Deriving QALY weights through discrete choice experiments: challenges and preliminary results

Rachel Baker, Cam Donaldson, Emily Lancsar, Mandy Ryan and John Wildman

Paper presented to Health Economists' Study Group
Norwich, 9 – 11 January 2007

On behalf of the Social Value of a QALY research team:

Rachel Baker[1,] Ian Bateman[2], Cam Donaldson[1], Michael Jones-Lee[1], Emily Lancsar[1,] Graham Loomes[2], Helen Mason[1], Maria Odejar[3] Angela Robinson[2], Mandy Ryan[3], Phil Shackley[4], Richard Smith[5], Kerry Sproston[6], Robert Sugden[2], Heather Wardle[6], John Wildman[1]

Newcastle University[1], University of East Anglia[2], University of Aberdeen[3], University of Sheffield[4], London School of Hygiene and Tropical Medicine[5],and National Centre for Social Research (NatCen)[6].

Corresponding author:
Rachel Baker
Institute of Health and Society
Newcastle University
21 Claremont Place
Newcastle upon Tyne
NE2 4AA (r.m.baker@ncl.ac.uk)

1. Introduction

Over the past 20 years, the quality adjusted life year (QALY) has become the dominant measure of benefit assessment in health economic evaluation (Torrance, 1986; Williams, 1985). Its use is now widespread, particularly in the various health technology assessment agencies around the globe, and most notably in the UK through the assessment procedures undertaken by the National Institute for Health and Clinical Excellence (NICE) (NICE, 2007). Nevertheless, almost since the QALY was introduced to the health economics literature, the importance of the context in which health gains are produced has been discussed, raising the question of whether 'a QALY is a QALY is a QALY' (Donaldson et al., 1988ab; Weinstein 1988). Williams (1996) has said: "there is nothing in the QALY approach that requires QALYs to be used in a maximising context… …more complex rules will almost certainly be needed if collective priority-setting is to reflect the views of the general public". This issue has come to the fore again recently through NICE. A prescribed task of NICE is to assess health interventions in terms of their health gains relative to their costs, and to make recommendations as to whether or not an intervention should be adopted by the rest of the NHS in England. However, it has been recognised that the Appraisal Committee at NICE will take characteristics of beneficiaries of such interventions into account in its deliberations (Rawlins and Culyer, 2006). This raises an important policy question as to whether quantitative estimates, reflecting the relative weight to be attached to health gains derived by different beneficiaries, can be derived from a survey of the general public and thus be used to assist the NICE process.

The Social Value of a QALY (SVQ) Team was contracted from October 2004 to January 2008 to undertake two studies, each based on a survey of the population in England. The first is the 'valuation study', with the aim of exploring the feasibility of estimating a baseline monetary value for a QALY. This is the subject of another paper which will be discussed in this session. The second is the 'relativities study', part of which we report in this paper.

The aims of the relativities study were:

- to identify characteristics of beneficiaries of health care over which relative weights were to be derived;
- to estimate the relative weights to be attached to health gains according to the characteristics of the recipients of these gains.

We adopted two methodological approaches to the estimation of relative weights (which nevertheless share some commonalities) and respondents to the relativities survey answered both matching (or person trade-off) questions and discrete choice questions. In this paper we report a preliminary analysis of the latter[1].

This rest of this paper is organised in six further sections as follows. We begin with a brief review of the literature on QALY weights, highlighting articles of particular interest, and concluding with challenges of this project. In Section 3 we describe the methods and results of exploratory and developmental work to identify attributes and the development of a diagrammatic approach to the presentation of survey questions. The design of a set of 286 discrete choice, pairwise questions is outlined in Section 4 together with accounts of the particular constraints around DCE design in the context of the selected attributes and our analytic strategy. In section 5, the results of an econometric analysis of the choice data are presented followed by section 6 in which two approaches to the estimation of relative weights are applied. The paper concludes with a brief discussion.

---

[1] An analysis of the matching data was presented at iHEA, Copenhagen 2007.

2. Weighting QALYs: literature and challenges

Several authors have discussed the theoretical, ethical and practical issues around distributional weights for QALYs (e.g. Nord et al., 1999; Dolan and Olsen, 2002) and there have been a number of attempts to estimate weights (e.g. Cropper et al., 1994; Johannesson and Johansson, 1996; Nord et al., 1996; Johannesson and Johansson, 1997).

Dolan et al. (2005) provide the most comprehensive review in this field to date. Using a 'citation pearl growing' search strategy they identify 78 papers, dated 2001 or earlier, 64 of which include empirical data. There is growing evidence from (mainly survey based) studies of the general population that the number of QALYs gained is likely to be traded off against *other factors*. Efforts to identify these other factors have indicated a wide range of possibilities but there remain inconsistencies and contradictory findings. The list of factors identified by Dolan et al. (drawn from both the empirical and theoretical literature) are as follows: age, severity of illness (or starting point health state), end point/final health state, culpability/responsibility for ill health, having dependents, socio-economic characteristics, gender, ethnicity, inequalities in health, and the concentration or dispersion of the distribution of a fixed health gain. The reviewers conclude that despite a growing body of literature, there are contradictory findings, many studies involve small samples and few attempt to estimate weights. Earlier, Schwappach (2002) described two categories of factors that could influence social value: i) characteristics of beneficiaries and ii) characteristics of the intervention's effect on patients' health, adding the factors of prior health consumption of patients the duration of benefit and whether the gain in health is an improvement or the prevention of a decline.

In the period since these reviews, there have been further contributions to the literature. Two of these used discrete choice data, examining attributes such as age, culpability (e.g. related to alcohol consumption), expected length of survival, time on waiting list and whether a previous transplant had been received (Ratcliffe, 2000) and lifestyle, socioeconomic status, age, life expectancy, quality of life after treatment, and level of past use of health care (Schwappach, 2003). The results indicate that several factors in addition to health gain influence people's choices. However, unlike the study described below, these studies were either condition specific or not based on a population sample of respondents.

Projects attempting to derive QALY weights are faced with three significant challenges: identifying characteristics of beneficiaries over which weights *should* be derived; designing and presenting questions so that respondents can understand complexities and make choices; elicitation of quantitative preference data from members of the general public to allow the estimation of QALY weights.

Designing questionnaires that respondents can engage with is a particular challenge in this study because our aim was to estimate the relative value of different types of de-contextualised, generic QALY gains. Without context, however, questions can seem overly abstract to respondents. There is also evidence that different 'types' of de-contextualised QALY (i.e. life saving or quality of life enhancing QALYs) will be regarded differently (see Johri et al., 2005; Mason 2007) and so the presentation of questions needed to be flexible enough to incorporate different QALY types.

