DO MEMBERS OF THE PUBLIC WISH TO ATTACH MORE WEIGHT TO SOME QALYs THAN TO OTHERS: THE 'QALY GRID' APPROACH

Angela Robinson, Rachel Baker, Cam Donaldson, Graham Loomes, Jose Luis Pinto & Bob Sugden

On behalf of the SVQ research team:

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

[1] University of East Anglia

[2] Newcastle University

[3] University of Aberdeen

[4] University of Sheffield (formerly Newcastle)

[5] London School of Hygiene & Tropical Medicine (formerly UEA).

[6] University of Seville (formerly on sabbatical at UEA).

For full project report see;

The SVQ Research Team (2008) *Weighting and valuing quality adjusted life years: preliminary results from the social value of a QALY project.* Report for the NIHR Methodology Programme, January 2008.

INTRODUCTION

This paper reports an empirical study that set out to determine whether the UK public wish to attach more weight to some quality adjusted life years (QALYs) than to others. In the standard 'unweighted' QALY model, all QALYs are of equal value. For example, the age of recipients does not matter, so long as the QALY gain is the same. Likewise, the standard model assumes that equal QALY gains are of equal value regardless of how severely ill the patients are prior to treatment. The aim of the research outlined in this paper was to address the question 'Is a QALY a QALY a QALY? i.e. is the 'standard' model correct? For example, should a particular QALY gain delivered to someone currently in very poor health be valued differently from the same number of QALYs delivered to people whose current health state is not nearly so bad? Should a paediatric QALY be valued differently from a geriatric QALY? And should either or both of these be given greater or lesser value than a QALY delivered to a typical cross-section of the population? The work was undertaken as part of the Societal Value of a QALY (SVQ) project, other aspects of which have been presented at previous HESGs (Mason et al, 2005, Baker et al, 2008, Loomes, 2008).

Briefly, the SVQ project consisted of three separate, but conceptually related, strands. The first of these involved the derivation of the monetary value of a QALY from the value that the UK Department for Transport (DfT) applies to a safety improvement that can be expected, on average, to prevent one fatality – this value being referred to as the "value of preventing a statistical fatality" (VPF). A second strand of the project assessed the *feasibility* of obtaining an estimate of the monetary value of a QALY by presenting members of the public with appropriately framed valuation questions in a survey. The third strand –the 'relativities' study-aimed to investigate the extent to which members of the public consider that there may be legitimate grounds for distinguishing between the social value to be accorded to QALYs gained by different types of patient.

Two general approaches were taken to the relativities study:

1. the 'QALY grid' approach based on Person trade off (PTO)questions.
2. a discrete choice experiment (DCE)

The results of the DCE study have been reported in a previous HESG (Baker et al, 2008), so we concentrate here on the PTO study, the aims of which were:

- to estimate the relative weights to be attached to a QALY according to the age of recipients.
- to estimate the relative weights to be attached to a QALY according to the severity of illness of recipients prior to treatment (i.e. quality of life if 'untreated') .

The review paper by Dolan and colleagues (Dolan 2005) cites a number of previous studies that have set out to estimate age and severity weights and report a range of values. With regard to age, Johannesson and Johanesson (1997) report weights for life years gained at ages 30, 50 and 70 years of; 1.0, 0.22 and 0.1 respectively- indicating a 10:1 differential between the most and least preferred age groups. Adjusting for quality of life affected these estimates somewhat with three QALYs gained for 50 year olds or 9 QALYs gained for 70 year old judged equivalent to one QALY gained for a 30 year old. Similarly, Nord and colleagues (1996) derived weights for life years gained at ages 10, 20, 60 and 80 of; 1.1, 1.0, 0.4 and 0.1 respectively – again indicating a more than 10:1 differential between the most and least preferred age groups. This pattern was very similar when life improving- rather than life- extending interventions were considered. Busschbach and colleagues (1993) compared quality of life improvements at ages 5, 10, 35, 60 and 70 and estimated weights for the utility of health at these ages to be; 0.2, 1.5, 1.0, 0.7 and 0.7 respectively, indicating a 7.5: 1 differential between the most preferred and least preferred age groups.

With regard to severity, the results appear to be mixed. are mixed. Nord (1996) and Ubel and colleagues (1998) asked subjects to compare improvements on a disability scale (where lower numbers indicate better functioning) with approximately equal distances between the levels. Respondents were asked to indicate how many patients moving from level 5 to level 1 was equivalent to moving a smaller number of patients from level 6 to level 4. Both studies report a marked preference for treating the more severely ill patients. In contrast, a study by Dolan and Tsuchiya (2005) found that respondents gave consistently higher priority to patients with better prospects without treatment indicating that the more severely ill attract *lower* priority than the better off.

