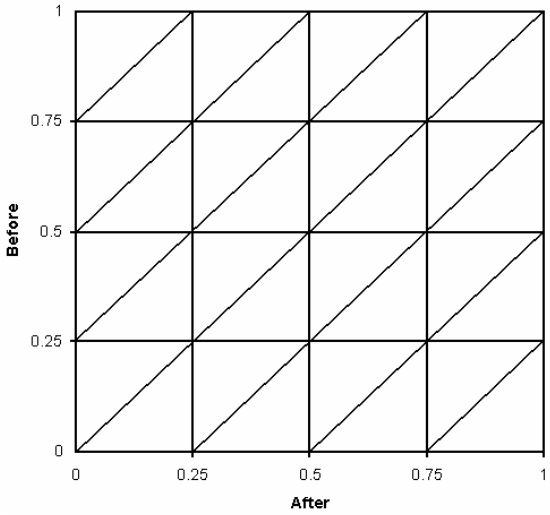
Any specific change in health can be represented by a single point in the 'health change matrix' (HCM) shown below (Figure 1). This is similar to the 'societal preference matrix' proposed for another purpose by Ubel, Richardson and Menzel [7]. The vertical axis shows the HRQL utility (as determined by existing methods of utility elicitation) before the health change, and the horizontal axis represents the HRQL utility after the change (again, as determined by existing methods). Any points lying on a 45° sloped line from the origin would show no change, any points above and to the left of this line show falls in HRQL, and points below and to the right show health improvements. The larger the health gain, the closer to the lower right extreme of the matrix, which represents an improvement in utility from 0 to 1. Point 'A' represents an improvement from 0.25 to 0.75.

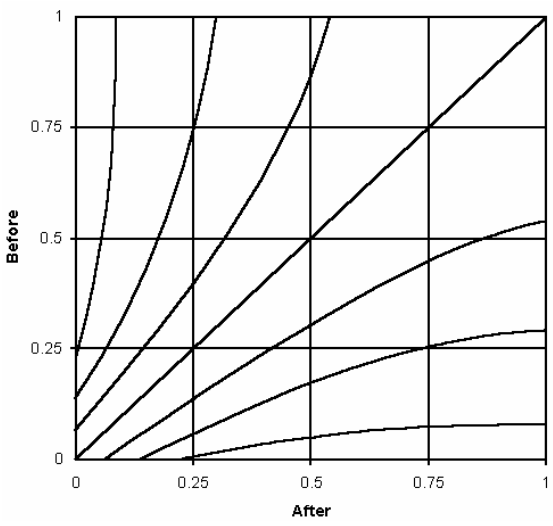
Figure 1: Health Change Matrix



Under expected utility theory (EUT), individuals should value changes of the same magnitude equally, and would therefore exhibit parallel indifference curves sloped at 45°, as shown in Figure 2a. If 'equal' improvements are valued more when the initial state is worse and 'equal' decreases are preferred when the initial state is higher, the indifference curves will appear to curve and 'fan out', as shown in Figure 2b.

Figures 2a and 2b: Parallel and 'Fanning' Indifference Curves





Two experiments were performed, each using different methods of representing health states. The first experiment used the visual analogue scale (VAS) method, whilst the second used the Health Utilities Index: Mark 3 (HUI:3) method of describing health states. This minimised the potential difficulties encountered by participants either: (i) being shown an exact utility estimation, but failing to associate it with an actual state of health; or (ii) reading a description of a health state, but placing a different valuation to that which has already been elicited. Since the weakness of each method is avoided by the methods of the other, it is hoped that the findings will have greater robustness if the two methods produce complementing results.

Experiment I.

This experiment (n=35, undergraduate students from the Universities of York and Newcastle) uses VAS methods to represent health states to participants. Prior to the experiment, participants were shown a number of descriptions of health states and asked to estimate the position of these states on the VAS in order to familiarise them with the VAS method of presentation. The presentation of the VAS states is shown in Appendix A.

Six questions were asked to each participant. Individuals were shown two health changes, 'X' and 'Y', each with different initial states, and were asked to choose which offered the 'best' change. It was explicitly stated that the participant should imagine himself/herself in those positions, as opposed to choosing between other individuals. The magnitude of change 'Y' was altered until the point of indifference between 'X' and 'Y' was determined. If the participant did not explicitly state indifference between any two specific changes, it was assumed that the point of indifference was exactly halfway between the two changes where the preference changed from 'X' to 'Y'. The

magnitude of change 'Y' in the initial presentation as well as subsequent questions was randomised in order to avoid 'start point' biases. Table 1 shows the comparator changes shown to the participants. The ordering of questions asked was randomised in order to reduce the chances of ordering biases such as fatigue or experience.

