Effect of age on health state valuations

Paul Dolan

Sheffield Health Economics Group and Department of Economics, University of Sheffield, Sheffield, UK

***Objectives*: Valuations for health states defined by the EQ-5D measure of health-related quality of life, which were elicited using the time trade-off method, from a large sample of the UK general population have previously been shown to be related to the age of the respondent. This paper presents population valuation tariffs for all EQ-5D health states for those aged 18–59 and those aged 60 and over.**

***Methods*: Health state valuations were elicited from a representative sample of over 3000 members of the UK general population using one-to-one interviews. The same regression technique which had previously been used to estimate the EQ-5D tariff for the whole population is used in the current analysis. Additional variables were created which were the product of the original variables and the dummy attached to the age group of the respondent. An earlier qualitative study had suggested that some of the observed differences in valuations might have resulted from more older respondents doubting the plausibility of the 'worse than dead' scenario. Therefore, the model was estimated both when the scores of those aged 60 and over were unmodified and when they were modified to account for this possibility.**

***Results*: The tariff based on values from those aged 60 and over, was considerably lower than that based on values from those aged 18–59. When changes between health states were calculated, the difference could be as much as 0.15 on a scale between −1 and 1. However, when the scores of older respondents were adjusted, on the assumption that their lower values for states rated as worse than dead were due to an experimental artefact, the differences between the valuation tariffs of the two age groups greatly diminished.**

***Conclusion*: These results suggest that whose values are used may have an important effect on health state valuations, and in turn, therefore, on resource allocation decisions, but more research is needed to understand the source of the differences between the values given by different population subgroups.**

Journal of Health Services Research & Policy Vol 5 No 1, 2000: 17–21

Introduction

There is little doubt that measures of health-related quality of life (HRQoL) are increasingly being used to evaluate the outcomes of health care interventions. Against this background, a large-scale study was conducted which elicited health state valuations from a representative sample of over 3000 members of the UK general public.[1] Even if it is accepted that the values of the general public should be given some weight in resource allocation decisions, a number of important theoretical and empirical issues are raised by datasets of this kind. For example, in a recent paper in this journal, it was shown that how valuations are aggregated could have a significant effect on subsequent resource allocation decisions.[2] This paper considers the question of whether different consequences might result from using the valuations of different groups of people.

Paul Dolan DPhil, Reader in Health Economics, Sheffield Health Economics Group and Department of Economics, University of Sheffield, 30 Regent Street, Sheffield S1 4DA, UK.

If it can be shown that different groups have similar preferences, then it may not matter whose valuations are used. However, if major differences exist between groups of raters, then this poses a question for policy-makers as to whose values are to be given the greatest weight in assessing the outcomes of health care. The views of the whole population are likely to be considered the most relevant when comparing interventions that affect many different population subgroups. When a comparison is being made between treatments for a particular subgroup, there are strong grounds for using the values of that subgroup.

In the general population study (referred to hereafter as the 'main study'), valuations were elicited for health states defined by the EQ-5D measure of health-related quality of life[3] using the time trade-off (TTO) method. The results of this study suggest that TTO valuations are primarily affected by the age of the respondent.[4] On average, valuations, particularly for more severe states, increase slowly from the age of 18 to about 40, begin to fall slowly from about 40 to 60 years and then fall sharply in later years. These results run counter to those of a number of earlier studies which had suggested that variation among population subgroups was not explained by the different demographic characteristics of respondents.[5] However, the large sample relative to some of the earlier studies may explain why differences between population subgroups were detected in the main study and not elsewhere.

In an attempt to gain a better understanding of the causes for such differences, a qualitative follow-up study was conducted in which 45 respondents who had taken part in the main study were re-interviewed and asked to 'think aloud' as they completed the interview.[6] Lower valuations from older respondents could have resulted from some of them perceiving that they were likely to face a shorter life-expectancy than the 10 years specified in the TTO question, but there was no evidence of this from the qualitative data. This suggests that any differences between the age groups in the values they give to states rated as better than dead are 'real' (i.e. representing genuine differences in preferences) rather than 'artefactual' (i.e. due to bias introduced by the elicitation procedure).