Comparing the results of studies directly, however, is not straightforward as different 'head to head' comparisons of ages were made and how 'severity' was operationalised differed across studies. And, of course we have no means of combining the data on age and severity into an overall weighting system to be applied to QALYs. This may be important as it is plausible that preferences over age and severity may not be independent. For example, it may be that the relative weight given to a return to full health, compared to a similar 'move' lower down the severity scale, may be different for the young and the old. That is, full health may be considered to be *relatively* more important at certain 'life stages' than others.

The 'QALY grid' approach outlined below sets out to offer a systematic means of eliciting preferences over age and severity that allows the two to be combined into a single set of weights.

METHODS

The basic approach

Briefly, PTO questions ask people to state the number of outcomes of one kind they consider to be 'just as good as' a specified number of outcomes of another kind. A typical question in this setting would be something along the following lines:

Consider two groups of 100 people: 100 of Type A and 100 of Type B (these types being described and differentiated by age and severity of illness prior to treatment). With treatment, each of these types could experience some given health benefit. If there were only enough resources to treat one Group, which would you prefer it to be?

Suppose the respondent chooses to give priority to Group A. The number of beneficiaries in Group B is held fixed at 100 while the number of A-types is reduced to the point where the respondent finds it hard to prioritise between that number (X) of A-types and 100 B-types. By systematically varying the age and health of the people described in Groups A and B in a series of questions, it is possible to investigate the relative weights attached to health gains to people of different ages and levels of health.

The exploratory work, which lasted for over a year, has been described previously (Baker et al, 2008). Briefly, methods for presenting information about age, severity and health gain were devised in focus groups and tested in cognitive interviews. The most successful technique was achieved through the use of diagrams which were first explained using an innovative animated Powerpoint presentation. The survey was administered by computer assisted personal interview (CAPI) 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 QALY grid

The animated Powerpoint introduction to the CAPI used a graph with Quality of Life (from 0% to 100%) on the vertical axis and Age (from 0 to 80) on the horizontal. Within this evaluative space, there are clearly an enormous number of potential PTO pairings that can be made and the challenge was to come up with the most parsimonious design that allowed all the research questions to be addressed. We were keen that the design was capable of testing the standard 'unweighted' QALY model – i.e. where all relative weights would be equal to one – as well as being able to detect systematic departures from this baseline assumption with respect to the age of patients and their 'untreated' quality of life. To that end we developed the 'QALY grid' approach summarised briefly below. By partitioning the vertical of the quality of life graph into five 20% ranges and dividing the horizontal into four 20-year intervals, we get twenty 'cells' (numbered 1-20 in Table 1 below).

Table 1 : The QALY grid

QoL	**A**	**B**	**C**	**D**	
80-100%	17	18	19	20	P
60-80%	13	14	15	16	Q
40-60%	9	10	11	12	R
20-40%	5	6	7	8	S
0-20%	1	2	3	4	T
Age	0-20	20-40	40-60	60-80	

Improving a person's health by any one cell – i.e. by 20% per year for 20 years – would, under the 'standard' model, give 4 QALYs. But are all QALYs weighted equally, irrespective of where they are located in the grid? For example, suppose that type A people were those whose QoL would, if untreated, be 60% at (and after) age 20 and treatment could raise that QoL to 80% for the next 20 years, whereas type B people would be at 80% from the age of 60 but could receive treatment that would restore them to 100% for the subsequent 20 years. Both types would gain 4 'standard' QALYs: but do respondents give both treatments equal priority? If not, which Group of 100 do they give priority to? And if the number of beneficiaries in that group is reduced, at what value of X < 100 do the two treatments receive equal priority? For example, if a respondent would find it hard to prioritise when the same total expenditure of resources could either treat 40 people of type A or else 100 people of type B, that would suggest that such a respondent gives two-and-a-half times as much weight to a QALY in cell 14 as in cell 20.