3. Identification and presentation of attributes for QALY weights

Qualitative techniques are increasingly used to establish appropriate attributes for DCEs (Coast and Horrocks, 2007). We have taken a predominantly qualitative approach supplementing conventional qualitative methods with other techniques. This exploratory phase of the project was an iterative process involving several phases and a range of methods including: open ended discussion in focus groups; simple ranking procedures; experimentation with sample questions; and a more complex ranking task involving card-sorting (Baker et al., 2006). Qualitative findings were interpreted alongside the results of the other methods used. Due to the focus in this paper on more quantitative aspects (especially estimation of weights), there is insufficient space in this paper to report all of these techniques in detail, but papers on this are available from the authors.

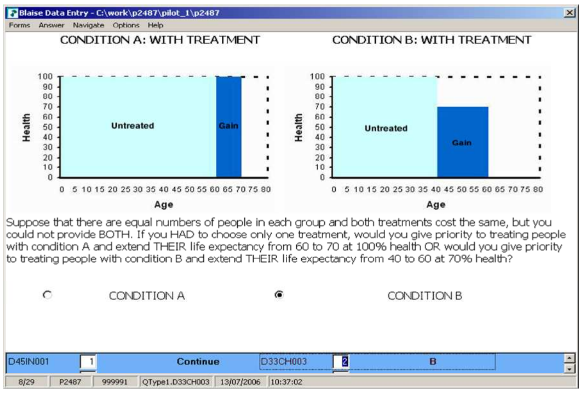
This exploratory work, which lasted for over a year, taken together with considerations of policy relevance in consultation with representatives of NICE, resulted in the selection of the following attributes: age (both at onset of illness and age at death) and severity of illness (with and without treatment). These attributes, and levels presented for each, are listed in Table 1.

Table 1: Attributes and levels

Description of attribute	**Levels**
Age at onset (years)	1 10 20 40 60 70
Age at death if untreated (years)	1 10 20 40 60 70 80
Gain in life expectancy (years)	0 1 5 10 20 40 60 79
Qol if untreated (represented as %)	0 30 60 90
Gain in Qol with treatment (%)	0 10 20 40 70 100

The presentation of the discrete choice questions, including a detailed introductory explanation, was also developed iteratively in focus groups. By far the most successful method was presentation of concepts of health, age, and health gains using diagrams. These diagrams were first explained by building them in small steps for respondents using an animated Powerpoint presentation (excerpts of which are reproduced in Appendix 1) which was ultimately incorporated into a computer assisted personal interview (CAPI). The diagrams were then presented as choice questions, an example of which is shown in Figure 1.

Figure 1: Diagrammatic choice question



Suppose that there are equal numbers of people in each group and both treatments cost the same, but you could not provide BOTH. If you HAD to choose only one treatment, would you give priority to treating people with condition A and extend THEIR life expectancy from 60 to 70 at 100% health OR would you give priority to treating people with condition B and extend THEIR life expectancy from 40 to 60 at 70% health?

○ CONDITION A ● CONDITION B

D45IN001 1 Continue D33CH003 2 B

The entire questionnaire instrument was then piloted using cognitive interviews with 42 respondents and was well received. Respondents followed the introduction and understood the questions. As a result of these, only minor adjustments to the wording of and number of questions in each version of the questionnaire were needed.

The discrete choice questions were part of a longer questionnaire (incorporating matching questions, attitudinal questions and socio-demographics) which was administered face to face using a CAPI. The survey was administered to a nationally-representative sample (of 587) of the population in England by the National Centre for Social Research (NatCen) during February-April 2007. The sample was generated by NatCen from the population of adults (18 and over) living in England. 30 addresses were selected from each of 40 postcode areas, which were stratified by Government Office Region (GOR 9 regions) and the proportion of manual/non manual households. Within each household, only one adult was eligible for inclusion in the study. In households with more than one eligible adult present, interviewers randomly selected one interviewee. 243 (41%) were male, the mean age being 52 years.

4. DCE design and analytic strategy

Basic design

A full factorial design using the attributes and levels listed in Table 1 would have resulted in 6(7)(8)(4)(6)= 8064 possible profiles but there are many combinations of levels on these attributes that result in implausible scenarios. For example, the age at death attribute can only take levels that are ‘older’ than the level of the attribute ‘age at onset’ of illness, the gain in life expectancy added to the age at death if untreated cannot exceed some reasonable maximum age, which was constrained to be 80 years. The full list of constraints is listed in Appendix 2. After imposing these constraints, 6572 of the possible 8064 profiles were implausible, leaving 1492 profiles (19% of the total).

Perhaps because of the number of constraints needed to exclude implausible pairings, attempts at experimental designs using software (SAS) resulted in over 200 profiles describing ‘age at onset’ as 1 year olds and only 12 profiles describing 70 year olds. At this stage the design was altered manually to improve the balance of questions about different age groups. Clearly, such a severely reduced set of available profiles, together with the manual alteration of the design to achieve greater balance, negatively affect the design properties which are desirable in DCEs. Before administering the questionnaire, data were simulated and used to ensure that a model could be estimated on the basis of the amended design. Nevertheless, none of the main effects or two way interactions are independent of each other, the efficiency of the main effects only model being 49%.

The amended design consisted of 286 pairwise choices presented in 36 versions of the questionnaire. Each respondent answered 8 choice questions as part of a questionnaire that also contained 6 matching questions, 4 attitudinal questions and a series of socio-demographic questions.

Functional form and empirical approach

If the standard QALY model is true, utility would simply be a function of QALYs. If, however, individuals are concerned about other characteristics then these will also be part of the utility function. In this simple case we assume that utility is a function of age of onset (AO), age at death without treatment (AD), quality of life lost without treatment (QL) and QALYs gained from treatmeant ($QALY$).This gives:

$$U=f(AO, AD, QL, QALY), \qquad [1]$$

The first two terms on the right hand side detect age effects, QL detects severity and $QALY$ is the health gain.

If we assume a multiplicative underlying model we may use a log-linear model of the form:

$$\log(U) = \beta_1 \log(AO) + \beta_2 \log(AD) + \beta_3 \log(QL) + \beta_4 \log(QALY) \qquad [2]$$

This is a standard log-linear utility function where the βs are parameters to be estimated. Given that response data to a DCE are modelled within a random utility framework and that DCEs involve choices over alternative combinations of the dependent variables, then, assuming that the β parameters are identical across all choices, a simple model of the following form can be estimated:

$$\begin{aligned}\Delta\log(U) = {} & \beta_1(\Delta\log(AO)) + \beta_2(\Delta\log(AD)) + \beta_3(\Delta\log(QL)) \qquad [3] \\ & + \beta_4(\Delta\log(QALY)) + \varepsilon\end{aligned}$$

where Δ represents the differences in levels of any given attribute reflected in the pairwise choices presented.

