EUROPEAN JOURNAL OF PUBLIC HEALTH 1997; 7: 441–448

ORIGINAL ARTICLES

Valuing health states

Interviews with the general public

CLAIRE GUDEX, PAUL H. DOLAN, PAUL KIND, ROGER THOMAS, ALAN H. WILLIAMS *

The objective of this study was to develop methods to elicit the general public's views on the comparative subjective value of different states of health. The resulting valuations form the basis for a set of British social preferences for use in clinical and economic evaluation of health care. The methods have proved extremely successful in generating complete data of high quality. Since the approach used is generally applicable for use in other national surveys, it is described here to encourage others to take the opportunity to generate comparable sets of social preferences. Face to face interviews, lasting approximately 1h, were conducted in the respondents' own homes. There were 3,395 interviews achieved (a response rate of 64%) and the sample was representative of the British general population in terms of age, sex, education, social class and geographical location. Each respondent valued 15 EuroQol health states using ranking, visual analogue scale (VAS) and time trade-off (TTO) methods, with 45 states being valued in all. Two hundred and twenty-one reinterviews were conducted after an average time of 10 weeks. Several methodological issues had to be confronted during the course of the study. These included the structure and format of the interview, the choice of health states to be valued, the determination of the sample size required, the achievement of a representative sample of the British adult population, interviewer training, data processing and data quality. Since few valuation studies have been undertaken on such a large scale, much time and effort was spent in resolving these issues. The methods used are recommended to others considering similar surveys.

Key words: health state valuations, social preferences, interview methods

The general public's views on resource allocation in health care are relevant to all concerned with health care services.[1] Their opinions can be sought on a range of issues such as the resources devoted to health care, the processes of health care and the outcomes from health care. In this study, outcome is the focus of interest but within a carefully defined context. As part of a priority setting exercise, the views of the general public are crucial in judging the comparative subjective value of different potential health states. The objective of this study was thus to develop methods to elicit, from a representative sample of the British general public, valuations for states of health which could act as a set of social preferences.

Previous studies have examined the nature of peoples' preferences for health states but none proved adequate for the present purpose. Many used small samples to investigate particular aspects of the valuation process[2–5] or the comparison of different methods[4,6–8] or elicited valuations from patients.[4,6,9–11] The valuations elicited in these studies, while invaluable to the further understanding of health state preferences, are unlikely to be representative of the views of the general public as a whole.

* C. Gudex[1], P.H. Dolan[1], P. Kind[1], R Thomas[2], A.H. Williams[1]
1 Centre for Health Economics, University of York, York, UK
2 Social & Community Planning Research, London, UK
Correspondence: Dr C. Gudex, MBChB, MPH, Tordenskjoldsgade 23, 3th, DK–1055 Copenhagen K, Denmark

Many of the studies which have sought to elicit health state preferences from the general population have been conducted in North America.[12–16] Efforts have been made to incorporate such weights into decision-making processes,[17] but the studies would need to be replicated to confirm the use of North American values in a British population or it would be necessary to generate new values if preferences were found to be different. At the time of its publication, the valuation matrix associated with the Rosser Classification of Illness States[18] was the only British-based set of health state values. However, the valuations were derived from a convenience sample of 70 and when similar methods were applied in a larger general population,[19] very different valuations to the initial set were found.

In order to establish a health state descriptive system for use in the current study, a survey of lay concepts of health[20] was undertaken to investigate the dimensions of health-related quality of life that would be relevant for a British population. It was concluded that the EuroQol Classification,[21] a standardized non-disease-specific instrument for describing and valuing health-related quality of life, would be the best choice, as it covered the major dimensions but was still parsimonious – with 5 dimensions, it has more than Rosser's 2, but considerably fewer than the North American systems, generating 'only' 243 potential health states. The use by the EuroQol Group of a visual analogue scale (VAS) approach to elicit prefer-

ences for health states in postal surveys[22–24] has generated criticism,[25–28] regarding the validity of the VAS approach and its ability to generate an interval scale, valuation of whole health states rather than individual components and low response rates. However, much of this debate centres on the methodology of health state preferences rather than the EuroQol classification itself, which is currently being used in a wide range of clinical areas as a measure of health status, with or without a valuation system.