However, whilst only 10% of respondents in the 18–59-year age group questioned the plausibility of the scenario for states rated as worse than dead (for details of this scenario, see Methods, below), 50% of respondents in the 60 or over age group considered it to be implausible. Those questioning the plausibility of the 'worse than dead' scenario had significantly lower values for states rated as worse than dead, suggesting that these respondents were somehow re-interpreting the scenario in order to provide a response. To control fully for this possible artefactual effect in the data of the main study, the distribution of values for each state was first established for the younger and older respondents separately. The values for states rated as worse than dead by older respondents were then substituted by a value drawn from the equivalent position in the distribution of values recorded by younger respondents (determined by converting the values for states worse than dead into standardised *z*-scores). Although reduced, the effect of age was still pronounced.

This means that the set of 243 EQ-5D health state valuations derived for the *whole* community from the results of the main study[7] might not be considered appropriate in all contexts. For example, if the population affected by a particular policy is exclusively elderly people, then it might be considered more appropriate to give greater weight to the preferences of such people. Thus, a separate set of valuations (referred to as an EQ-5D tariff), based on the valuations of respondents over the age of 60, might be required. This paper uses data from the main study to estimate valuation tariffs for all EQ-5D health states based on the age group of the respondents. In addition, it looks at the effect of modifying the values of older respondents in the way described above.

Methods

Since the main study has been detailed elsewhere,[2,7] only a brief summary is provided here. Using one-to-one interviews conducted by trained interviewers, the aim was to elicit health state valuations from a representative sample of the UK general population. The overall sample comprised 3395 respondents, but to enable modelling of the data at the individual level only the 2997 respondents with complete data were included in the analysis. Each respondent was asked to value 12 EQ-5D health states (drawn from a wider subset of 42 states) using the TTO method.

For a state that was regarded as better than being dead, the respondent was led by a process of 'bracketing' to select a length of time (x) in full health that s/he considered to be equivalent to 10 years in that state. For a state that was regarded as worse than being dead, the respondent was led by a similar process to select a length of time ($10-x$) in that state followed by x years in full health that s/he considered to be equivalent to dying immediately. If full health and dead are assigned scores of 1 and 0 respectively, then TTO scores for states rated as better than dead are given by the formula $x/10$. Valuations for states rated as worse than dead were transformed using the formula $[(x/10)-1]$, so that the health state valuations lay in the range $[-1, 1]$.[8]

The same generalised least-squares regression technique employed in estimating the EQ-5D tariff for the whole population[6] was used for the current analysis. The analysis is based on individual level data in which the functional form is additive. The dependent variable is defined as $(1-S)$, where S is the value given to a particular health state. Besides the intercept, the specification of the remaining independent variables derives from the ordinal nature of the EQ-5D descriptive system. The model contains two dummy variables for each dimension: one to represent the (assumed equal) move between levels within health states and one to represent the (possibly different) move from level 2 to level 3.

To differentiate the tariffs generated, and to enhance their practical and policy relevance, respondents were split into two age groups: those aged 18–59 and those aged 60 and over. Since the effect of age may differ across levels and/or dimensions of health, additional variables were created which were the product of the original 12 dummies and the dummy attached to the age group of the respondent. The model was estimated both when the scores of those aged 60 and over were unmodified and when they were modified to account for the possibility that different valuations might have arisen partly because older respondents considered the 'worse than dead' scenario to be implausible.

Since each respondent valued 12 states, it is likely that these 12 scores were related to one another. To account for this, a random effects (RE) model was used with an overall intercept and an error term with two components: $e_{il}+u_i$. The e_{il} is the traditional error term unique to each observation. The u_i is an error term representing the extent to which the intercept of the ith respondent differs from the overall intercept. The models were tested for general heteroskedasticity in the error terms

Table 1 Decrements for each level on each dimension

Dimension	18–59 years	≥60 years	≥60 years, modified
Constant representing any move away from full health	0.076	0.090	0.090
Mobility			
1. No problems in walking about			
2. Some problems in walking about	0.067	0.075	0.059
3. Confined to bed	0.304	0.345	0.313
Self-care			
1. No problems with self-care			
2. Some problems washing/dressing self	0.083	0.150	0.099
3. Unable to wash/dress self	0.199	0.248	0.201
Usual activities			
1. No problems performing usual activities			
2. Some problems performing usual activities	0.031	0.049	0.040
3. Unable to perform usual activities	0.104	0.069	0.092
Pain/discomfort			
1. No pain or discomfort			
2. Moderate pain or discomfort	0.117	0.131	0.115
3. Extreme pain or discomfort	0.391	0.358	0.389
Anxiety/depression			
1. Not anxious or depressed			
2. Moderately anxious or depressed	0.072	0.066	0.057
3. Extremely anxious or depressed	0.246	0.208	0.258
Decrement when any dimension is at level 3	0.236	0.346	0.255