We may also want to take account of higher order nonlinearities. The log transformation introduces nonlinearity into the functional form, but it is still monotonic. In order to allow further nonlinearities we also estimate the models using squared and cubic terms.

5. Results

587 respondents yielded 4696 usable responses to the discrete choice questions.

The estimated models are shown in Table 2. The model with the squared and cubed terms is preferred according to the Akakie information criterion.

Table 2: Simple and powered models

	Coefficients	Z	Coefficients	z
log AO	-0.02	-1.03	-0.31	-1.18
log AO^2			0.24	1.44
log AO^3			-0.04	-1.61
log AD	-0.07	-2.12	1.28	4.06
log AD^2			-0.76	-4.34
log AD^3			0.11	4.22
log QL	-0.14	-3.80	-0.64	-1.72
log QL^2			-0.43	-0.88
log QL^3			-0.09	-0.58
log QALY	0.75	22.37	0.45	8.32
log QALY^2			-0.03	-1.18
log QALY^3			0.03	4.88

Examining the coefficients for the simple model we see that increasing age at onset reduces the probability of choice, as does increasing the age at death. This suggests that the young are preferred to the old. The coefficient on the quality of life lost is also negative suggesting that as the health state is more severe the respondents are less likely to choose that group. Finally the QALY is positive, as we would expect. The shapes of the functional forms for the simple model are given in Figures 2-5, with those for the powered models in Figures 6-9.

