Quality of Life Research, 5, pp. 521–531

Health state valuations from the general public using the Visual Analogue Scale

C. Gudex,* P. Dolan, P. Kind and A. Williams

Centre for Health Economics, University of York (C. Gudex, P. Kind, A. Williams); Department of Economics, University of Newcastle (P. Dolan)

In the clinical and economic evaluation of health care, the value of benefit gained should be determined from a public perspective. The objective of this study was to establish relative valuations attached to different health states to form the basis for a 'social tariff' for use in quantifying patient benefit from health care. Three thousand three hundred and ninety-five interviews were conducted with a representative sample of the adult British population. Using the EuroQol health state classification and a visual analogue scale (VAS), each respondent valued 15 health states producing, in total, direct valuations for 45 states. Two hundred and twenty-one re-interviews were conducted approximately 10 weeks later. A near complete, and logically consistent, VAS data set was generated with good test–retest reliability (mean ICC = 0.78). Both social class and education had a significant effect, where higher median valuations were given by respondents in social classes III–V and by those with intermediate or no educational qualifications. These effects were particularly noticeable for more severe states. The use of such valuations in a social tariff raises important issues regarding the use of the VAS method itself to elicit valuations for hypothetical health states, the production of separate tariffs according to social class and/or education and the appropriate measure of central tendency.

Key words: EuroQol; health status measurement; social preferences; visual analogue scale.

Introduction

In the clinical and economic evaluation of health care there is general agreement on two key issues. Firstly, the benefit gained from different forms of health care should not be measured solely by survival but also by the 'health-related quality of life'** of that survival. Secondly, weight should be given to the views of the general public, in their capacity both as taxpayers and as potential patients, in determining how good or bad potential health states are.

The achievement of both these conditions requires a generic (*i.e.* non disease-specific) classification of health status where every state is given a single numerical value reflecting the relative preference for being in that state. These values should reflect the views of the general population and should preferably lie on an interval scale.[2] The classification and its associated 'social tariff' could then be used to describe and quantify health status before and after treatment, thus providing a measure of benefit from health care. They could also be used in association with population surveys to measure levels and trends in community health and to provide the quality adjustment in calculating Healthy Life Expectancy.

It is conventional to have health state valuations range from 1.00 (representing 'full' or 'good' health) down to 0.00 (representing death), with negative values possible for states considered worse than death. Several methods have been used to generate these valuations, the principal ones being Category Rating (CR), Time Trade-off (TTO) and Standard Gamble (SG).[2] Each of these methods makes different assumptions about how individuals choose between health states, and debate continues over which is the most appropriate, particularly as the different methods appear to produce different valuations.[3–5]

In the health economics field, it is often argued that health state valuations for use in a social tariff should be generated by a method, such as TTO or SG, which requires the respondent to make some sort of choice between health states, involving a notion of sacrifice. Thus the resulting valuations are said to

Work attributed to: Centre for Health Economics, University of York, York YO1 5DD, England.

* To whom correspondence should be addressed at Tordenskjoldsgade 23, DK-1055 Copenhagen K, Denmark.

** Defined by Patrick and Erikson[1] as 'the value assigned to duration of life as modified by the impairments, functional states, perceptions and social opportunities that are influenced by disease, injury, treatment or policy'.

be 'real' preferences for different health states. This requirement excludes CR methods, such as the Visual Analogue Scale (VAS), which ask a respondent to rate a series of health states but without making any choices or decisions between health states.[6]

The VAS approach has, however, considerable advantages. Its origin lies in the field of psychometrics (especially psychophysics). A review of the use of the VAS in pain measurement reports that the method is readily understood by most people and that it is reliable, sensitive and concurrently valid with other methods.[7] Good sensitivity to change has also been reported for a VAS mood scale,[8] while in the measurement of general well-being and health status there is evidence for its ease of use and high reliability, both internally and on test–retest.[9–12] The VAS method has been used in several further studies to assess change in patient health status with treatment.[13–15]

This paper reports on data obtained as part of a major project designed to produce a social tariff for use in the clinical and economic evaluation of health care in Britain. The primary purpose of the project was to obtain health state valuations from a representative sample of the non-institutionalized adult population of England, Scotland and Wales. Health states were described in terms of the EuroQol classification, a standardized generic instrument for describing and valuing health-related quality of life[16] which has been used in a range of health care settings[14,17] and in general population surveys.[18–21] In addition the study sought to identify any population subgroups with markedly different valuations.