If we have data on respondents' relative values for the 20 cells and the value of a 'reference' cell – is set equal to 1 (suppose we call that v_1), we can derive a set of relative weights (v_2-v_{20}) for the remaining 19 cells in the grid. Our basic aim was to estimate these relative values. If each cell is paired with every adjacent cell, we can estimate ratios such as v_1/v_2 and v_1/v_5. This requires 31 pairings. An example of a question respondents were actually presented with is reproduced in Figure 1. This represents the comparison of cells 14 and 15 in the grid and shows the situation both the without (Figure 1a) and with treatment (Figure 1b)

Figure 1a : Screen shots from CAPI- without treatment

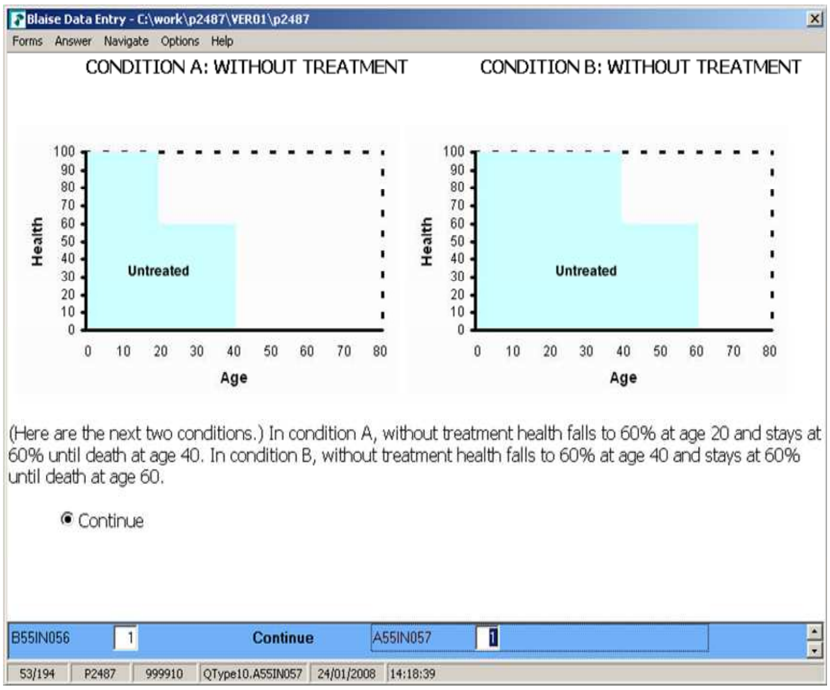
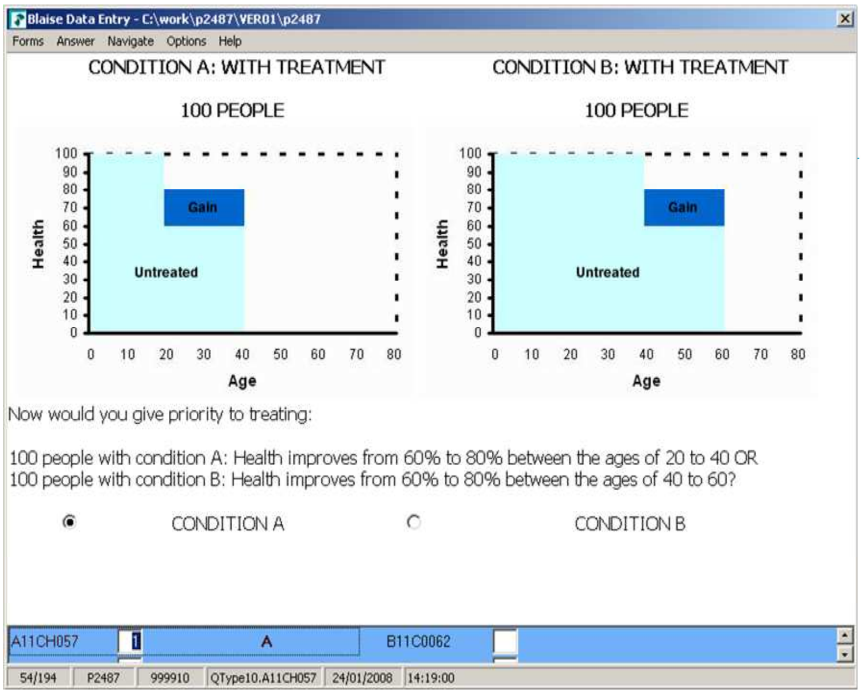