Figure 2: **Age at onset vs utility**

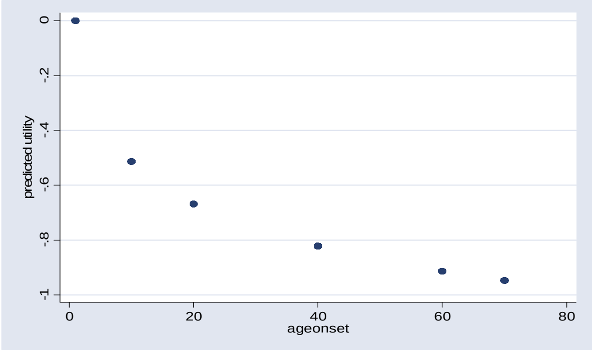


Figure 3: **Age at death vs utility**

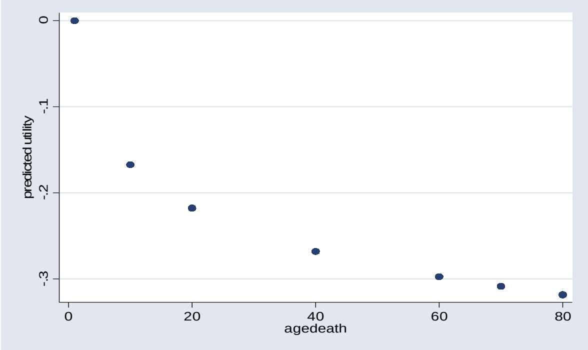


Figure 4: **Quality of life lost vs utility**

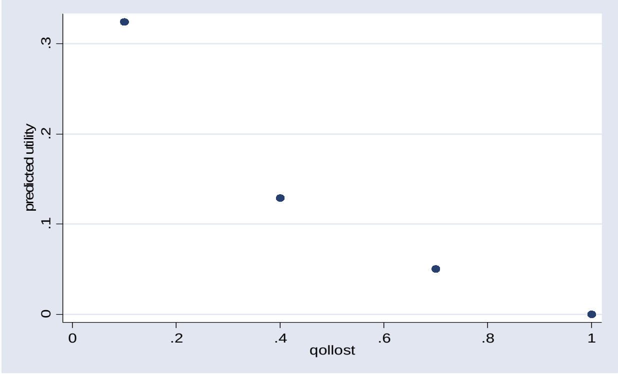


Figure 5: **QALY vs utility**

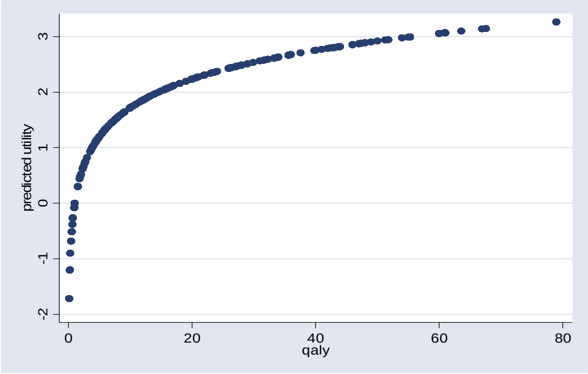


Figure 6: **Age at onset vs utility (powered functions)**

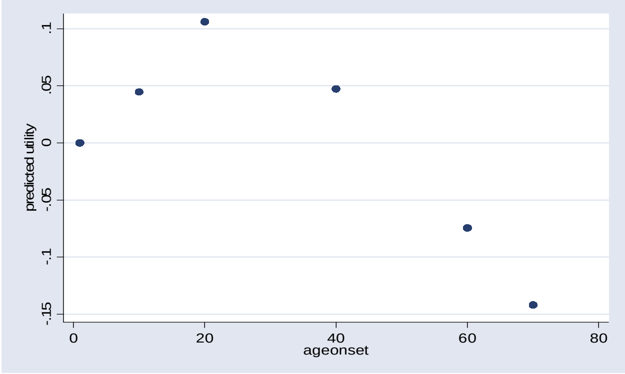


Figure 7: **Age at death vs utility (powered functions)**

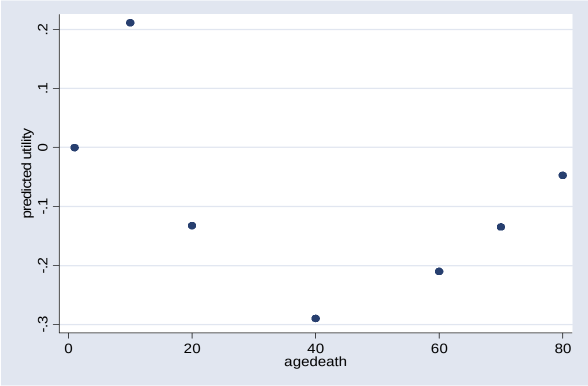


Figure 8: **Quality of life lost vs utility (powered models)**

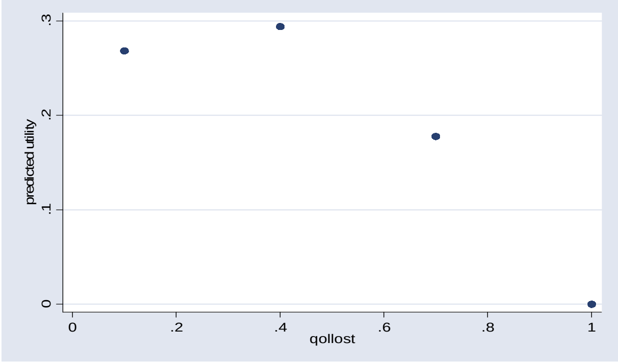


Figure 9: **QALY vs utility (powered model)**

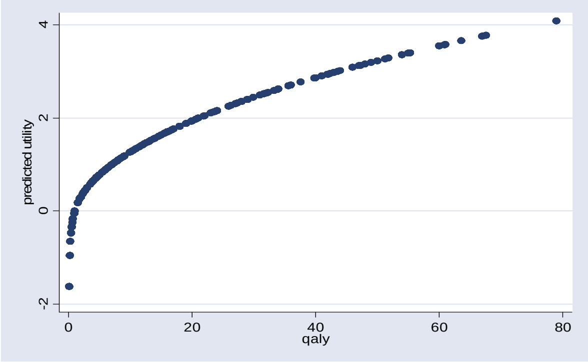


Figure 6 shows that as age at onset increases the predicted utility increases but then falls away. Individuals who are around 10 to 40 are clearly preferred to the very young and the very old. Figure 7 indicates that the age of death has a rather complicated functional form. There is a clear preference to save those who die young and a slight preference to save those who will die old rather than the middle aged. Figure 8 demonstrates a preference for individuals with lower severity, with the maximum at 0.4. The predicted utility then slopes downwards quite sharply. This shows that there is no preference for purely life saving interventions. Figure 9 shows that increasing QALYs are always preferred. Concavity is still observed, but less so than in the simple model. The increase in predicted utility is very large for the very small QALY gains.

6. Weights

Basic approach

We estimate weights for both models from Table 2. The weights are estimated using the estimated probability of choice, using a base case scenario, as our reference case, of AO = 40, AD = 60, QL = 0.7, and QALY = 4 (i.e. individuals fall ill at 40, lose 0.7 of their quality of life, will die at 60 without treatment, and are then given 4 QALYs with treatment). The way in which the 4 QALYs have been allocated is unspecified in this model. If the QALY model were true then the composition of the QALY should be irrelevant. It may however be the case that the gain in utility from gains in life years may be greater (or less) than gains in utility from quality of life gains. We will return to this issue later.

The base case is compared to an alternative scenario. To illustrate, consider the following: if AO = 1, AD = 1, and QL = 1 what is the number of QALYs required of the comparison group to reach a point of indifference between these health states (or the probability of choosing the base case is 0.5)? At this point we take the ratio of QALYs in the base case compared to the comparator group, to derive the weight, so:

$$weight = \frac{QALY_{COMPARISON}}{QALY_{BASECASE}}$$

Weights greater than 1 favour the base case, weights less than 1 favour the comparison and weights equal to 1 show indifference between the two scenarios. Weights for a number of scenarios based on the simple model are presented in Table 3.

There are a number of patterns we can observe. For those scenarios (1, 6, 11 and 14) with life expectancy gains only (age at onset equals age at death and quality of life lost equals 1), the weights are in favour of the comparator (the life saving option) but the weights, compared to the base case, move towards 1 as the comparison group gets older, until the weight is greater than 1 (favouring the base case).

As age of onset increases, the weights move towards 1. For example, looking at comparisons 4 and 7, we see that for the same age of death and the same severity the weights move from 0.8125 to 0.86 as age of onset increases from 1 to 10.

Table 3: Weights based on simple model

	Age onset	Age death	Qollost	Qaly gained	weight
Base	40	60	0.7	4	
Comparisons 1	1	1	1	2.55	0.64
2	1	10	0.7	3	0.75
3	1	10	0.1	2.1	0.525
4	1	20	0.7	3.25	0.8125
5	1	20	0.1	2.23	0.56
6	10	10	1	3.69	0.92
7	10	20	0.7	3.45	0.86
8	10	20	0.1	2.38	0.6
9	10	40	0.7	3.67	0.92
10	10	40	0.1	2.55	0.64
11	40	40	1	4.1	1.025
12	40	60	0.7	4	1
13	40	60	0.1	2.77	0.69
14	70	70	1	4.4	1.1
15	70	80	0.7	4.2	1.05
16	70	80	0.1	2.9	0.725

The weights for severity move to zero as severity reduces. So as the amount of quality of life lost falls from 0.7 to 0.1 the weight for the comparison group gets smaller, suggesting that it is weighted more favourably.

Another way to consider these weights is to look at scenarios where severity is fixed at a certain level and we allow the age at onset and age at death to vary. Table 4 fixes QL at 0.7 and then varies the age onset and age at death. The base case is as before. This clearly demonstrates that the weights move towards 1 (favouring the base case) as the age of onset increases, or the age of death increases.

Table 4: Weighting by age of onset and age at death

AO	*1*	*10*	*20*	*40*	*60*	*70*
AD						
1						
10	0.75					
20	0.8	0.86				
40	0.86	0.92	0.94			
60	0.89	0.96	0.975	1		
70	0.91	0.975	1	1.01	1.03	
80	0.92	1	1.01	1.03	1.04	1.05

Alternatively, we can fix age at death and allow quality of life lost and age at onset to vary, as in Table 5. We can see that the weights move closer to 1 as the age at onset increases (as was shown above). The weights also move towards 1 as the severity increases.

Table 5: Severity vs age at onset

AO	1	10	20	40	60
QL					
0.1	0.62	0.66	0.68	0.69	
0.4	0.805	0.86	0.88	0.9	
0.7	0.89	0.96	0.98	1	
1					1.09

Overall, these weights demonstrate that younger individuals are preferred to older individuals, and those with less severity, as measured by quality of life lost, are preferred to the more severe. However, as we noted above, the log transformation is monotonic. The weights were recalculated with the higher order terms allowing more nonlinearities in the weights. The results are given in Table 6.

Table 6: Weights based on powered model

	Age onset	Age death	Qollost	Qaly gained	weight
Base	40	60	0.7	4	
Comparison 1	1	1	1	4.1	1.025
2	1	10	0.7	1.85	0.463
3	1	10	0.1	1.52	0.38
4	1	20	0.7	3.8	0.95
5	1	20	0.1	3.16	0.79
6	10	10	1	2.48	0.62
7	10	20	0.7	3.5	0.875
8	10	20	0.1	2.9	0.725
9	10	40	0.7	4.6	1.15
10	10	40	0.1	3.95	0.9875
11	40	40	1	6.15	1.54
12	40	60	0.7	4	1
13	40	60	0.1	3.38	0.845
14	70	70	1	6.45	1.61
15	70	80	0.7	4.2	1.05
16	70	80	0.1	3.55	0.89

The nonlinearities have a large impact on the estimated weights. For those who are receiving only life expectancy gains, instead of the weights approaching 1 as the individual gets older the weights start greater than 1, fall to 0.62 for 10 year olds and then rise well above 1 for 70 older groups. In this case the very young and the very old are weighted lower than the ages in between.