Note: Higher numbers correspond to a greater decrement in health state value since they are subtracted from 1 (i.e. the score for full health).

by regressing the value of the squared residual term on the predicted values. The modelling was carried out using the LIMDEP statistical package.[9]

Results

The R^2 for both models (i.e. when the scores of those aged 60 and over were unmodified and modified) was 0.47, which is high given the type of data analysed here. In addition, the estimated values from the models came very close to the actual (mean) values of the 42 states directly valued in the study. For the 18–59-year age group, the difference between the estimated and actual values was within 0.05 for 67% of the states and within 0.10 for 95% of the states. The corresponding figures for those aged 60 or over were 57% and 88% when their scores were unmodified and 60% and 93% when they were modified. However, both models suffered from general heteroskedasticity. This is difficult to overcome given the categorical nature of the independent variables and, in the context of this analysis, it is likely to have resulted in inefficient rather than biased parameter estimates.

Rather than present the coefficients for each variable in the regression equation, Table 1 presents the coefficients for each age group in terms of their effect on each dimension and level within dimension. This enables health state valuations to be readily calculable and facilitates a comparison of the differential effects of age on the decrements attached to the dimensions, and to levels within the dimensions. From Table 1, it can be seen that the constant for both age groups is highly significant, suggesting that *any* move away from full health is associated with a substantial loss of utility. The coefficient on 'any level 3' highlights the aversion that respondents (particularly those in the older age group when their scores are unmodified) have to being at the lowest level on any of the dimensions.

With regard to attitudes towards different dimensions, it appears that all respondents are most concerned about being in pain and least concerned about having problems with their usual activities. One of the biggest differences between the under- and over-60s is in their attitude towards self-care: when the scores of those aged 60 or over are unmodified, some problems with washing or dressing are considered to be much worse by older respondents than by younger ones. Overall, it would appear that older respondents consider all levels and all types of dysfunction to be worse than do younger respondents.

When estimated values are calculated for the full set of EQ-5D health states, it is found that older respondents (when their scores are unmodified) have estimated values that are lower than those of younger respondents by at least 0.05 for 93% of the EQ-5D states and lower by at least 0.10 for 75% of the states. When the scores of the older respondents are modified, the differences between the age groups are less dramatic: no differences are in excess of 0.10 and for only 22% are the estimated values of older respondents more than 0.05 lower than those of younger respondents.

This indicates that the extent to which the benefit from various treatments differs according to which set of values is used will depend crucially on whether the

Table 2 Illustrative example of differences between tariffs by age group: mean health state values for patients in each group

Group[a]	18–59 years	⩾60 years	⩾60 years modified
1	0.850	0.816	0.836
2	0.718	0.648	0.704
3	0.291	0.130	0.256
4	0.079	−0.069	0.043
5	−0.182	−0.314	−0.221

[a]Definition of groups: group 1, five states with only one dimension at level 2; group 2, 26 states with any combination of levels 1 and 2; group 3, 80 states with one dimension at level 3; group 4, 80 states with two dimensions at level 3; group 5, 51 states with three or more dimensions at level 3.

valuations of those aged 60 and over are modified or not. It will also depend on the initial and final health states. To illustrate this, suppose that all patients with a particular condition fall into one of five groups of health states. These are identical to the groups used in the earlier paper[2] and are shown in Table 2. This table shows the mean health state value associated with patients in each group according to whether the values of those aged 18–59 are used or whether the (unmodified or modified) values of those aged 60 or over are used.

Comparing the values of those aged 18–59 with the unmodified values of those aged 60 or over, it can be seen that moving patients up one group (including the move from group 1 to full health) in four out of five cases would yield about the same amount of benefit irrespective of whose values are used. The exception is the move from group 3 to group 2, which yields about 0.10 more benefit if the valuations of older rather than younger respondents are used. Any patients in groups 3, 4 or 5 who could be restored to full health would yield about 0.15 more benefit if the valuations of older respondents are used. When the valuations of those aged 60 and over are modified, however, the differences between the age groups are dramatically reduced. Indeed, the difference in benefit from using the values of younger as opposed to older respondents is less than 0.05 for all movements between any two groups.