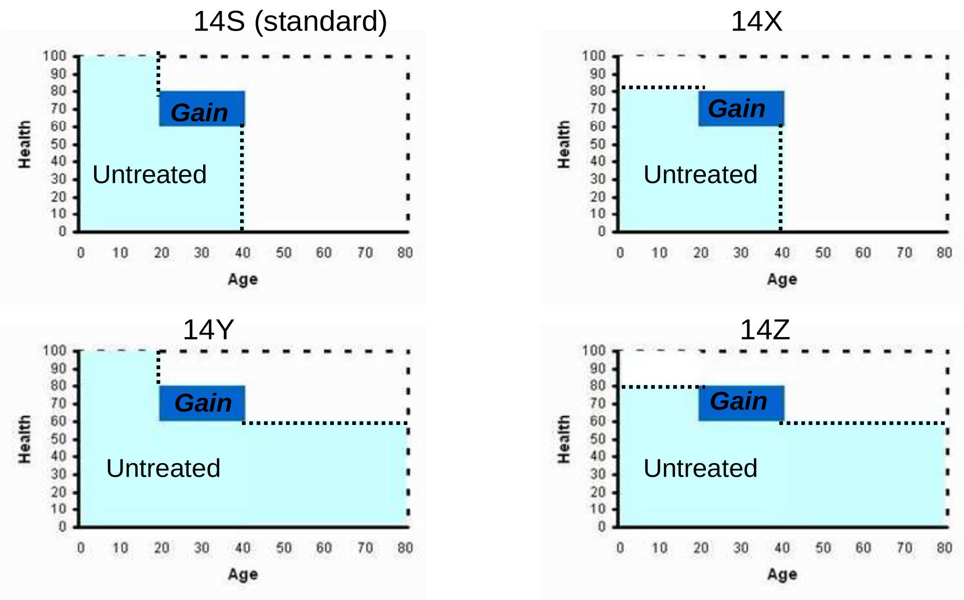
Figure 1b : Screen shots from CAPI- with treatment



Whilst the focus of the current paper is the results of these 31 adjacent pairing questions, another 41 questions in the overall design were included to allow a number of different tests of consistency. For example, some of these questions look at blocks of cells: for example, each row P-T is paired with every adjacent row, and each column with every adjacent column. The design also allows us to test for transitivity via chaining with equal numbers of steps, e.g. $(v_1/v_5)*(v_5/v_9)$ should equal (v_1/v_9). Space does not permit a detailed account of this data here.

One set of tests we do discuss briefly here are the tests for independence of the profile in which a cell appears. Consider the different ways in which cell 14 may appear in a scenario. In Figure 2 below, 14S shows the patient to be in good health until aged 20 when – without treatment -their quality of life falls from 100% to 60% and they die at the age of 40.

Figure 2 : The profile tests (cell 14)



This depicts the 'standard' representation that applies to all other single cell comparisons used here. In 14X the patient is at 80% of good health until aged 20 when their quality of life falls to 60% and again they die at the age of 40. In 14Z, the patient is at 80% of good health until aged 20 when their quality of life falls to 60%, but now they live until age 80. Finally, in 14Y the patient is in good health from birth and lives to aged 80. Each profile test compares the 'standard' representation with one of the 3 other representations of that cell (i.e. 14S vs 14X, 14S vs 14Z, 14S vs 14Y). Three different sets of profile tests were carried out involving cells 7, 11 & 14. 'Fair innings' type arguments would suggest a tendency – all else equal- to favour those patients whose total lifetime QALYs were lower. Table 2 gives the total lifetime untreated QALYs in each representation (S,X,Y&Z) of the 3 cells used in the profile tests.

Table 2: The profile tests- number of QALYs if untreated.

	S	**X**	**Z**	**Y**
Cell 7	44	20	24	48
Cell 11	48	32	40	56
Cell 14	32	28	52	56

For each of the 3 cells used in the profile tests, a 'fair innings' type argument would predict that *more* weight would be given to S than to Y (as the total lifetime QALYs is less in S than Y) , but S would be given *less* weight than X (as the total lifetime QALYs is greater in S than X). Finally, S would be given less weight than Z in the cases of cells 7 & 11, but S would be given more weight than Z in the case of cell 14.

The 72 questions were spread evenly across 12 versions of the questionnaire, each respondent answering 6 PTO questions, 8 DCE questions as well as attitudinal and socio-demographic questions.

Aggregating the results.

As outlined previously (Robinson et al, 2008) there are a number of ways to aggregate PTO results and different aggregation methods yield different results. It is not our intention to go into detail here, but two aggregation methods- ratio of means and medians of individual ratios -are reported below.

The sampling frame

The population surveyed was the population, aged 18 and over, living in private households in England. The sampling frame was the small users Postcode Address File from which 40 postcode sectors were selected as the primary sampling units. 30 addresses were randomly selected from each postcode sector with 1200 addresses selected in total. Within each household, only one adult aged 18 and over was eligible to be included in the study. Where there was more than one adult aged 18 and over present in a household, interviewees were selected randomly.