Regarding methods for valuing health states, none can be regarded as superior to all others, since no 1 set of values constitutes the 'gold' standard. In these circumstances, the selection of a method for use in a population survey will depend upon many factors including acceptability, feasibility, reliability and cost, as well as methodological issues. An earlier study comparing time trade-off (TTO) and standard gamble[29] led to the choice of TTO as the method most likely to generate the best quality data in a general population survey. The VAS method was also incorporated to allow comparison between a non-choice-based method and a choice method (TTO). The resulting general population survey is unique in its use of both TTO and VAS to value EuroQol health states through face to face interviews.

This paper is aimed at several audiences: i) those wishing to conduct similar surveys and thus who are interested in our experiences and recommendations, ii) those wishing to compare the data from our methods with data from their own and iii) those wishing for more detail about our methods in order to assess the robustness of our data. The main issues that had to be confronted are discussed in turn, including the structure of the interview, the choice of health states, sample size and selection, interviewer training, data processing and data quality.

STRUCTURE OF INTERVIEW

Two developmental pilot studies were conducted and these generated crucial results.[30] Each involved interviews with a separate group of 40 respondents. The first, with 4 interviewers (all of whom had been involved in the previous valuation survey), tested a modification of the traditional 'ping-pong' presentation of the TTO task[8] and investigated the maximum number of states that could be evaluated by each respondent, assuming a maximum interview length of approximately 1 h. There was unanimous support among the interviewers for changes in the script, moving away from the line-by-line and often repetitive instructions, to a more dynamic script with just the basic TTO 'rules' on a single page for each state. This format made it possible for each respondent to value 13 states on TTO, but required a more detailed understanding of the TTO procedure on the part of interviewers.

The second pilot, with 4 new interviewers, tested changes to the TTO script to address the lack of discrimination between states, particularly at the mild end of the severity range. These changes worked well and were accepted into the final protocol. The script also took a format more familiar to the interviewers, allowing questions to be followed sequentially. The involvement of 4 new interviewers emphasized the need for a good briefing, with each interviewer having the opportunity to conduct a practice interview with a dummy respondent familiar with the objective of the survey.

The final interview contained 5 main elements: self-reported health, ranking of states, VAS rating of states, TTO rating of states and personal background data. Further details, including materials required and verbatim instructions, are available.[31,32]

- Self-reported health

Self-reported health, using the EuroQol classification system and a 20 cm visual analogue scale, was elicited partly to collect the information itself to see whether current health status affected valuations, but also to familiarize respondents with the descriptive system and the VAS that they would use later in the interview.

- Ranking of states

The ranking of states was intended primarily as a warm-up exercise to familiarize the respondent both with the range of states to be assessed and with the subjective nature of the interview. Health state descriptions printed on cards were handed to the respondent in random order, to be placed in rank order from best to worst. The respondent was then handed a further card, 'immediate death', which was to be located within the ranked set. Ranking was often 1 of the longer sections of the interview, taking approximately 20 min. Interviewers uniformly took the view that time spent on ranking was a good investment, leading to a quicker understanding and completion of the later valuation tasks.

- VAS rating of states

With the cards still set out in the rank order chosen, the respondent was asked to indicate where, on a 20 cm visual analogue scale with end-points of 100 (best imaginable health state) and 0 (worst imaginable health state), they would rate each state. As on the ranking, it was stated that each state would last for 10 years without change, followed by death. All cards remained on the table and respondents were permitted to change the order if they wished.

- TTO rating of states

The TTO exercise involved the use of a specially designed board, one side of which was used to value states regarded by the respondent as better than death and the other for states worse than death. In the former case the respondents were led by a process of 'bracketing' to select a length of time in full health (state 11111) that they regarded as being equivalent to 10 years in a dysfunctional state: the shorter this length of time, the worse must be the target state and the lower its value. Respondents were given an opportunity to refuse to trade-off any length of time in order to improve its quality. In the case of states worse than death, the choice was between dying immediately and spending a length of time (x) in the target state followed by (10-x) years in full health. The more time required in full health to compensate for quite a short time

in the target state, the worse must be the target state. Health states (excluding 'full health' and 'immediate death') were valued one by one, using whichever side of the board was appropriate.

- Personal background data

At the end of the interview, background data were collected to assess the representativeness of the achieved sample and to enable any systematic influences on health state valuations to be tested. Respondents were also asked if they would be willing to be reinterviewed at a later date. With the addition of some questions relating to recent experience of illness, the second interview was identical to the first, with the same health states being used.