Focussing on the comparison groups 4 and 7, we can see that the weight falls from 0.95 to 0.875 as the age at onset increases from 1 to 10. For age at death increases we focus on comparisons 2 and 4, and 7 and 9. As age of death increases from 10 to 20 the weight changes from 0.463, strongly favouring the comparison group, to 0.95 which means it is weighted very similarly to the base case. We see similar movements when age at death increase from 20 to 40 in comparisons 7 and 9.

As above, the increase in severity means that the weight moves towards zero. However, Table 7 clearly shows that as age of onset increases the weights move towards 0 up to age 20 but then move towards 1. Age at death moves towards 1 up to age 40 but then falls again. This shows that there is a clear preference for treating people who fall ill at ages between 10 and 40 rather than the very old or the very young and a preference for treating those who will die either very young or very old.

Table 7: **Weights by age of onset and age at death (powered)**

AO	1	10	20	40	60	70
AD 1						
10	0.47					
20	0.95	0.87				
40	1.25	1.16	1.04			
60	1.09	1.00	0.90	1.00		
70	0.95	0.88	0.78	0.86	1.09	
80	0.80	0.74	0.65	0.73	0.93	1.05

A different approach: compensating-variation-based weights

Another approach to calculating the relative weights attached to different types of QALYs, or beneficiaries of QALYs, is to use the Hicksian compensating variation approach to welfare measurement. The method for calculating the compensating variation using discrete data is due to Small and Rosen (1981) and was introduced to the health economics literature in the context of discrete choice experiments by Lancsar and Savage (2004).

In general, the compensating variation (CV) is calculated by valuing in monetary terms the change in expected utility due to a policy change (e.g. change in price or quality of a good/service) as the change in income required to return the individual to their initial level of utility, that is, to compensate them for the change.

The CV for discrete choice data takes the following form:

$$CV = -\frac{1}{\lambda}\left[\ln\sum_{j=1}^{J} e^{V_j^0} - \ln\sum_{j=1}^{J} e^{V_j^1}\right] \qquad [4]$$

where V_j^0 and V_j^1 are the value of the indirect utility function (equation (4) above) for each choice option j before and after the policy change, respectively; J is the number of options in the choice set; and λ is the usually the marginal utility of income, or its proxy.

While a monetary value is the most convenient to turn the change in expected utility into a common metric, in fact any quantitative metric could be used as the numeraire. In the current study, instead of using income it is possible to calculate the CV for a move from one health state to another using the marginal utility of QALY gains as the numeraire. That is, we calculate the number of additional QALYs required to be given to or taken away in the new health state to equalise the utility derived from the base and new health states.

Like the probability approach described above, the CV approach can be used to value changes in, and create weights for, whole scenarios (i.e. entire health states). In addition, the CV can also be calculated to value and derive weights for individual characteristics such as age of onset, age of death and severity. To calculate both types of weights we value changes from an initial health sate, namely the same reference case as used in the calculation of the probability weights: AO = 40, AD = 60, QL = 0.7 and QALY = 4.

From this initial health state we can calculate the compensating variation, measured in numbers of QALYs, associated with a move to a new health state, described by new levels on the above attributes. This can involve an entire scenario or the valuation of each attribute one at a time. We use these CV measures to calculate weights for entire health states and for individual characteristics that describe the health states, by taking the ratio of the total QALYs required in the new health state to equalise utility and the original number of QALYs in the base model. This welfare theoretic approach also allows investigation of the strength of preference, as indicated by the magnitude of the CV.

The CV and relative weights reported in this section used the clogit results for the log model containing age of onset, age of death, QOL lost and QALY gains along with their higher order effects reported in Table 2. The CV and relative weights for the entire health states described in Table 3 are reported in Table 8. In each of these QALYs are held constant at 4.

Table 8: **Weights based on compensating variation approach per scenario (Version I)**

	Age onset	Age death	QOL lost	QALY gain with treatment	CV	Weight
Base	40	60	0.7	4		
Comparisons						
1	1	1	1	4	0.022365	1.005591
2	1	10	0.7	4	-0.5788	0.855301
3	1	10	0.1	4	-0.72801	0.817999
4	1	20	0.7	4	-0.04353	0.989118
5	1	20	0.1	4	-0.17872	0.955321
6	10	10	1	4	-0.36501	0.908748
7	10	20	0.7	4	-0.10969	0.972579
8	10	20	0.1	4	-0.2468	0.9383
9	10	40	0.7	4	0.118974	1.029743
10	10	40	0.1	4	-0.01122	0.997195
11	40	40	1	4	0.359147	1.089787
12	40	60	0.7	4	0	1
13	40	60	0.1	4	-0.13389	0.966528
14	70	70	1	4	0.405025	1.101256
15	70	80	0.7	4	0.039008	1.009752
16	70	80	0.1	4	-0.09369	0.976577

Before discussing the results in Table 8, it is important to note that "QALY gain with treatment" represents the number of QALYs offered in the base case and in the comparison cases as described in the health state scenarios. This is not the same as the "QALY gain" in the probability weight tables which is the result of the calculation of the number of QALYs required to equalise the probability of choosing the base case and alternative health states.

Both the CV and weights are interpreted relative to the reference case. The same reference case is included as comparison 12 for which the CV is zero, as would be expected since the health state has not changed, and the weight is 1. Relative to this base case, a negative CV indicates that individuals are willing to pay to secure the new health state – see for example, comparison 2. A positive CV indicates that individuals require compensation (in terms of being given additional QALYs) in the new health state – see for example, comparison 1. A weight greater than one indicates a preference for the reference health state whereas a weight less than one indicates a preference for the new health state. For example, the weight of 0.855 indicates that a QALY given in the base health state is worth only 0.855 of a QALY given in the new health state.

For the 4 comparison health states that involve instant death without treatment (comparisons 1, 6, 11, 14) there is a preference for giving the QALYs to the base case except for comparison 6 which favours the alternative. This is in accord with the underlying utility values derived from each health state. That is, the utility derived from the base case is greater than the utility derived from the alternatives except for comparison 6 (a 10 year old who dies instantly without treatment) which yields a higher utility than the base case.

Looking at the magnitude of the CV relative to the base case, respondents were willing to pay the largest amount (in terms of QALYs) for comparison 3, namely a health state in which the individual becomes unwell at age 1, without treatment QOL drops by 90% and they die at age 10 whereas with treatment they gain 4 QALYs. The least preferred beneficiary of 4 QALYs is someone who becomes unwell at age 70 and without treatment they die instantly at age 70.

Welfare measures and relative weights can also be calculated for the same complete health states described in Table 8 above but where QALYs, in addition to the other attributes, are allowed to change between the initial and new health states. These results are presented in Table 9.