RESULTS

The sample

A total of 587 interviews were completed. 243 of 587 (41%) respondents were male, 344 (59%) female. Mean (and median) age was 52 years with a range of 18 to 92. 527 of the 587 (90%) classified themselves as 'white', whilst 326 (56%) left school at aged 16 or under. A total of 414 (70.5%) classified their health as either 'very good' or 'good' whilst 26 respondents (4.4%) classified their health as either 'bad' or 'very bad'. 283 (49%) had no problems on any EQ-5D dimension, whilst 31(5.3%) had severe problems on at least one dimension.

Responses to first PTO question

Before going on to look at the weights derived from the aggregate PTO data, we begin by looking at the number of respondents choosing to give priority to one group or the other in the first iteration i.e. when the number of patients treated was 100 in both groups. These data are the most straightforward to interpret and are arguably the simplest way of illustrating the *general patterns* to emerge from the relativities study.

With regard to the general pattern relating to age, a total of 269 respondents answered a PTO question involving an equivalent gain going either to 40 to 60 year olds or to 60 to 80 year olds. Of these 269, 173 (64%) preferred to give priority to the younger patients. Similarly, a total of 294 respondents answered a question involving an equivalent gain going either to 20 to 40 year olds or to 40 to 60 year olds. Of the 294, 182 (62%) preferred to give priority to the younger patients. In contrast, a total of 295 respondents answered a PTO question involving an equivalent gain going either to 0 to 20 year olds or to 20 to 40 year olds. Of the 295, only 107 (36%) preferred to give priority to the younger patients. This suggested that the pattern relating to age is not a simple one and that ‘non-linearities’ may exist with respect to age weights.

This pattern is borne out if we look only at responses to the whole column questions – detailed above – which involved gains of 20 QALYs. Of the 51 respondents who were asked to compare column A – 20 years in full health for a new born – with column B – 20 years in full health for a 20 year old – 17 (33.3%) preferred to treat the younger patients. In contrast, of the 54 respondents who compared column C – 20 years in full health for 40 year olds – with column D – 20 years in full health for 60 year olds – 36 (66.7%) preferred to treat the younger patients.

With regard to the general pattern relating to severity, a total of 294 respondents answered a PTO question involving a move either from 60% to 80% health or from 80% to 100% health (for a given age group). Of the 294, 170 (58%) preferred to give priority to patients in the more severe health state. In contrast, of the 280 respondents who answered a question involving a move either from 0% to 20% health or from 20% to 40% health (for a given age group), only 105 (38%) preferred to give priority to those patients in the more severe health state. Again, observing these initial choices suggested that non-linearities may exist with respect to severity weights. The pattern is borne out by observing responses to the ‘whole row’ questions. Of the 44 respondents who were asked to compare row T – a move from 0% health to 20% health between birth and aged 80 - and row S - a move from 20% health to 40% health between birth and aged 80 – 17 (38.6%) preferred to treat those patients who were worse off initially. In contrast, of the 44 respondents who were asked to compare row Q and row P, 32 (71.1%) preferred to treat those patients who were worse off initially.

As it seemed plausible that the desire to prioritise one age group over another may be related to the age of respondents, we also broke these data down by age of respondent. The sample was divided into 3 broad age groups: the under 40 year olds, the 40 to 60 year olds and the over 60 year olds. These categories were chosen as they best coincided with those used in the QALY grid (as respondents had to be 18 or over to participate in the survey, there are too few respondents to have a separate category for 0 to 20 year olds). The results showed that responses are significantly different according to age of respondent. The general pattern to emerge from the data is that older respondents seem more inclined than younger respondents to give preference to older patients, but this does not appear to be related just to a self- interested desire to prioritise their own age group (as they are also more likely to give preference to 20 to 40 year olds than to 0 to 20 year olds). The data on severity suggests that older respondents appear to place greater weight on returning patients to full health than their younger counterparts.

The relative weights

Recall that our 'QALY grid' partitions the vertical axis into five 20% ranges and the horizontal axis into four 20-year intervals, giving 20 'cells', each worth 4 QALYs under the 'standard' model. We then set out to determine whether all QALYs were weighted equally, irrespective of where they are located in the grid. We begin by presenting the results of the 31 single cell comparisons and the ratio of means aggregation method (see Robinson et al,2008).

Table 3: Horizontal (age) weights: ratio of means aggregation method.