CHOICE OF HEALTH STATES

The EuroQol classification (*box 1*) generates 243 theoretically possible health states. To these need to be added 2 further states, 'unconscious' and 'immediate death', which are not technically defined by the EuroQol dimensions and thus need to be valued directly. It was clear, however, that only a subset of the 245 states could be directly valued in this study. Factorial designs, based on balanced incomplete blocks, limit the investigation of individual differences, a key feature of this study. In choosing the states for direct valuation, there were several considerations in mind. Three states had to be included: 'unconscious' and 'immediate death' because they could not be estimated from valuations given to other states and the state 11111 ('full health', *box 1*) because it was essential to the rescaling (onto a 0–1 scale) of the VAS data. The remaining states should be widely spread over the valuation space in terms of severity (as indicated by previous valuation data), should include all plausible combinations of levels across each of the 5 EuroQol dimensions to allow testing for any significant interaction effects, should appear plausible to respondents and should be as close as possible to the set of states used in an earlier Finnish study.[33]

Finally, 45 states were chosen thus.

- Group 1, valued by every respondent, included states 11111, 33333, 'unconscious' and 'immediate death'. The first 2 ensured a common frame of reference for all respondents.
- Group 2 included the mildest of the EuroQol states: 11112, 11121, 11211, 12111 and 21111. Each respondent rated 2 of these.
- Groups 3 to 5 aimed at balancing, at an individual level, the remaining states between 'mild' (group 3), 'moderate' (group 4) and 'severe' (group 5). Each respondent valued a random selection of 3 states out of each group of 12.

Computer software was specially written to select 2 states from group 2 and 3 from each of groups 3–5. The programme applied sampling without replacement and chose states at random from each of the groups independently. There were 6,080 unique combinations of 11 states produced which, when combined with the 4 states common to all respondents, generated lists of 15 states per interview. Appropriate numbers of these lists ('card allocation sheets') were provided to each interviewer who was instructed to use 1 sheet per interview. In the event, slight discrepancies between the intended and actual coverage of states did occur. This was firstly because the interviewers did not need to use all the card allocation sheets assigned to them (and, thus, the different states within each group were not rated exactly the same number of times) and, secondly, there were 41 cases where the interviewer used the wrong state(s). In 15 of these only 1 state was used incorrectly, while in another case 2 states were chosen incorrectly. In the other 25 cases *all* the states were incorrect indicating that the interviewer had used the wrong card allocation sheet. The net effect of these errors within and between groups of states was small since in the majority of cases a state was replaced by another belonging to the same group. Overall there were 4 fewer states from group 3 than intended, 3 more states from group 4 and 1 more state from group 5.

SAMPLE SIZE

Two considerations were dominant here. Sufficient observations were required firstly to detect small differences between the valuations given to different states and, secondly, to detect significant differences in valuations for the same state between different subgroups of the population.

Box 1 The EuroQol descriptive system

Mobility
1 No problems walking about
2 Some problems walking about
3 Confined to bed

Self-care
1 No problems with self-care
2 Some problems washing or dressing self
3 Unable to wash or dress self

Usual activities
1 No problems with performing usual activities (e.g. work, study, housework, family or leisure activities)
2 Some problems with performing usual activities
3 Unable to perform usual activities

Pain/discomfort
1 No pain or discomfort
2 Moderate pain or discomfort
3 Extreme pain or discomfort

Anxiety/depression
1 Not anxious or depressed
2 Moderately anxious or depressed
3 Extremely anxious or depressed

Note: For convenience each composite health state has a five digit code number relating to the relevant level of each dimension, with the dimensions always listed in the order given above.
Thus 11232 means:
1 No problems walking about
1 No problems with self-care
2 Some problems with performing usual activities
3 Extreme pain or discomfort
2 Moderately anxious or depressed

The sample size needed to detect a difference between 2 means is determined by the standard formula

$$N = 2\,SD^2/DIFF^2 \times (\S\,(SIG + POWER))^2$$

where N is the size of the sample, SD is the standard deviation, DIFF is the absolute size of the difference to be detected, § (from standard normal tables) is a function of the desired power and the desired significance level, SIG is the desired significance level and POWER is the desired power of the test.

Here, these parameters were POWER = 80%, SIG = 0.05/0.01, DIFF = 0.025/0.05/0.10 and SD = 0.35 (based on the mean SD for valuations obtained from pilot studies). The formula assumes a normal distribution of data however. Previous experience showed that such distributions were unlikely to occur and therefore non-parametric tests would be used. To take account of the lower efficiency of these tests, the power calculations were adjusted by 1/0.95.