Table 9: Weights based on compensating variation approach per scenario (Version II)

	Age onset	Age death	Qollost	QALY gain with treatment	CV	Weight
Base	40	60	0.7	4		
Comparison						
1	1	1	1	1	0.856822	1.214205
2	1	10	0.7	1	0.385326	1.096331
3	1	10	0.1	1	0.264184	1.066046
4	1	20	0.7	1	0.806518	1.20163
5	1	20	0.1	1	0.702215	1.175554
6	10	10	1	1	0.556146	1.139036
7	10	20	0.7	1	0.755657	1.188914
8	10	20	0.1	1	0.649138	1.162285
9	10	40	0.7	1	0.929927	1.232482
10	10	40	0.1	1	0.831226	1.207806
11	40	40	1	1	1.108226	1.277057
12	40	60	0.7	1	0.839787	1.209947
13	40	60	0.1	1	0.736962	1.184241
14	70	70	1	1	1.14171	1.285427
15	70	80	0.7	1	0.869471	1.217368
16	70	80	0.1	1	0.767985	1.191996
17	1	1	1	2	0.489543	1.122386
18	1	10	0.7	2	-0.04604	0.988491
19	1	10	0.1	2	-0.1813	0.954675
20	1	20	0.7	2	0.431636	1.107909
21	1	20	0.1	2	0.312186	1.078047
22	10	10	1	2	0.146231	1.036558
23	10	20	0.7	2	0.373286	1.093322
24	10	20	0.1	2	0.25171	1.062928
25	10	40	0.7	2	0.57406	1.143515
26	10	40	0.1	2	0.460053	1.115013
27	40	40	1	2	0.782104	1.195526
28	40	60	0.7	2	0.469911	1.117478
29	40	60	0.1	2	0.351889	1.087972
30	70	70	1	2	0.821494	1.205374
31	70	80	0.7	2	0.504136	1.126034
32	70	80	0.1	2	0.387411	1.096853
33	1	1	1	10	-0.95935	0.760162
34	1	10	0.7	10	-1.65684	0.58579
35	1	10	0.1	10	-1.82564	0.543589
36	1	20	0.7	10	-1.03736	0.740661
37	1	20	0.1	10	-1.1961	0.700975
38	10	10	1	10	-1.41221	0.646947
39	10	20	0.7	10	-1.11525	0.721187
40	10	20	0.1	10	-1.27542	0.681144
41	10	40	0.7	10	-0.84419	0.788951
42	10	40	0.1	10	-0.99917	0.750209
43	40	40	1	10	-0.55345	0.861638

44	40	60	0.7	10	-0.98588	0.753531
45	40	60	0.1	10	-1.14365	0.714088
46	70	70	1	10	-0.49713	0.875719
47	1	1	1	15	-1.61616	0.595961
48	1	10	0.7	15	-2.35655	0.410863
49	1	10	0.1	15	-2.53376	0.366561
50	1	20	0.7	15	-1.6997	0.575074
51	1	20	0.1	15	-1.8691	0.532724
52	10	10	1	15	-2.09848	0.475381
53	10	20	0.7	15	-1.78293	0.554267
54	10	20	0.1	15	-1.95346	0.511636
55	10	40	0.7	15	-1.49242	0.626895
56	10	40	0.1	15	-1.65882	0.585294
57	40	40	1	15	-1.17774	0.705566
58	40	60	0.7	15	-1.64459	0.588853
59	40	60	0.1	15	-1.81322	0.546695
60	1	1	1	30	-3.20126	0.199686
61	1	10	0.7	30	-4.00362	-0.00091
62	1	10	0.1	30	-4.19263	-0.04816
63	1	20	0.7	30	-3.29298	0.176755
64	1	20	0.1	30	-3.47795	0.130512
65	10	10	1	30	-3.72645	0.068388
66	10	20	0.7	30	-3.38402	0.153994
67	10	20	0.1	30	-3.56959	0.107603
68	10	40	0.7	30	-3.06477	0.233808
69	10	40	0.1	30	-3.24814	0.187964
70	40	40	1	30	-2.71385	0.321537
71	40	60	0.7	30	-3.23251	0.191872
72	40	60	0.1	30	-3.41708	0.145731

Again a weight greater (less) than 1 indicates a preference for the reference (alternative) health state. Looking at these results in aggregate, the number of QALYs gained is driving the preference between the reference and comparison health states. Without exception, for QALY gains in the new health state of 10, 15 and 30, the new health state is always preferred to the reference (which only contains 4 QALYs). For all comparison health states involving only 1 QALY gain, the reference case (containing 4 QALYs) is always preferred. For comparison health states with 2 QALY gains, the reference case is generally preferred to the comparison case except for when those 2 QALYs are given to 1 year olds whose QOL drops by 0.1 and die at age 10 or 1 year olds whose QOL drops by 0.7 and die at age 20.

The CV and relative weights associated with the individual attributes (age at onset, age of death and severity) are reported in Table 10. These calculations used the same initial base health state of age of onset of 40, age of death of 60, QOL lost of 0.7 and 4 QALYs. The CV and weights per attribute level were calculated by changing one attribute at a time in the new health state, holding all else constant.

Relative to the base levels used in the reference (age onset = 40, age death = 60 and QOL lost = 0.7), more importance is placed on higher levels of severity and this is monotonic, as would be expected. In particular, respondents would have to be compensated a relatively large amount in absolute value for a health state that involved the highest level of severity. WTP (in terms of QALYs) to improve the initial level of severity, that is to only loose 10 or 40 percent of QOL rather than the base case of 70 percent is also monotonic. Age of onset of 20 and 40 appear to be most important which could relate to considerations of the ages at which individuals are most productive to society (and also more likely to have dependents). Higher preference is given to those who die young (aged 10, 1 and 20).

Table 10: Compensating variations per attribute

Attribute	Level	CV	Weight
Age of onset	1	0.068797	1.017199
	10	0.004333	1.001083
	20	-0.08623	0.978442
	40	0	1
	60	0.174144	1.043536
	70	0.267529	1.066882
Age of death	1	-0.31725	0.920688
	10	-0.65684	0.83579
	20	-0.11413	0.971467
	40	0.114759	1.02869
	60	0	1
	70	-0.11082	0.972295
	80	-0.24383	0.939042
Severity (QOL loss)	1	0.251673	1.062918
	0.7	0	1
	0.4	-0.17301	0.956748
	0.1	-0.13389	0.966528

7. Issues for discussion

The main purpose of this paper has been to expose the reader to three things: the challenges of deriving relative weights for QALYs using DCE methods; two novel approaches to eliciting relative QALY weights from a DCE; and some initial empirical weights.