QoL				
80-100%	*17* 0.837	*18* 1.559	*19* 1.834	*20* N/A
60-80%	*13* 0.781	*14* 1.131	*15* 1.564	*16* N/A
40-60%	*9* 0.676	*10* 1.039	*11* 1.383	*12* N/A
20-40%	*5* 0.716	*6* 1.298	*7* 1.433	*8* N/A
0-20%	*1* 0.867	*2* 1.228	*3* 1.687	*4* N/A
Age	0-20	20-40	40-60	60-80

The numbers in each cell of Table 3. signify the weight attached to that cell relative to the cell to the right. Values greater than one indicate that more weight is attached to that cell than the one to the right. Values less than one indicate that less weight is attached to that cell than to the one to the right. For example, a value of 1.564 in cell 15 indicates that a gain (in this case from 60% to 80% health) accruing to 40 to 60 years old is valued at 1.564 times the equivalent gain accruing to 60 to 80 years old. In contrast, the value of 0.676 in cell 9 indicates that a gain (in this case from 40% to 60% health) accruing to 0 to 20 year olds is worth 0.676 times the equivalent gain accruing to 20 to 40 years old.

Following this through for all other adjacent horizontal cell comparisons, the general pattern to emerge with respect to age (holding severity constant) is as follows:

- Less weight is given to treating 0 to 20 year olds than 20 to 40 year olds.
- More weight is given to treating 20 to 40 year olds than 40 to 60 year olds.
- More weight is given to treating 40 to 60 year olds than 60 to 80 year olds.

Though not shown in detail here, this broad pattern is borne out in the 'whole column' comparisons, indicating that this pattern holds for larger gains too.

Turning now to the vertical - or severity – weights the numbers in each cell of Table 4 signify the weight attached to that cell relative to the cell *above* it. Values greater than one indicate that more weight is attached to that cell than to the one above it and vice versa. For example, a value of 1.618 in cell 16 indicates that, for a given age group of patient, a gain from 60% health to 80% health is valued at 1.618 times that from 80% to 100% health. In contrast, a value of 0.801 in cell 4 indicates that a gain from 0% to 20% health is valued at 0.801 times that from 20% to 40% health (again keeping age of patient constant).

Table 4: Vertical (severity) weights: ratio of means aggregation method

QoL				
80-100%	*17* N/A	*18* N/A	*19* N/A	*20* N/A
60-80%	*13* 1.064	*14* 1.369	*15* 1.203	*16* 1.618
40-60%	*9* 0.792	*10* 1.144	*11* 1.179	*12* 1.247
20-40%	*5* 0.955	*6* 1.185	*7* 1.235	*8* 1.247
0-20%	*1* 0.786	*2* 0.759	*3* 0.857	*4* 0.801
Age	0-20	20-40	40-60	60-80

The *general* pattern to emerge with respect to severity (holding age constant) is as follows;

- Less weight is given to a move from 0% to 20% than from 20% to 40%.
- More weight is given to a move from 20% to 40% than from 40% to 60%.
- More weight is given to a move from 40 % to 60% than from 60% to 80%.
- More weight is given to a move from 60% to 80% than from 80 to 100%.

(The exception to this pattern is for the 0-20 age range, where a move from 60% to 80% is weighted most highly and where the weights drop away progressively above and below that cell.)

Given the broad regularity of the patterns for both age and severity – in both cases, the general shape is a (somewhat off-centre) inverted-U – one simple way of combining the two into a single set of weights for the 20 cells is as follows.
(i) Identify the average pattern showing how weights vary just with age, independent of severity.
(ii) Likewise, identify the average pattern showing how weights vary just with severity, independent of age.
(iii) Compute the weight for any cell as the cross-product of the appropriate age and severity weights, normalising overall so that the most highly weighted cell is indexed at 1.
Following this procedure, we get the implied weights given in Table 5.

Table 5: Implied relative weights: ratio of means method

QoL				
80-100%	*17* 0.533	*18* 0.688	*19* 0.560	*20* 0.362
60-80%	*13* 0.658	*14* 0.848	*15* 0.690	*16* 0.446
40-60%	*9* 0.679	*10* 0.876	*11* 0.713	*12* 0.461
20-40%	*5* 0.775	*6* **1.00**	*7* 0.844	*8* 0.527
0-20%	*1* 0.613	*2* 0.791	*3* 0.644	*4* 0.417
Age	0-20	20-40	40-60	60-80

Table 5 gives the estimated weight of each cell relative to the cell with the highest value - cell 6, where a gain from 20% to 40% health accrues to 20 to 40 year olds. The general pattern is that the valuation function peaks at cell 6, but then 'falls away' in all directions. Since the top right hand corner cell combines the lowest average weight assigned to severity with the lowest average weight assigned to age, it carries the lowest weight of any cell: on this basis, the differential between the highest and lowest valued cells is roughly 2.75:1.