Comparison of differences in valuations given to different states

In this study the smallest difference that can be expressed on the TTO is 0.025, which is equivalent to 3 months. To detect such small differences at the 5% significance level would require 3,235 valuations for each state *(table 1)*. The complication here was that it was impossible for each respondent to value all states, so the number of valuations obtained for most states would be much less than the survey population size, e.g. 36 of the chosen states would be valued by only 25% of respondents. A survey size of 3,235 would allow, at the 5% level, a 0.05 difference to be detected between these 36 states and a 0.025 difference to be detected between states that had been valued by all respondents.

Comparison of differences in valuations given to the same state by different subgroups of the population

The complication here was that it could not be assumed that the subgroups of interest would be of equal size. Although the population could be divided into 2 age groups, a comparison of the elderly versus the non-elderly or employed versus non-employed, for example, may generate subgroups comprising of only approximately 25% of the survey population. A survey size of 3,235 would generate a minimum of 809 valuations for each state and, thus, a minimum of approximately 200 per smallest subgroup. This would allow, at the 5% level, a 0.10 difference to be detected between valuations given by different subgroups.

SELECTION OF SAMPLE

Anticipating a response rate of approximately 55%, 6,080 addresses in England, Scotland and Wales were selected from the Post Code Address File, using a 3-stage design. Firstly, 80 postcode sectors were selected, with a probability proportionate to the number of addresses in each sector (small sectors being grouped with neighbouring ones to give a minimum sector size of 1,000 addresses). Before selection, sectors were stratified by region (14 regional health authority areas of England plus Wales and Scotland), socioeconomic group (using 2 bands) and, within each of these 32 bands, by population density. Secondly, 76 addresses were selected systematically from throughout each of the 80 sampled postcode sectors and the resulting 6,080 addresses were issued to interviewers. Thirdly, at each address 1 adult (aged ≥18 years) was selected at random using a Kish grid. People in institutions, hostels, homes for the elderly or bed and breakfast accommodation were not included. Respondents for re-interview were selected to be representative of the full sample on age, sex and education and to give collectively a balanced coverage of each of the 45 states. Key items of data were entered on a computer as soon as questionnaires were returned so that the retest sample could be issued.

Response rate

Of the 6,080 addresses selected, 756 (12%) were outside the scope of the survey, being non-residential, empty/derelict, untraceable or even not yet built. There were 3,395 interviews achieved, giving a response rate of 64% on in-scope addresses *(table 2)*. Eighty-three per cent of the respondents were willing to participate again and a subsample of 221 were reinterviewed.

Representativeness of sample achieved

In order to correct for the effect of varying household size on selection probabilities, data from each respondent

Table 1 Estimation of sample size

Difference to be detected	Significance level	
	0.01	0.05
0.025	4,827	3,235
0.05	1,207	809
0.10	302	20

Table 2 Response rate

	n	%
Total addresses sampled	6,080	
Out of scope addresses	756	
Total in-scope addresses	5,324	100
Total completed interviews	3,395	64
Refusal by selected person	732	
Refusal of all information	402	
Ill/away/senile	224	
No contact at address	175	
Broken appointment	136	
Proxy refusal	127	
No contact with selected person	74	
Inadequate English	53	
Other	6	
Total successful interviews	1,929	36

were weighted according to the number of adults living in the household. The achieved sample has nearly identical characteristics to the general population (*table 3*). Good geographical coverage was achieved, with percentages of the weighted sample in each of the 8 English standard regions and the 14 regional health authorities and in Scotland and Wales, being virtually identical to those from the 1992 General Household Survey.[34] The retest sample differed significantly from the main sample only in a higher educational level, with 29% of retest respondents having no qualifications (cf. 37% at test).

INTERVIEWER TRAINING

All interviewers attended small group briefings which involved intensive and closely supervised training in the valuation methods. All necessary explanations and questions were included in the script provided and it was emphasized that interviewers must follow the script exactly (e.g. repetition of 'Would you choose life A or life B or are they the same?') and to use only those prompts provided. Written project instructions containing a scripted example interview were also provided and before starting the fieldwork assignment each interviewer completed 2 further practice interviews on family or friends. Interviewers who appeared to be having problems with their first 10 interviews attended a rebriefing before they were permitted to continue. The second assignment of 66 addresses per interviewer was issued approximately 4 weeks after the first 10. At retest, respondents were issued to the same interviewer who had carried out the initial interview.