On the last of these, a first point for discussion is what are people’s initial reactions to the weights. Overall, the estimated weights are similar across different regression models and different weighting procedures. There are obviously differences between the simple model and that including higher order polynomials due to the imposed restriction of monotonicity in the former. Comparing the CV weights with those from the probability models we can see that the range of weights is smaller (0.818-1.1) for the former, and are generally nearer to 1, suggesting no difference between the base case and the comparison scenario. More generally, this would indicate that age and severity have little impact on QALY gains. The range for the probability models is 0.38–1.61. Since the weights are based on the same regression model it is unsurprising that same scenarios result in the lowest and highest weights in both methods.

There are clear patterns, as the age of onset increases the weights move towards 0 (until age 20) and then move towards 1. This shows a preference for those who fall ill between the ages of 10 and 40, with 20 being the most preferred. As age of death increases the weights move towards 0 (until age 10) and then move towards 1 (up to age 40) before moving back towards 0. This shows a preference for individuals who ill die at 1, 10 and 80 but less so for those between 20 and 70. As severity increases there is a preference for those who are slightly ill, a quality of life loss of 0.4 and 0.1, compared to those who lose 0.7 and 1. Of course, the advantage with these approaches is that it is possible that age effects outweigh severity effects, or vice versa, meaning that generalisations are harder to make unless we can invoke ceteris paribus assumptions. In other words, different weights on QALYs are required for a variety of contexts!

With respect to more methodological issues, we have demonstrated two novel approaches to deriving weights from our empirical data. What issues arise from people’s perceived similarities or differences in the weights? At a conceptual level, the compensating variation and probability approaches to deriving relative weights are similar. However, instead of equalising the probability of choice between the two health states, the CV method equalises the utility associated with the two health states; a subtle but important difference. Does this difference matter? One possibility, of course, is that using the CV approach overcomes one limitation of the probability approach in that the CV can be used to calculate weights per attribute in addition to weight for entire health state scenarios.

There are then several issues to discuss in relation to challenges:

- despite a poorly-performing design, can we place any reliance in what appear to be quite consistent results across approaches to weighting QALYs? Indeed, a point not mentioned up to now is that the approaches came up with the exact same ordering of scenarios, although this may be regarded as unsurprising given the similarities in the weighting methods.
- can we really say anything sensible about weights given the inter-relationships between all of the dependent variables?
- Because of such inter-relationships and unobservalables, are we, for example, getting a weight for age at onset or is it more to do with potential QALYs to be gained, some kind of ‘fair innings’, QALYs experienced or some other factor?
- A key missing element is that of uncertainty. While the coefficients are estimated with confidence intervals, the weights are simply point estimates. If our coefficients from the regressions are imprecise, which we know they are given the challenges of the design, then the weights will have wide confidence intervals. How has this challenge been met in past valuation and relativities studies?

References

Baker R, Thompson C and Mannion R (2006) Q Methodolgy in Health Economics. Journal of Health Services Research and Policy 11: 38-45.

Coast J and Horrocks S (2007) Developing attributes and levels for discrete choice experiments using qualitative methods. Journal of Health Services Research and Policy 12: 25-30.

Cropper ML, Aydede SK and Portney PR (1994) Preferences for life saving programs: how the public discounts time and age. Journal of Risk and Uncertainty 8: 243-265.

Dolan P and Olsen JA (2002) Distributing health care: Economic and Ethical Issues. Oxford University Press, Oxford.

Dolan P, Shaw R et al. (2005) QALY maximisation and people's preferences: a methodological review of the literature. Health Economics 14: 197-208.

Donaldson C, Atkinson A, Bond J, Wright I (1988) Should QALYs be programme-specific? Journal of Health Economics 7: 239-257.

Donaldson C and Wright KG (1989) Programme-specific QALYs: A reply. Journal of Health Economics 8: 489-491.

Johannesson M and Johansson P-O (1996) The economics of ageing: on the attitude of Swedish people to the distribution of health care resources between the young and the old. Health Policy 37: 153-161.

Johannesson M and P-O Johansson (1997) Is the valuation of a QALY gained independent of age? Some empirical evidence. Journal of Health Economics 16: 589-599.

Johri M, Dmaschroder LJ, Zikmund-Fisher BJ and Ubel PA (2005) The importance of age in allocating health care resources: does intervention-type matter? Health Economics 14: 669-678.

Lancsar E and Savage E (2004) Deriving welfare measures from discrete choice experiments: inconsistency between current methods and random utility and welfare theory. Health Economics Letters 13: 901-907.

Mason H (2007) Monetary valuation of health outcomes for use in national policy formulation. PhD Thesis, Newcastle University.

National Institute for Health and Clinical Excellence (2007) Guide to the Methods of Technology Appraisal. NICE, London.

Nord E (1999) Cost-Value Analysis in Health Care: Making Sense out of QALYs. Cambridge University Press, Cambridge.

Nord E, Street A et al. (1996) The significance of age and duration of effect in social evaluation of health care. Health Care Analysis 4: 103-111.

Ratcliffe J (2000) Public preferences for the allocation of donor liver grafts for transplantation. Health Economics 9: 137-148.

Rawlins M. and Culyer AJ (2004) National Institute for Clinical Excellence and its value judgements. BMJ 329: 224-227.

Schwappach DL (2002) Resource allocation, social values and the QALY: a review of the debate and empirical evidence. Health Expectations 5: 210-22.

Schwappach DL (2003) Does it matter who you are or what you gain? An experimental study of preferences for resource allocation. Health Economics 12: 255-67.

Small K and Rosen H (1981) Applied welfare economics with discrete choice models. Econometrica 49: 105 - 130.

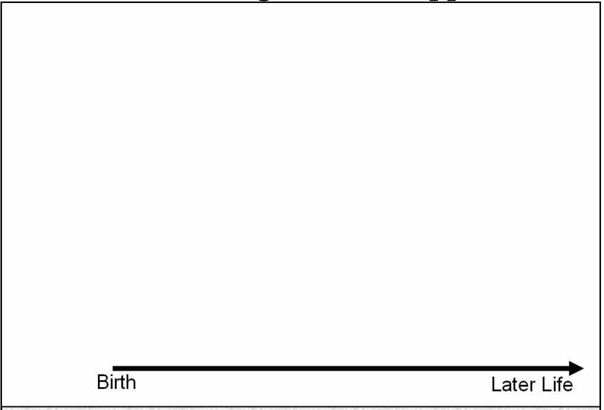
Torrance GW (1986) Measurement of health state utilities for economic appraisal. Journal of Health Economics 5: 1-30.

Weinstein M (1988) A QALY is a QALY is a QALY - or is it? Journal of Health Economics 7: 489-491.