Table 1: Comparisons used in Experiment I

Health gains	Comparator Improvement		Equivalent Improvement	
	Initial Utility	**Final Utility**	**Initial Utility**	**Final Utility**
G1A	0.75	1.00	0.50	?
G1B	0.75	1.00	0.25	?
G1C	0.75	1.00	0.00	?
G1D	0.50	1.00	0.25	?
G1E	0.50	1.00	0.00	?
G1F	0.25	1.00	0.00	?
Health losses	**Comparator Reduction**		**Equivalent Reduction**	
	Initial Utility	**Final Utility**	**Initial Utility**	**Final Utility**
L1A	1.00	0.75	?	0.50
L1B	1.00	0.75	?	0.25
L1C	1.00	0.75	?	0.00
L1D	1.00	0.50	?	0.25
L1E	1.00	0.50	?	0.00
L1F	1.00	0.25	?	0.00

In accordance with the transitivity axiom of EUT[1], all of the improvements in responses 1A, 1B, 1C and the comparator are considered to be of equal value, and as such will lie upon the same indifference curve. Responses to for the gains in questions 1D, 1E and their comparator will provide a second (and 'better') indifference curve, whilst the response to 1F and its comparator will show a ('better still') third. Similar curves will be

[1] Transitivity states that if A is considered to be equal to B, and that B is considered to be equal to C, then A must be considered to be equal to C.

inferred for the questions involving health losses. As such, six indifference curves will be elicited.

Experiment II.

The second experiment (n=35, undergraduate students from the universities of York and Newcastle) follows the same method as the first. The health states are represented this time by the HUI:3 method using more descriptive methods [8; 9]. The HUI:3 method describes each health state using eight dimensions, with various levels of symptom for each factor. 18 combinations were chosen to represent HRQL utilities at equal intervals, ranging from 0 to 1. To avoid the problem of ranking inconsistencies, care was taken to ensure that, for each card, all eight indicators were either equal or better than for the corresponding dimensions in the adjacent lower card. Appendix B shows an example of a card, as shown to participants. Before the experiment, participants were asked to read a small number of cards (usually two or three), and show on a VAS where they felt each particular state would be placed. The purpose of this was to observe whether the participant fully understood the descriptions on the cards, and that 'better' states were placed higher on the scale than 'worse' states.

Again, six questions were asked to each participant, and the comparator changes are shown in Table 2. Six indifference curves will be produced: the first showing the improvements in questions 2A, 2B, 2C and the comparator, the second represented by questions 2D, 2E and their comparator, and the third by question 2F and its comparator. The fourth, fifth and sixth curves will represent those elicited in the questions involving health losses.

Table 2: Comparisons used in Experiment II

Health gains	Comparator Improvement		Equivalent Improvement	
	Initial Utility	**Final Utility**	**Initial Utility**	**Final Utility**
G2A	0.76	1.00	0.52	?
G2B	0.76	1.00	0.28	?
G2C	0.76	1.00	0.00	?
G2D	0.52	1.00	0.28	?
G2E	0.52	1.00	0.00	?
G2F	0.28	1.00	0.00	?
Health losses	**Comparator Reduction**		**Equivalent Reduction**	
	Initial Utility	**Final Utility**	**Initial Utility**	**Final Utility**
L2A	1.00	0.76	?	0.52
L2B	1.00	0.76	?	0.28
L2C	1.00	0.76	?	0.00
L2D	1.00	0.52	?	0.28
L2E	1.00	0.52	?	0.00
L2F	1.00	0.28	?	0.00

3. Results

Experiment I.

The responses of each participant were analysed for differences using the Wilcoxon Signed Rank Test. The results of Experiment I are presented in Tables 3 (direct comparisons) and 4 (indirect). It is clear that for all twelve direct comparisons, respondents have indicated a greater weight towards 'equal' improvements with lower initial states. In other words, smaller improvements are required from lower initial states to make them 'just as good' as larger improvements from higher initial states. It is also possible to compare the responses to questions 1A, 1B and 1C, since all are implied to be of equal value. This is also true for questions 1D and 1E. In these eight indirect comparisons, respondents again prefer the improvement with the lower initial state. All findings are significant at a 5% level or less. For illustrative purposes, the

implied indifference curves using the mean responses are shown in Figure 3, where a clear 'fanning out' effect is apparent.