As noted above, there are other ways of aggregating PTO data. Starting with weights based on medians of individual ratios and following the same procedure as outlined above, yields the implied weights in Table 6. .

Table 6: Implied relative weights: median of individual ratios method

QoL				
80-100%	*17* 0.47	*18* 0.72	*19* 0.54	*20* 0.24
60-80%	*13* 0.55	*14* 0.85	*15* 0.64	*16* 0.29
40-60%	*9* 0.60	*10* 0.92	*11* 0.69	*12* 0.31
20-40%	*5* 0.65	*6* **1.00**	*7* 0.75	*8* 0.34
0-20%	*1* 0.49	*2* 0.75	*3* 0.56	*4* 0.26
Age	0-20	20-40	40-60	60-80

Yet another method (geometric means) yielded weights that were more extreme again than those shown above. However, although the specific weights may vary according to the measure of central tendency chosen and the particular method of aggregation, the *general pattern* of weights dropping as we move in any direction away from cell 6 is robust to these differences.

The 'profile' tests

Finally, we turn to the tests for independence of the profile in which a cell appears. Recall that Figure 2 illustrated the alternative ways in which we might have presented the QALY gains embodied in the cells in the QALY grid. Profile tests were carried out on three different cells; 7, 11 and 14 and the total lifetime untreated QALYs under each representation given in Table 2. On this basis, we argued that a 'fair innings' type argument – in which the total lifetime QALYs was relevant- would always predict that *more* weight would be given to the 'standard' profile -S- than to profile Y (as the total lifetime QALYs is less in S than Y), and *less* weight to profile S than profile X. Finally, profile S would be given less weight than Z in the cases of cells 7 & 11, but S would be given more weight than Z in the case of cell 14.

The results are given in Table 7 where values greater than one indicate that the gain was valued more highly when embedded in the 'standard' profile –S- than in the alternative and vice versa. For example, the first column of Table 7 gives the results of comparing the standard profile -S- with profile X (see Figure 2). As the values are each greater than one, there appears, if anything, to be a tendency to favour those with higher lifetime health which is the opposite of what a 'fair innings' argument would suggest. Column two shows that for two of the three cells - 7 and 14 - the standard profile was valued less highly than profile Y which had lower lifetime health - a finding which does fit with a 'fair innings' type argument, but the opposite was true in the case of cell 11. The results of the S vs Z comparison also appear to contradict the 'fair innings' argument.

To sum up, it is difficult to detect any clear pattern in responses to the profile tests conducted here, but, if anything, the data appear to contradict the 'fair innings' argument, which is puzzling.

Table 7: Results of the 'profile' tests

	S vs X	**S vs Y**	**S vs Z**
Cell 7	1.33	0.74	1.24
Cell 11	1.22	1.36	1.08
Cell 14	1.06	0.55	0.78

DISCUSSION

We reported here the results of a computer assisted PTO valuation study carried out in a sample of 587 members of the public in the UK. The results showed that, when asked questions of the type posed here, respondents *do* differentiate between ‘types’ of QALYs and are willing to trade off (sometimes considerable) numbers of patients treated in order to prioritise according to age and severity of illness. The data shows a *general* tendency to give more weight to younger patients and to those in poorer health, but the pattern is not a simple one. In particular, less weight is given to the *youngest* patients and to those in *poorest* health, reversing the pattern found elsewhere in the data. A number of important methodological issues remain to be resolved before these results may be used to inform policy.

First there are well documented methodological issues surrounding the PTO method itself. There is evidence to suggest that respondents are susceptible to anchoring effects and ‘ratio bias phenomenon’- in which subjects focus on the *absolute* numbers in the numerator and ignore the denominator (Pinto- Prades et al, 2006). It is important to note, however, that such biases affect *all* elicitation procedures and are not specific to the PTO. Further, it has been shown that PTO responses do not satisfy the transitivity principle and that responses are contingent upon the absolute numbers treated in each group (Dolan and Tsuchiya, 2003). Whilst our method of deriving the weights in the QALY grid does not itself rely on transitivity, the implications of these findings are nevertheless worrying as they suggest that starting with a different number of patients in each group (say 1000, rather than 100) may have yielded different results.

Second, the partitions used in the QALY grid are obviously fairly crude and it is possible that patterns are hidden within the age ranges used here. This seems particularly likely at the lower end of the age range, where, for example, a 10 year old may well be regarded differently than either a 5 year old or a 15 year old. Future work on the QALY grid approach ought to consider the use of a more finely partitioned grid.