Two valuation methods were used: VAS, as used by the EuroQol Group in postal surveys, and TTO, chosen on the basis of a previous empirical comparison between SG and TTO.[22] It was recognized from the outset that the VAS valuations would offer the opportunity to produce a second social tariff to be compared with that from the TTO. Presented here are the VAS valuations for 45 different health states elicited from the general public. The comparison with TTO forms (the subject of further work) will be reported elsewhere.

Figure 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

Method

Sample selection

A three-stage sampling procedure was adopted with the aim of achieving a nationally representative sample of at least 3,235 respondents.* Eighty postcode sectors, stratified by 16 health regions (14 English Regional Health Authorities plus Scotland and Wales) and two socioeconomic bands, were selected from the Post Code Address File. Within each of these 32 groupings, postcode sectors were ranked in order of population density. The second stage was to select 76 addresses from across the whole of each selected postcode sector, thus generating 6,080 addresses. As a final stage, interviewers randomly selected one adult at each address.

Selection of states

The EuroQol classification describes health status in terms of five dimensions of mobility, self-care, usual activities, pain/discomfort and anxiety/depression, each of which has three levels (Figure 1). Health states are defined by combining one statement from each

* A sample size of 3,235 was set in order to detect a 0.10 difference in valuations (where 1.00 = 'full health' and 0.00 = 'death') between population subgroups at the 5% level. This was based on a desired power of 80%, a mean SD of 0.35, and the use of non-parametric tests having 95% of the efficiency of parametric tests.

of the five dimensions, generating 243 possible health states. To these need to be added two further states which are not technically defined by the EuroQol classification: 'unconscious' and 'immediate death'. Pilot studies had established that with direct valuations for approximately 40 states, it was possible to model the valuations for the remaining states. Since each respondent could not handle more than 15 states in an hour-long interview, states were selected to ensure a wide spread in severity of state and coverage of all plausible combinations of levels across the five dimensions. All respondents valued 11111, 33333, 'unconscious' and 'immediate death'; each respondent also valued two of the five mildest states (11112, 11121, 11211, 12111 and 21111), together with three states from each of 12 'mild', 12 'moderate' and 12 'severe' states.

Interview procedure

Each respondent was first asked to describe his/her own current health state using the EuroQol classification. Fifteen EuroQol health states printed on cards were then presented to the respondent who was asked to rank them from best to worst, assuming that each state would last for 10 years without any change, followed by death. A visual analogue scale consisting of a 20cm vertical line, similar to a thermometer, was then used to record values for these health states. The endpoints of the 'thermometer' were labelled 100 ('best imaginable health state') and 0 ('worst imaginable health state'). A process of 'fractionation'[23] was used to encourage respondents to treat the scale as having interval properties. As on the ranking task, respondents were told that each state would last for 10 years without any change followed by death. After all 15 states had been rated using a method of 'fractionation', the respondent then rated his/her own health on a separate VAS. At the end of the interview personal background data were collected including age, occupation, education and experience of illness. Respondents were asked about their own personal experience of serious illness, whether they had worked with ill people, and whether someone close among family or friends had had serious illness. Further details were elicited regarding onset and length of illness. Respondents were interviewed in their own homes between August and November 1993.

Retest interviews

Two hundred and twenty-one respondents were re-interviewed an average of 10 weeks later by the same interviewer who had conducted the first interview. The same protocol was followed, using the same health states but with the addition of some background questions asking about any new experiences of illness since the original interview.

Data analysis

A limited volume of data was excluded for reasons of practicality, such as missing data or unusable valuations (*e.g.* death given the highest valuation). VAS data from 107 respondents (3.2% of total) were excluded from further analysis.*

'Raw' valuations were read off directly from the VAS with its endpoints of 100 and 0. In order to compare valuations from different respondents, these raw valuations were adjusted relative to 11111 and death using the formula:

$$\frac{V_x = R_{(x)} - R_{(Death)}}{R_{(11111)} - R_{(Death)}}$$

Where V_x = Adjusted valuation for health state x; $R_{(x)}$ = raw valuation given to state x; $R_{(Death)}$ = raw valuation given to 'immediate death' and $R_{(11111)}$ = raw valuation given to state 11111

Thus the state 11111 and 'immediate death' always had adjusted valuations of 1.0 and 0.0 respectively and the 'unit of health' was the same across all individuals. Adjusted valuations for the remaining states were either greater than zero (if considered better than death) or less than zero (if considered worse than death).