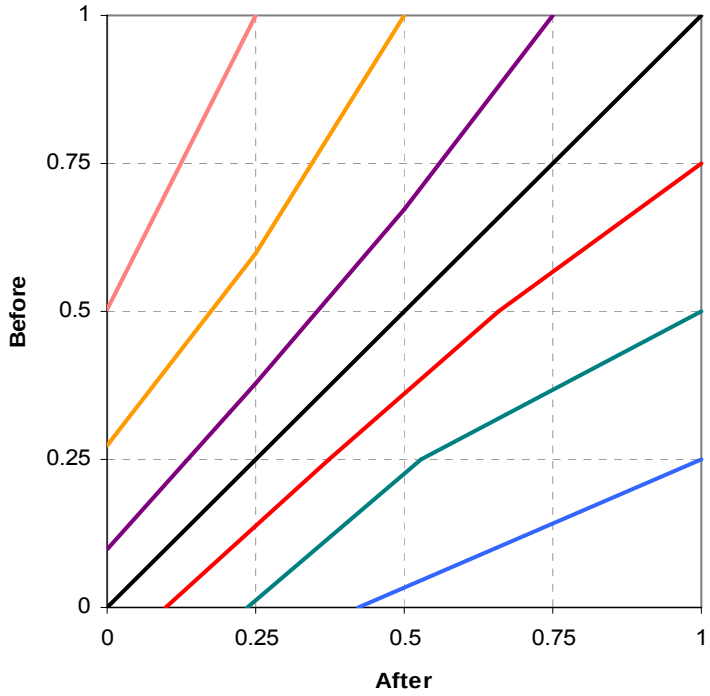
Table 3: Experiment I Results – Direct Comparison

Question	Comparator	Initial Utility	Mean Final Utility	Higher Value for Change With Higher Initial Utility	Higher Value for Change With Lower Initial Utility	T-Statistic	Critical Value	Significant (α<0.05)	P-Value
G1A	0.75→1.00	0.50	0.658	1 (2.9%)	30 (85.7%)	2.5	163	Yes	7.90E-07
G1B	0.75→1.00	0.25	0.374	0 (0.0%)	34 (97.1%)	0	199	Yes	1.91E-07
G1C	0.75→1.00	0.00	0.099	1 (2.9%)	34 (97.1%)	7	211	Yes	2.37E-07
G1D	0.50→1.00	0.25	0.528	0 (0.0%)	34 (97.1%)	0	199	Yes	1.91E-07
G1E	0.50→1.00	0.00	0.236	0 (0.0%)	35 (100%)	0	211	Yes	1.30E-07
G1F	0.25→1.00	0.00	0.423	0 (0.0%)	35 (100%)	0	211	Yes	1.30E-07
L1A	1.00→0.75	0.671	0.50	3 (8.6%)	27 (77.1%)	54	151	Yes	2.51E-05
L1B	1.00→0.75	0.379	0.25	2 (5.7%)	29 (82.9%)	20.5	163	Yes	9.00E-06
L1C	1.00→0.75	0.098	0.00	1 (2.9%)	32 (91.4%)	9	187	Yes	1.00E-06
L1D	1.00→0.50	0.599	0.25	3 (8.6%)	27 (77.1%)	12.5	151	Yes	6.00E-06
L1E	1.00→0.50	0.274	0.00	1 (2.9%)	31 (88.6%)	18.5	175	Yes	5.00E-06
L1F	1.00→0.25	0.505	0.00	1 (2.9%)	29 (82.9%)	8	151	Yes	4.00E-06