The main fieldwork was conducted between August and November 1993, with reinterviews during December 1993. Ninety-two interviewers conducted the 3,395 interviews with an average of 37 per interviewer; 13 interviewers were required to attend a rebriefing. Interviews took just under 1 h (54 min, SD 14.6 min) while reinterviews were slightly shorter at 44 min (SD 12.8 min). The TTO method appears complicated at first, but in essence one basic question is repeated throughout (i.e. a choice between 2 options) and after the first 2–3 states, most respondents could value the rest very quickly, particularly as they were familiar with the states from the previous tasks. In general, ranking took approximately 20 min, VAS 5–10 min and TTO 20–30 min.

Table 3 Representativeness of the sample

	Unweighted survey %	Weighted survey %	1992 GHS[34] %	1991 census[35] %
Women[a]	57	54	53	52
Men				
18–34 years	14	15	15	17
35–59 years	17	19	20	20
≥60 years	13	12	12	11
Women				
18–34 years	17	18	16	17
35–59 years	22	23	21	20
≥60 years	18	14	15	15
Qualifications				
Higher/degree	20	20	18	
A level/GCSE	40	41	45	–
Other/none	40	37	38	
Tenure[b]				
Own outright	26	26	25	23
Mortgage	40	44	42	47
Rent/other	34	31	34	29
Social class[c]				
I–II	29	30		30
IIIN/M	44	46	–	43
IV–V/other	26	25		24
Housing[d]				
Detached house	22		20	20
Semi/terraced	57	–	59	58
Flat/other	21		21	22
Economic position				
Paid/seeking work	54	59		61
Retired	22	19	–	20
Sick/disabled/other	24	23		19
Marital status				
Single	17	17	21	
Married/ cohabitating	60	68	64	–
Widowed/divorced	23	15	15	

a: GHS data: adults ≥18 years
b: Unweighted data comparable with GHS (households) but weighted data with census (individuals)
c: Census data: adults ≥16 years
d: GHS and census data based on households not persons
N=3,395; % not summing to 100 due to rounding

DATA PROCESSING

All questionnaires were manually edited and coded by Social and Community Planning Research coding staff, who were themselves carefully monitored. Coding was a complex task, involving the transfer of ranking and rating scores to a coding booklet and the calculation of scores based on responses to the TTO task. Any coding queries that arose during the second assignment were referred to the researchers for a final decision. Data were keyed into the computer and subjected to full verification with further edit checks for consistency and accuracy, with reference to the relevant questionnaire where necessary.

DATA QUALITY

The data set from the 3,395 respondents was virtually complete. There were few missing responses on sociodemographic variables, the most being on social class (2.9%). The missing data on self-rated health status was 0.2–0.4%. Ranking was the most complete of the 3 valuation methods but even on TTO the maximum missing data was only 2.7% for any one state. The states with the most missing data were not always the most severe ones.

Data for all 3 valuation methods were examined to determine whether responses from any respondents (or interviewers) should be excluded from further analysis (*table 4*).

Four separate data sets have been generated, 1 each for ranking, VAS and TTO when each is analysed inde-

pendently of the other methods and a VAS/TTO combined data set for contexts where an identical set of respondents with a complete set of data is required on each method, for example when examining the relationship between VAS and TTO methods.

Criteria for exclusion were as stringent as possible so as to exclude data only when absolutely necessary.

- Respondents with a large amount of missing data.

On ranking, the rank for each state was determined by its position within the set of 15 presented, so that any missing data compromised the information content; for the separate VAS and TTO data sets at least 3 states besides 11111 and 'death' must be valued; and for the combined VAS/TTO data set, only respondents with complete data on both methods were included.

- Respondents with rankings or VAS scores that could not be adjusted to a 0–1 scale, i.e. where 'death' was set equal to or greater than 11111 or where 11111 and/or 'death' was missing.

This consideration does not apply to TTO as the scores generated are already on a 0–1 scale.

- Respondents in the top 5% for logical inconsistency on VAS (or TTO) and who also had missing or unusable data on TTO (or VAS).

Logical inconsistency here refers to the ordinal nature of levels within each EuroQol dimension that requires some states to be given a higher ranking and score than other states. Each respondent valued a different set of states and thus had a different number of possible comparisons. An inconsistency rate was calculated, based on the proportion of possible inconsistencies actually encountered.