Due to the ordinal nature of levels within each EuroQol dimension, there are some states that should logically be given a higher valuation than other states. The definition of logical consistency used here is a 'strong' one, in which the logically better state is required to be valued higher than the worse state and not just equal to it. Each respondent valued a different set of states and thus had a different number of possible comparisons. An inconsistency rate was

* Fifty-five had less than the minimum data set required (three states valued plus 11111 and death);48 had rated death equal to or higher than 11111 or higher than all other states except 11111; two had both high logical inconsistency on VAS and missing data on TTO; and two were excluded because of poor interviewer performance. These 107 respondents were significantly older and less educated than the other 3,288 respondents. They were also more likely to have past or current personal experience of illness and to have more difficulty with the interview.

calculated, based on the proportion of possible inconsistencies actually encountered.

For Mann–Whitney and Kruskal–Wallis tests, statistical significance was set at $p < 0.01$ in view of the large number of tests conducted. Test–retest comparisons were made on an individual-by-individual basis using an intraclass correlation coefficient (ICC),[24] calculated for each respondent. The closer the ICC was to 1, the greater the reliability.

Results

Representativeness of sample

Interviews were achieved with 3,395 members of the general public, representing a 64% response rate for in-scope addresses. After exclusion of the 107 respondents with missing or unusable data, the sample

Table 1. Characteristics of respondents. Unless otherwise stated, GHS data[25] are for adults 16+ yrs and Census data[26] for 18+ yrs of age. For study population, $n = 3,288$. Percentages not summing to 100 are due to rounding.

	Survey (unweighted) %	Survey (weighted) %	1992 GHS %	1991 Census %
Sex[a]				
Men	48	46	47	48
Women	52	54	53	52
Age				
18–24	9	11	11	13
25–34	22	22	20	20
35–49	25	26	27	26
50–59	14	14	15	14
60–64	7	7	7	7
65+	23	19	21	21
Qualification				
Degree	9	10	8	—
Higher education	11	11	10	—
A level	9	10	11	—
GCSE A–C	20	20	23	—
GCSE D–G	11	11	11	—
Other	3	3	3	—
None	37	34	35	—
Tenure[b]				
Own	25	26	25	23
Mortgage	41	44	42	47
Rent LA/HA	24	21	25	21
Rent private	8	8	7	6
Other	2	2	2	2
Social class[c]				
I–II	29	30	—	30
IIIN	24	25	—	22
IIIM	21	21	—	21
IV–V	25	24	—	21
Other	1	1	—	3

Survey respondents were also virtually identical to GHS/Census on type of accommodation, marital status and economic position.

[a] GHS data based on adults aged 18 and over.
[b] Unweighted survey data should be compared with GHS data (based on households), weighted survey data with Census (based on individuals 18+ yrs).
[c] Census data based on adults aged 16 and over.

N.B. The survey data were weighted to correct for the effect of varying household size on selection probabilities. Each respondent was given a weight according to the number of adults living in the household.

remained representative of the general British population with respect to sociodemographic characteristics (Table 1) and geographical coverage across the 8 standard economic regions, the 14 English RHAs, Scotland and Wales.

Raw and adjusted valuations

Median raw valuations varied between 100 (11111) and 0 (death) with no state being valued below death (Table 2). Only one state, 33333, had a mean raw valuation less than zero. Adjusting the raw valuations