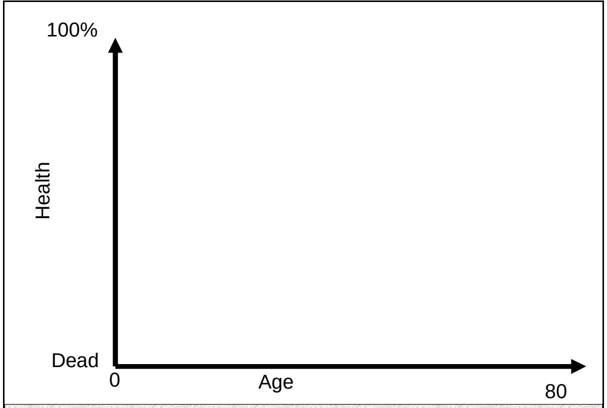
Williams A (1985) Economics of coronary artery bypass grafting. British Medical Journal, 291: 326-329.

Willams A (1996) QALYS and ethics: a health economist’s perspective. Social Science & Medicine 43: 179-1804.

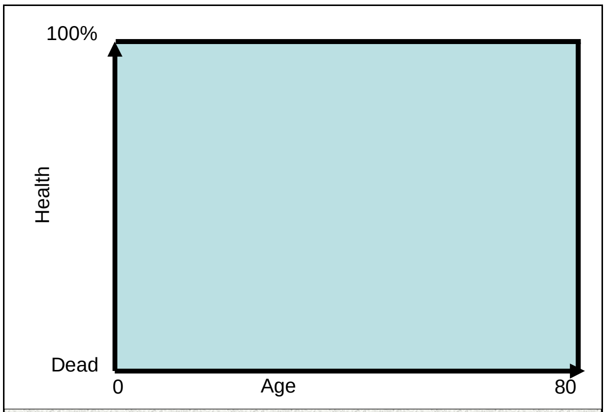
APPENDIX 1: Diagrammatic approach to explaining QALYs



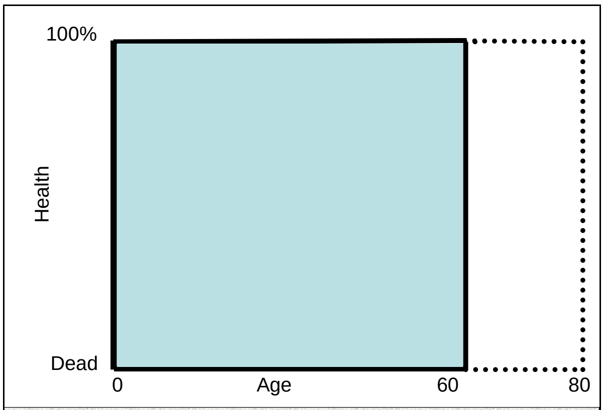
We can represent how long someone lives with a line starting from their birth and going on into later life.



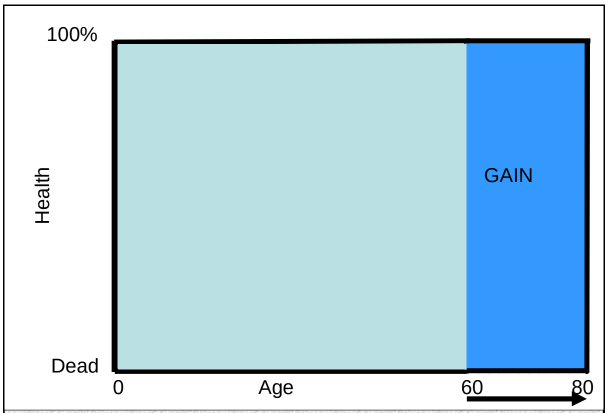
We can show this with what you might think of as a ‘health thermometer’, where dead is zero and normal good health is marked as 100%.



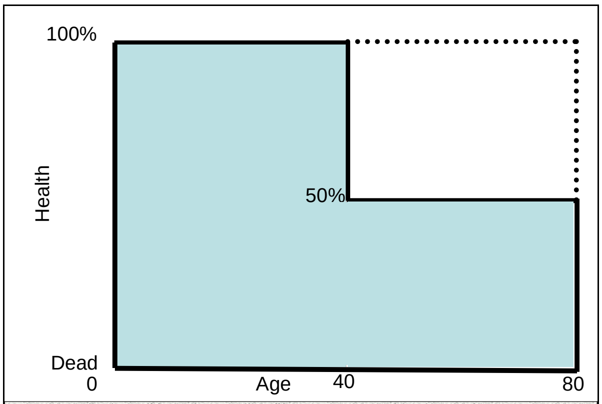
This picture shows someone in normal good health throughout their life up to age of 80. The sudden drop of the black line to zero shows death at aged 80.



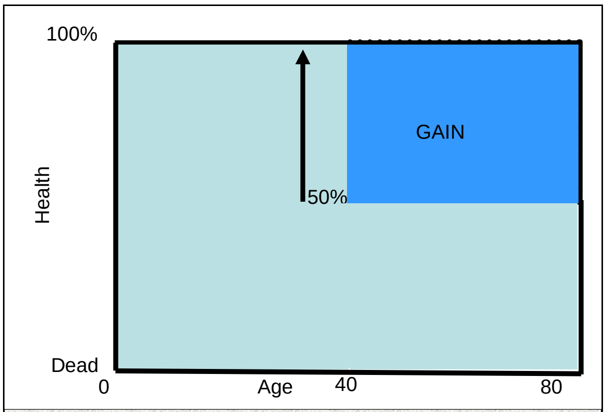
This picture shows someone with a life threatening medical condition at the age of 60. This is shown by the line dropping to zero at age 60.



This picture shows a successful treatment for that condition. The treatment gives them an extra 20 years life expectancy.



Some medical conditions do not affect life expectancy. This picture shows someone aged 40 who suffers a condition which reduces their health from 100% to 50% for the rest of their life.



With successful treatment their health can be restored to 100%.

APPENDIX 2: Constraints on DCE design

1. Age at death must be no less than age of patients.

2. The sum of expected age at death if untreated and gain in life expectancy must not exceed 80 years.

3. If age at death if untreated is equal to age of patients, and gain in quality of life with treatment is 0, then gain in life expectancy must be 0.

4. If age at death if untreated is equal to age of patients, then quality of life if untreated is 0. If age at death if untreated is greater than age of patients, then quality of life if untreated is not 0.

5. Since the gain in quality of life with treatment has a maximum of 100%, the sum of quality of life if untreated and gain in quality of life with treatment is less than or equal to 100%.

6. Gain in quality of life with treatment and gain in life expectancy cannot be both 0.

7. If age at death if untreated is equal to age of patients, and gain in life expectancy is 0, then gain in quality of life with treatment is 0.

8. If gain in life expectancy is not 0, then the sum of quality of life if untreated and gain in quality of life with treatment cannot be 0.