A third point- related closely to the second- is that gains of the magnitude of 4 QALYs are actually fairly large and more than many interventions could hope to achieve. Applying the weights derived here when the QALY gains are more marginal requires the additional assumption that the general pattern will hold when the overall gain is much lower. Again this is an issue for future research.

Certain of our results may be surprising to readers but the qualitative data from the preparatory stages offers some explanation for the pattern of results uncovered here. In the focus group discussions, age was found to be important to focus group participants for a variety of reasons including life expectancy, which would explain the *general* tendency to favour the young over the old. Age was also, however, used as a 'proxy' for a range of other characteristics including economic activity and the existence of dependants. One possible explanation of the tendency to favour 20 to 40 year olds over younger patients, therefore, is that such considerations outweigh the impact of life expectancy around those ages.

Respondents in the preparatory focus groups also expressed the view that an improvement in quality of life for people in poor health would be more important than an identical improvement in QoL for people in relatively good health. Again, this would seem to be consistent with the *general* tendency to favour those in more severe health states initially. Some concern was also expressed that prolonging life in a really bad health state was not desirable. One possible explanation of why respondents favoured a move from 20% to 40% health over one from 0% to 20% is that 20% was still considered to be not worth living. This raises issues about how the QoL axis - which necessarily relied on a 'VAS-like' representation in order to put the concept across to respondents-was being interpreted, particularly towards the lower end.

It is also possible that tendency for the weights to 'dip' towards the top right hand corner of the grid may be linked to how respondents were interpreting the QoL scale. Whilst the diagrammatic representation of QoL over the patient's lifetime made it clear that the quality of life related to the relevant stage of the patient's life, such that 100% health at the age of 70 meant in good health *for a 70 year old*, it may be that certain respondents were considering QoL as an *absolute* measure. If this were the case, then it might have been considered that 80% health, for example, was the best that old people could reasonably expect to achieve and little was to be achieved by trying to improve their health beyond this. If there were a certain degree of this going on when respondents were answering the matching questions, this raises the possibility that the weights estimated here are perhaps penalising the ‘healthy old’ rather too heavily.

On the other hand, the weights we derived here are, if anything, *more* conservative than those uncovered previously. Recall that both Johannesson and Johanesson (1997) and Nord and colleagues (1996) derived age weights between the most and least preferred groups of the magnitude of around 10:1. It is possible that differences in the magnitudes of the weights derived are due to differences in methods used to aggregate the data as we have shown that different methods yield different results.

We look forward to hearing the views of HESG members on the work presented here. We conclude by offering some potential points for discussion. If we were to take a (somewhat giant) leap of faith and suppose that we did have a robust set of weights of the kind derived here;

- How would such a set of weights be operationalised in informing policy?
- Would valuing some QALYs three times (or more) more highly than others be ‘acceptable’ to a) NICE, b) the public?
- Are most actual decisions concentrated in certain parts of the grid- i.e. the top right hand corner?

References

Baker R, Donaldson C, Lancsar E, Ryan M, Wildman J, Deriving QALY weights through discrete choice experiments: challenges and preliminary results, paper presented to HESG, Norwich 2008.

Busschbach JJV, Hessing DJ, De Charro FT. The utility of health at different stages in life: A quantitative approach. Social Science and Medicine. 1993;37(2):153-8.

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

Dolan P, Tsuchiya A, The person trade-off and the transitivity principle: an example from preferences over age weighting, Health Economics Letters, Health Economics, 2003, 12, 505-510 .

Dolan P, Tsuchiya A. Health priorities and public preferences: the relative importance of past health experience and future health prospects. Journal of Health Economics. 2005;24 703-14.

Johannesson M, Johansson P-O. 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. 1996;37:153-61.

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.

Loomes G (on behalf of the SVQ team) The feasibility- or otherwise- of estimating the monetary value of a QALY, paper presented at HESG, Norwich Jan 2008

Mason H, Marshall A, Jones-Lee M. Donaldson C, Estimating a monetary value of a QALY form existing UK values of prevented fatalities and serious injuries, paper presented to HESG, Newcastle 2005.

Nord, E, Street A, Richardson, J, Huhse, H, Singer, P ‘The significance of age and duration effect in social evaluation of health care’, *Health Care Analysis*, 1996; 4: 103-111.

Robinson A, Covey J, Jones-Lee M, Loomes G, Comparing the results of face-to-face and web-based PTO exercises, paper presented to HESG, Aberdeen 2008.

Ubel, P.A, Spranca M.D, DeKay, M.L, Hershey J.C, Asch D.A, Public preferences for prevention versus cure: What if an ounce of prevention is worth only an ounce of cure? Med Decis Making 1998; 18 (2) 141-148