- All data from interviewers with a high rate of respondents with missing/unusable data and/or logical inconsistency and/or incomplete interviews.
- Respondents giving the same score to all states on TTO. These people often mentioned religious beliefs and were considered to be akin to 'conscientious objectors' to the TTO method.

REPRESENTATIVENESS OF FINAL DATA SETS

Respondents excluded from the ranking data set were more likely than the remaining respondents to have recorded some problems with mobility ($p<0.01$) or to have reported difficulty with ranking, rating or TTO procedures ($p<0.001$). According to the interviewers' comments, approximately one-third mentioned death (usually recent bereavement) or were in poor physical or mental health.

Respondents excluded from VAS and TTO data sets were more likely than the remaining respondents to be older, retired, separated/divorced/widowed or (in the case of TTO) with no qualifications (all $p<0.001$). They also reported worse health status, i.e. problems with mobility ($p<0.001$), pain/discomfort (TTO, $p<0.01$), depression/anxiety (VAS, $p<0.001$), rated own health status lower (VAS, $p<0.01$) and reported more past/current serious illness (TTO, $p<0.001$). They also reported more difficulty on all valuation methods and fewer of them agreed to a reinterview (both $p<0.001$). Despite these trends however, the respondents remaining in each data set were still representative samples of the general population (*table 5*).

DISCUSSION

The methods used here have been successful in achieving a high response rate (64%) and a sample representative of the British general population in terms of age, sex, education, social class and geographical location. Good quality data were obtained, despite the absence of a cut-off point at the top of the age range, as used elsewhere.[12,14,22,24] Indeed, older people's preferences are important in a society where health care resources are increasingly being directed towards the growing elderly population. The survey required much preparation, including the development of techniques to select a representative set of health states (i.e. spread across the potential valuation space and including all plausible combinations of levels across the 5 EuroQol dimensions), to determine sample size and selection, to train interviewers and to clean and process data (i.e. writing programmes to read data, checking and creation of variables and calculation of scores). These methods can be generalized to other studies and, particularly considering the time and effort required in developing them, are recommended to others considering similar surveys. The methods are not specific to a British population and could be used in other countries, both inside and outside Europe. Both the data and the SPSS files are lodged with the Economic and Social Research Council archive[36] and are freely available for others' use.

A large investment was made in interviewer training, emphasizing the need for data that were not only complete, but also meaningful and reliable. This could only

Table 4 Number of respondents excluded from final data sets

	Ranking	VAS	TTO	VAS/TTO
Missing data	31	55	28	316[d]
Unusable data	39 (death ranked top)	48 (death in top 2)	20[a]	68
Inconsistent and incomplete data	–	2[b]	8[c]	10
Interviewer problems	–	2	2	4
Total	70	107	58	398

a: Thirteen with all states given the same value, 7 with all states worse than death
b: High logical inconsistency on VAS and missing/unusable data on TTO
c: High logical inconsistency on TTO and missing/unusable data on VAS
d: Fifty-five plus 28 already excluded plus a further 233 with any state missing on VAS or TTO
VAS: visual analogue scale
TTO: time trade-off

be achieved by ensuring that each interviewer understood the concepts behind ranking, VAS and TTO and could confidently present the tasks to respondents. The importance of an intensive training session aimed at basic principles and 'hands-on' practice, as well as a concise and detailed script, cannot be overemphasized. The overall impression from interviewers was that both they and their respondents found the exercise stimulating and rewarding.

The process of identifying those respondents whose data should be excluded from the final data sets was not straightforward and it could clearly be biased by researchers' or interviewers' subjective opinions. Thus, a criterion of interviewer judgement was not included as in other studies,[5,16] but rather criteria related specifically to the capacity of the data to generate social preferences. Hence, the emphasis on missing data and the requirement for the 'healthy' state to have the highest valuation, as used by others.[5,15] Beyond these 'errors', it was assumed that valuations given by respondents reflected true preferences and it was not the researcher's role to question these further. Thus, data were not excluded solely on the basis of logical inconsistency as has sometimes been the practice.[14,16]

Although excluded respondents tended to be older with no educational qualifications, respondents in each of the 4 final data sets remain representative of the general British population. These data sets form the basis for further analysis of ranking, VAS and TTO valuations. The next step after successful collection of a complete and consistent data set is to investigate the validity and reliability of such measurement and there remain many questions as to the suitability of even the TTO approach, let alone a VAS approach, in eliciting health preferences to be used in health policy. Another major question concerns the use of aggregated societal data rather than individual data and how to account for factors that influence individual reponses. Since this paper has focused on the more practical aspects of study design and implementation, these aspects are discussed in more detail in later papers, e.g. implications of the differences found between population subgroups on both VAS and TTO valuations, the relationship between valuation methods, technical aspects of the modelling procedure used to generate social tariffs and the limitations and potential uses of these tariffs.