Table 4: Experiment I Results – Indirect Comparisons

Comparison	Improvement in First Question	Improvement in Second Question	Higher Value for Improvement With Higher Initial Utility	Higher Value for Improvement With Lower Initial Utility	T-Statistic	Critical Value	Significant (α<0.05)	P-Value
G1A v G1B	0.158	0.124	10 (28.6%)	22 (62.9%)	137.5	175	Yes	0.009
G1A v G1C	0.158	0.099	5 (14.3%)	28 (80.0%)	87	187	Yes	0.0003
G1B v G1C	0.124	0.099	6 (17.1%)	19 (54.3%)	96.5	100	Yes	0.036
G1D v G1E	0.278	0.236	8 (22.9%)	20 (57.1%)	94.5	130	Yes	0.007
L1A v L1B	0.171	0.129	9 (25.7%)	19 (54.3%)	95	130	Yes	0.014
L1B v L1C	0.129	0.098	3 (8.6%)	20 (57.1%)	40	83	Yes	0.003
L1A v L1C	0.171	0.098	4 (11.4%)	22 (62.9%)	33	110	Yes	0.0003
L1D v L1E	0.394	0.274	7 (20%)	20 (57.1%)	60	119	Yes	0.002

Figure 3: Experiment I Health Improvement Matrix



Experiment II.

Tables 5 and 6 show the results of Experiment II. Of the direct comparisons, eight of the twelve questions yielded statistically significant preferences for the health change with the lower initial state (five of the six concerning health gains). Only one comparison did not produce a significant finding. Of the eight possible indirect comparisons, six show significant indications of a greater weight for the changes with the lower initial state, whilst two show no significant preference. The indifference curves implied by the mean responses are shown in Figure 4. Again, a 'fanning out' effect is evident.

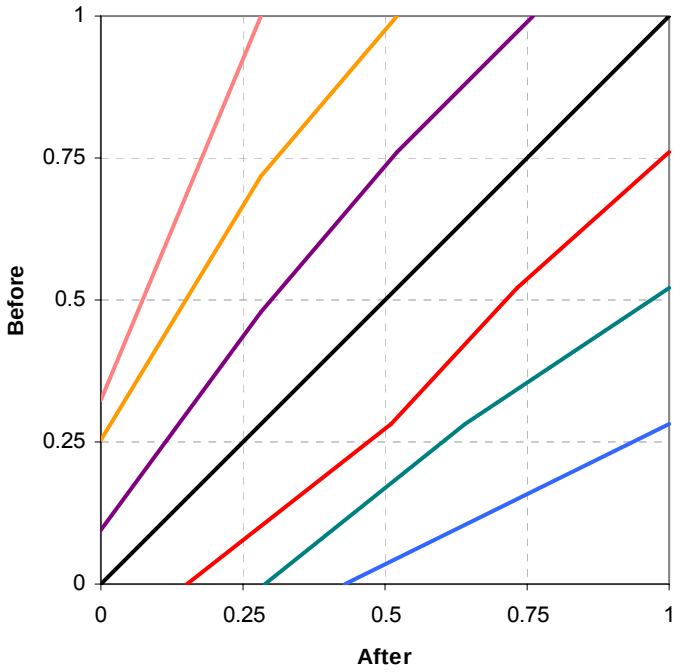
Table 5: Experiment II Results – Direct Comparisons

Question	Comparator	Initial Utility	Mean Final Utility	Higher Value for Improvement With Higher Initial Utility	Higher Value for Improvement With Lower Initial Utility	T-Statistic	Critical Value	Significant (α<0.05)	P-Value
G2A	0.76→1.00	0.52	0.732	8 (22.9%)	22 (62.9%)	131.5	151	Yes	0.019
G2B	0.76→1.00	0.28	0.509	11 (31.4%)	22 (62.9%)	225.5	187	No	0.165
G2C	0.76→1.00	0.00	0.153	6 (17.1%)	29 (82.9%)	77	211	Yes	5.0E-05
G2D	0.52→1.00	0.28	0.641	4 (11.4%)	30 (85.7%)	44.5	199	Yes	7.9E-06
G2E	0.52→1.00	0.00	0.287	6 (17.1%)	29 (82.9%)	48	211	Yes	6.4E-06
G2F	0.28→1.00	0.00	0.431	3 (8.6%)	32 (91.4%)	8	211	Yes	2.6E-07
L2A	1.00→0.76	0.761	0.52	21 (60%)	14 (40%)	311	214	No	0.954
L2B	1.00→0.76	0.479	0.28	13 (37.1%)	21 (60%)	210.5	201	No	0.139
L2C	1.00→0.76	0.094	0.00	3 (8.6%)	32 (91.4%)	32	214	Yes	4.00E-06
L2D	1.00→0.52	0.720	0.28	18 (51.4%)	16 (45.7%)	258	201	No	0.505
L2E	1.00→0.52	0.253	0.00	5 (14.3%)	30 (85.7%)	49	214	Yes	1.40E-05
L2F	1.00→0.28	0.323	0.00	7 (20%)	28 (80%)	30	214	Yes	3.00E-06