Discussion

This paper has provided insights into the relative weights that different age groups attach to the different dimensions of health in the EQ-5D classification system. Whilst some problems with self-care are seen as being considerably worse by older respondents, it would appear that older people generally consider all levels and all types of dysfunction to be worse than younger people do. When valuations for all EQ-5D health states are calculated, older respondents have estimated values that are lower than those of younger respondents by at least 0.10 for 75% of the states. In a decision-making context, *changes* in valuations are important.[10] Depending on the initial and final health states, the differences between the inferred change in values of older and younger respondents can be as large as 0.15 on a scale from −1 to +1.

However, there is some evidence that the much larger negative values elicited from older people in the main study might be artefactual.[5] In response, a tariff has been estimated for those aged 60 and over on data modified to control fully for this potential source of bias. This has the effect of dramatically reducing the differences between the two age groups. Modifying valuations in this way assumes that *all* the observed differences between the age groups in their valuations for states rated as 'worse than dead' are the result of older respondents considering the 'worse than dead' scenario to be implausible (although it does still treat the differences in the *likelihood* of a state being rated as 'worse than dead' as genuine). A more realistic conclusion may be that *some* of the difference is artefactual and *some* is probably due to older people genuinely being more concerned about helplessness and dependency than their younger counterparts are.

Therefore, whether using the values of older or younger people makes any difference to decisions depends crucially on the extent to which the differences in scores (for *any* health state) are considered to be a genuine reflection of different preferences or the result of biases introduced by the procedures used to elicit them. It is difficult to answer this central question at this stage with any real degree of confidence. An important avenue for future research in this area, and in the measurement of HRQoL generally, will be to gather qualitative data from respondents. This will provide insights into how they think in order to arrive at their responses, thus enabling researchers to gain a better understanding of *why* valuations differ, in addition to *how* they differ. With such information, policy-makers will be better placed to make informed decisions as to the appropriateness of using one set of values rather than another.

Of course, the key question for real-world decision-making is whether using one set of values as opposed to another makes any difference in practice. At the moment, there is little evidence in the literature regarding the size difference likely to alter decisions.[11] However, now that tariffs for different age groups have been generated, it should be possible to identify those resource allocations decisions that could be affected by the choice of valuations. Ultimately, the decision as to whose values should be used is a political not a scientific one, but empirical evidence of the kind presented in this paper can highlight the consequences of the choices made.

Acknowledgements

The author would like to thank Alan Williams, one of the editors and two anonymous referees for their helpful comments on earlier drafts of this paper. Thanks are also due to all those at SCPR for their help in making the survey such a success.

References

1. Gudex C, Dolan P, Kind P, Williams A. Valuing health states: interviews with the general public. European Journal of Public Health 1997; 7: 441–448

2. Dolan P. Aggregating health state valuations. Journal of Health Services Research & Policy 1997; 2: 160–165
3. Brooks R. EuroQol: the current state of play. Health Policy 1996; 37: 53–72
4. Dolan P, Gudex C, Kind P, Williams A. Valuing health states: a comparison of methods. Journal of Health Economics 1996; 15: 209–231
5. Froberg DG, Kane RL. Methodology for measuring health state preferences III: population and context effects. Journal of Clinical Epidemiclogy 1989; 42: 585–592
6. Robinson A, Dolan P, Williams A. Valuing health states using the VAS and TTO methods: what lies behind the numbers? Social Science and Medicine 1997; 8: 1289–1297
7. Dolan P. Modelling valuations for EuroQol health states. Medical Care 1997; 11: 1095–1108
8. Patrick DL, Starks HE, Cain KC, Uhlmann RF, Pearlman RA. Measuring preferences for health states worse than death. Medical Decision Making 1994; 14: 9–18
9. Greene WH. LIMDEP version 6.0: user's manual and reference guide. New York: Econometric Software Inc., 1992
10. Fitzpatrick R. A pragmatic defence of health status measures. Health Care Analysis 1996; 4: 265–272
11. O'Brien BJ, Drummond MF. Statistical versus quantitative significance in the socioeconomic evaluation of medicines. PharmacoEconomics 1994; 5: 389–398