In summary, the methodological challenges posed by this study appear to have been successfully met and have generated new and revised procedures that offer the opportunity for others to conduct similar studies in different settings and populations. It is hoped that the health state valuations elicited from this study will form a valuable basis for use in the clinical and economic evaluation of health care.

Financial support was received from the Department of Health and the ESRC.

The authors thank the SCPR staff and interviewers for their valuable and enthusiastic collaboration in this study and the participants from the general public who made this study possible. Thanks also to Sally Baker and Kerry Atkinson for their secretarial assistance and to many colleagues who provided advice, support and guidance along the way.

REFERENCES

1 Richardson A, Charny M, Hanmer-Lloyd S. Public opinion and purchasing. BMJ 1992;304:680-2.

2 Kaplan RM, Debon M, Anderson B. Effects of number of rating scale points upon utilities in a quality of well-being scale. Med Care 1991;29:1061-4.

3 Kind P, Gudex C, Dolan P, Williams A. Practical and methodological issues in the development of the EuroQol: the York experience. Adv Med Sociol 1994;5:219-53.

4 O'Connor A, Boyd M, Warde P, Stolbach L, Till J. Eliciting preferences for alternative drug therapies in oncology: influence of treatment outcome description, elicitation technique and

Table 5 Characteristics of respondents

	Ranking n=3,325	VAS n=3,288	TTO n=3,337	VAS/TTO n=2,997
Mean age years (SD)	47.8 (18.3)	47.5 (18.2)	47.7 (18.3)	47.1 (18.1)
% female	56.7	56.7	56.8	56.9
% separated/divorced/ widowed	23.0	22.6	22.8	22.0
% no qualifications	36.8	36.5	36.5	35.5
% retired	22.4	22.0	25.6	24.6
% rent home	32.7	32.9	33.1	33.3
% current smokers[a]	30.9	31.0	30.9	30.9
% illness in past/now	31.6	31.4	31.5	30.8
% problems with				
Mobility	18.1	17.9	18.1	17.7
Self-care	4.2	4.3	4.2	4.1
Usual activities	16.2	16.0	16.2	16.1
Pain/discomfort	32.8	32.7	32.8	32.7
Anxiety/depression	20.8	20.5	20.8	20.4
Own health: median[b]	90.0	90.0	90.0	90.0
IQR	(75–95)	(75–95)	(75–95)	(75–95)
% major difficulty				
On Ranking	11.0	10.4	10.6	9.8
On VAS	8.3	7.6	8.0	6.7
On TTO	5.0	4.8	4.7	3.6
% agreed to retest	15.6	15.2	16.7	15.8

a: Twenty-eight per cent in general population (GHS 1992[34])
b: VAS, 0 (worst) to 100 (best)
VAS: visual analogue scale
TTO: time trade-off
SD: standard deviation
IQR: interquartile range