Table 6: Experiment II Results – Indirect Comparisons

Comparison	Improvement in First Question	Improvement in Second Question	Higher Value for Improvement With Higher Initial Utility	Higher Value for Improvement With Lower Initial Utility	T-Statistic	Critical Value	Significant (α<0.05)	P-Value
G2A v G2B	0.212	0.229	17 (48.6%)	13 (37.1%)	268.5	151	No	0.233
G2A v G2C	0.212	0.153	11 (31.4%)	21 (60.0%)	127.5	175	Yes	0.005
G2B v G2C	0.229	0.153	7 (20.0%)	25 (71.4%)	87.5	175	Yes	0.0004
G2D v G2E	0.361	0.287	12 (34.3%)	22 (62.9%)	157.5	199	Yes	0.009
L2A v L2B	0.24	0.199	13 (37.1%)	18 (51.4%)	175.5	163	No	0.158
L2B v L2C	0.199	0.094	4 (11.4%)	19 (54.3%)	33	83	Yes	0.0014
L2A v L2C	0.24	0.094	2 (5.7%)	29 (82.9%)	20	163	Yes	<0.0001
L2D v L2E	0.44	0.253	6 (17.1%)	27 (77.1%)	66.5	187	Yes	0.00013

Figure 4: Experiment II Health Improvement Matrix



4. Application

It is possible to use this technique in order to elicit unit valuations for an individual's preferences towards health improvements. Assuming a 'full improvement' from 0 to 1 is worth 1 unit of benefit to the individual, using the following methodology, where improvements between states are denoted by '→', can determine the benefit measure of an improvement.

First, determine the state α, where:

$$0 \rightarrow \alpha = \alpha \rightarrow 1. \qquad (1)$$

We can assume that:

$$0 \rightarrow 1 \equiv 0 \rightarrow \alpha + \alpha \rightarrow 1 \qquad (2)$$

If we assign a unit of 1 to the improvement 0→1, then from (1) and (2) we can show that:

$$0 \rightarrow \alpha = \alpha \rightarrow 1 = ½ \qquad (3)$$

Thus, any improvements from 0 to α, or α to 1 can be valued at 0.5 units. This valuation will also apply to any other health improvements lying on the same indifference curve as $0 \rightarrow \alpha$. The method can be repeated, determining the state β, where:

$$0 \rightarrow \beta = \beta \rightarrow \alpha \qquad (4)$$

Since:

$$0 \rightarrow \alpha \equiv 0 \rightarrow \beta + \beta \rightarrow \alpha \qquad (5)$$

We can see from (3), (4) and (5) that:

$$0 \rightarrow \beta = \beta \rightarrow \alpha = ¼ \qquad (6)$$

Again, this valuation will apply to all improvements upon that indifference curve. This process is not restricted to simply valuing those states equal to ½, ¼, ⅛ and so forth. For example, specific improvements can be summed together, and this method can be repeated indefinitely until all health improvements have been assigned a value. Furthermore, if it is found that responses conform to a predictable pattern, it will be possible to elicit a small number of valuations, and simply calculate the remaining figures.

5. Discussion

It is important to remember that the methods proposed here do not undermine existing methods of health state utility elicitation such as the standard gamble, time trade-off, person trade-off or VAS. These methods can be used to determine the individual valuations of the initial and final states of health as the first stage in a two-stage process, the second stage being the adjustment for health improvement preferences.

The dimension of time can be incorporated into the analysis in the same way that it is used in normal QALY estimation. If a health improvement is estimated to be worth say, 0.5 units of utility, then the benefit would be 0.5 multiplied by the length of time expected in that state. If more than one change is predicted, this can also be easily allowed for.

The initial measurement of a health improvement matrix would be an empirically daunting prospect. However, it can be considered to be a one-off task. Once the indifference curves and utility ratings have been assigned, any health improvement can be valued relatively easily. Any new utility valuations for specific health states can be incorporated by simply using the new utility value for either the 'before' or 'after' utility.

The analysis has assumed that participants in Experiment II 'agreed' with the existing HRQL utility valuations associated with those descriptions. It would be possible to avoid this by initially asking each patient to explicitly state a valuation for each description and thereafter use those values. It was felt in this study however, that this procedure would increase the likelihood of other biases distorting the results. In any case, current policy making techniques also assume that patients 'agree' with the present valuations of health states.

It should be emphasised once more that this study has explicitly studied preferences at an individual level, rather than preferences for the allocation of health utility. It is possible to utilise the valuations elicited by this method in conjunction with existing strategies of healthcare allocation.

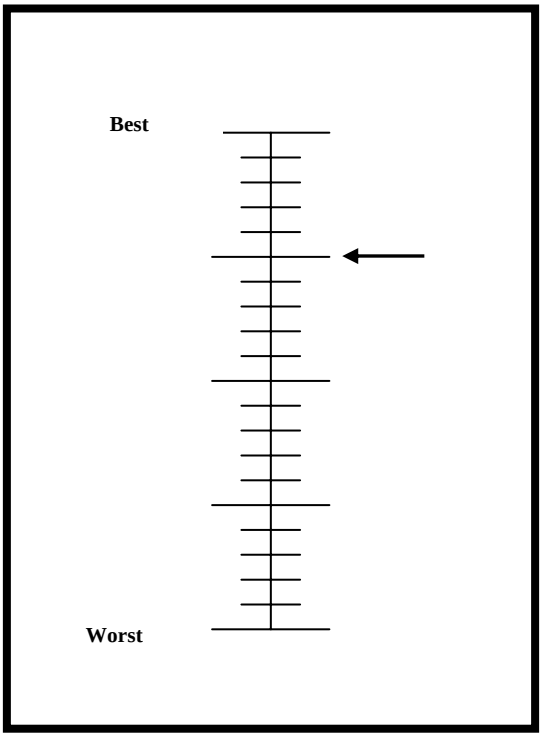
6. Conclusion

Significant evidence is presented here towards the argument that individuals value 'equal' health changes more if they occur from a lower initial state. This effect is especially apparent when the initial HRQL utility approaches zero. The results are consistent between two different methods of presenting health state utilities to individuals. Methods are also presented, proposing a technique to place values on these preferences and incorporate these adjusted values into normal cost-utility analysis.

Acknowledgements

The authors are grateful for comments made in the early stages of this study by Peter Ubel (Director of the Program for Improving Health Care Decisions, University of Michigan). The monetary funding provided for participant payment incentives was provided from the University of Newcastle upon Tyne Business School.

Appendix A: Example of VAS presentation



Appendix B: Example of HUI:3 presentation

Imagine that you are experiencing the following symptoms:	
SIGHT	Able to see well enough to read ordinary newsprint and recognize a friend on the other side of the street, without glasses or contact lenses.
Hearing	Unable to hear at all.
Speech	Able to be understood completely when speaking with strangers or friends.
Mobility	Able to walk around the neighbourhood without difficulty, and without walking equipment.
Dexterity	Full use of two hands and ten fingers.
Emotion	Somewhat unhappy.
Cognition	Somewhat forgetful, but able to think clearly and solve day to day problems.
Pain	Mild to moderate pain that prevents no activities.

References

1. Gold MR, Siegel JE, Russell LB, Weinstein MC. *Cost-effectiveness in health and medicine.* Edition Oxford University Press, New York, 1996.

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

3. Ubel P, Richardson J, Pinto-Prades J-L. Life saving treatments and disabilities: are all QALYs created equal? *International Journal of Technology Assessment and Health Care*. 1999; **15** 738-748.

4. Nord E. *The Person Trade-Off Approach to Valuing Health Care Programs*. Centre for Health Program Evaluation, Working Paper 38. 1994.

5. Dolan P. The Measurement of Individual Utility and Social Welfare. *Journal of Health Economics*. 1998; **17** 39-52.

6. Nord E, Pinto-Prades J-L, Richardson J. Incorporating Societal Concerns for Fairness in Numerical Valuations of Health Programmes. *Health Economics*. 1999; **8** 25-39.

7. Ubel P, Richardson J, Menzel P. Societal Value, the Person Trade-Off, and the Dilemma of Whose Values to Measure for Cost-Effectiveness Analysis. *Health Economics*. 2000; **9** 127-136.

8. Feeny D, Furlong W, Boyle M, et al. Multi-Attribute Health Status Classification Systems: Health Utilities Index. *Pharmacoeconomics*. 1995; **7** 490-502.

9. Health Utilities Group. *Health Utilities Index*. http://www.fhs.mcmaster.ca/hug/. Date Accessed: 26 July 2006.