treatment experience on preferences. J Chronic Dis 1987;40:811-18.
5 Patrick DL, Starks HE, Cain KC, Uhlmann RF, Pearlman RA. Measuring preferences for health states worse than death. Med Decis Making 1994;14:9-18.
6 Bass E, Steinberg E, Pitt H, et al. Comparison of the rating scale and the standard gamble in measuring patient preferences for outcomes of gallstone disease. Med Decis Making 1994;14:307-14.
7 Read JL, Quinn RJ, Berwick DM, Fineberg HV, Weinstein MC. Preferences for health outcomes: comparison of assessment methods. Med Decis Making 1984;4(3):315-29.
8 Torrance GW. Social preferences for health states: an empirical evaluation of three measurement techniques. Socio-Econom Plan Sci 1976;10:129-36.
9 Balaban DJ, Sagi PC, Goldfarb NI, Nettler S. Weights for scoring the quality of well-being (QWB) instrument among rheumatoid arthritics: a comparison to general population weights. Med Care 1986;24:973-80.
10 McNeil BJ, Weichselbaum R, Pauker SG. Speech and survival: tradeoffs between quality and quantity of life in laryngeal cancer. N Engl J Med 1981,305:982-7.
11 Tsevat J, Dawson NV, Matchar DB. Assessing quality of life and preferences in the seriously ill using utility theory. J Clin Epidemiol 1990;43(Suppl):73S-77S.
12 Fryback D, Dasbach E, Klein R, et al. The Beaver Dam Health Outcomes Study: initial catalog of health state quality factors. Med Decis Making 1993;13:89-102.
13 Kaplan RM, Anderson JP. A general health policy model: update and application. Hlth Serv Res 1988;23:203.
14 Sackett DL, Torrance GW. The utility of different health states as perceived by the general public. J Chronic Dis 1978;31:697-704.
15 Torrance GW, Boyle MH, Horwood SP. Application of multi-attribute utility theory to measure social preferences for health states. Oper Res 1982;30:1043-69.
16 Torrance GW, Zhang Y, Feeny D, Furlong W, Barr R. Multi-attribute preference functions for a comprehensive health status classification system. CHEPA Working Paper Series No.92-18. Hamilton, Canada: McMaster University, 1992.
17 Klevit H, Bates A, Castanares T, Kirk P, Sipes-Metzler P, Wopat R. Prioritization of health care services: a progress report by the Oregon Health Services Commission. Arch Int Med 1991;151:912-6.
18 Rosser R, Kind P. A scale of valuation of states of illness: is there a social consensus? Int J Epidemiol 1978;7:347-58.
19 Gudex C, Kind P, van Dalen H, Durand M, Morris J, Williams A. Comparing scaling methods for health state valuations: Rosser revisited. Centre for Health Economics, University of York: Discussion Paper 107, 1993.
20 van Dalen H, Williams A, Gudex C. Lay people's evaluations of health: are there differences between subgroups? J Epidemiol Commun Hlth 1994;48:248-53.
21 EuroQol Group. EuroQol: a new facility for the measurement of health-related quality of life. Hlth Policy 1990;16:199-208.
22 Brooks RG, Jendteg S, Lindgren B, Persson U, Bjork S. EuroQol health-related quality of life measurement: results of the Swedish questionnaire exercise. Hlth Policy 1991;18:37-48.
23 Essink-Bot ML, Bonsel GJ, Maas PJ van der. Valuation of health states by the general public: feasibility of a standardised measurement procedure. Soc Sci Med 1990;31:1201-6.
24 Nord E. Euroqol: health-related quality of life measurement. Valuations of health states by the general public in Norway. Hlth Policy 1991;18:25-36.
25 Carr-Hill RA. Health related quality of life measurement: Euro style. Hlth Policy 1992;20:321-8.
26 EuroQol Group. EuroQol: a reply and reminder. Hlth Policy 1992;20:329-32.
27 Gafni A, Birch S. Searching for a common currency: challenges to the European harmonization of the measurement of health related quality of life (Euroqol). Hlth Policy 1993;23:219-28.
28 Dolan P. Search for a criticial appraisal of EuroQol: a response by the EuroQol Group to Gafni and Birch. Hlth Policy 1994;28:67-9.
29 Dolan P, Gudex C, Kind P, Williams A. Valuing health states: a comparison of methods. J Hlth Econ 1996;15:209-31.
30 Thomson K. Health related quality of life: the 1993 pilots. London: Social and Community Planning Research, 1993.
31 Erens R. Health related quality of life: general population survey. London: Social and Community Planning Research, 1994.
32 Gudex C. Time trade-off user manual: props and self-completed methods. University of York: Centre for Health Economics, 1994.
33 Ohinmaa A, Sintonen H. The effect of duration on the values given to the EuroQol states. In: Busschbach J, Bonsel G, de Charro F, editors. Proceedings of the EuroQol Meeting, Rotterdam 1993. Rotterdam: Erasmus University of Rotterdam, Department of Public Health, 1994;51-9.
34 OPCS. General Household Survey. Series GHS No.23. London: OPCS, 1992.
35 OPCS. Key population and vital statistics. Series VS No.17, PP1 No.13. London: HMSO, 1990.
36 Williams A, Kind P, Gudex C, Dolan P. Health state valuations from the British general public. Archive number 3444. London: Economic and Social Research Council (ESRC) Data Archive, January 1996.

Received 18 September 1995, accepted 26 March 1996