Table 2. Raw and adjusted VAS valuations

		Raw valuations		Adjusted valuations	
State	***n***	**Median (IQR)**	**Mean (SD)**	**Median (IQR)**	**Mean (SD)**
11111	3,288	100.0 (100–100)	98.7 (4.8)	1.00 (—)	1.00 (—)
11112	1,308	87.0 (75–95)	82.4 (15.2)	0.87 (0.75–0.94)	0.81 (0.23)
11121	1,320	85.0 (75–94)	82.8 (13.6)	0.86 (0.75–0.94)	0.81 (0.21)
11211	1,322	85.0 (75–92)	82.2 (14.0)	0.85 (0.75–0.92)	0.81 (0.21)
21111	1,307	85.0 (75–90)	81.4 (14.2)	0.85 (0.74–0.91)	0.79 (0.24)
12111	1,310	85.0 (75–90)	80.7 (14.5)	0.84 (0.75–0.90)	0.79 (0.18)
12211	828	75.0 (63–80)	71.3 (15.1)	0.73 (0.60–0.80)	0.68 (0.23)
11122	814	75.0 (60–80)	70.8 (15.0)	0.72 (0.57–0.80)	0.66 (0.40)
12121	831	75.0 (60–80)	70.0 (15.9)	0.71 (0.56–0.80)	0.66 (0.23)
22121	832	65.0 (50–75)	63.1 (17.0)	0.64 (0.49–0.75)	0.57 (0.35)
22112	825	65.0 (50–75)	63.3 (17.2)	0.63 (0.48–0.75)	0.59 (0.29)
21222	815	60.0 (50–75)	58.8 (17.7)	0.56 (0.43–0.71)	0.51 (0.50)
12222	827	60.0 (50–71)	58.1 (17.6)	0.55 (0.43–0.70)	0.52 (0.38)
22122	814	57.0 (49–70)	57.2 (17.6)	0.53 (0.40–0.70)	0.50 (0.44)
11312	824	55.0 (45–70)	55.9 (19.5)	0.53 (0.38–0.69)	0.50 (0.35)
11113	818	55.0 (40–75)	55.0 (22.7)	0.51 (0.35–0.72)	0.47 (0.47)
21312	802	51.0 (40–65)	52.0 (17.8)	0.50 (0.35–0.63)	0.43 (0.53)
22222	834	50.0 (41–65)	52.3 (17.3)	0.50 (0.35–0.61)	0.45 (0.37)
13212	818	50.0 (35–60)	48.3 (18.6)	0.45 (0.30–0.60)	0.40 (0.46)
11131	806	50.0 (30–67)	48.4 (22.9)	0.45 (0.25–0.65)	0.39 (0.53)
13311	809	45.0 (30–56)	43.8 (19.5)	0.40 (0.24–0.55)	0.34 (0.59)
12223	817	41.0 (30–55)	43.0 (19.7)	0.37 (0.22–0.53)	0.32 (0.52)
21232	822	38.0 (25–50)	39.0 (19.2)	0.33 (0.19–0.50)	0.31 (0.34)
32211	817	35.0 (20–50)	36.3 (19.5)	0.30 (0.15–0.46)	0.28 (0.38)
11133	823	35.0 (20–50)	36.9 (21.2)	0.30 (0.14–0.49)	0.25 (0.55)
21323	822	35.0 (25–50)	35.9 (18.3)	0.30 (0.15–0.45)	0.21 (0.85)
23321	823	30.0 (21–45)	33.4 (17.2)	0.26 (0.15–0.41)	0.24 (0.38)
22331	819	30.0 (19–45)	31.7 (17.9)	0.25 (0.12–0.40)	0.16 (0.68)
21133	826	30.0 (15–46)	32.7 (19.9)	0.25 (0.10–0.43)	0.19 (0.81)
22323	816	28.0 (15–40)	30.2 (17.8)	0.25 (0.11–0.38)	0.13 (0.97)
33212	829	25.0 (15–40)	28.3 (16.7)	0.22 (0.10–0.35)	0.14 (0.64)
23232	808	25.0 (15–40)	28.3 (16.7)	0.21 (0.10–0.35)	0.18 (0.44)
23313	823	25.0 (15–35)	26.5 (14.5)	0.20 (0.08–0.32)	0.13 (0.65)
22233	825	24.0 (13–35)	25.6 (17.1)	0.17 (0.07–0.31)	0.12 (0.64)
32232	820	20.0 (10–30)	23.4 (15.9)	0.17 (0.05–0.28)	0.06 (0.77)
13332	805	20.0 (10–32)	23.9 (16.6)	0.16 (0.05–0.30)	0.11 (0.57)
32313	833	20.0 (10–30)	23.5 (15.6)	0.16 (0.06–0.29)	0.11 (0.51)
32223	825	20.0 (10–30)	22.8 (15.5)	0.15 (0.05–0.27)	0.10 (0.56)
33321	813	20.0 (10–30)	22.0 (15.6)	0.15 (0.05–0.25)	0.08 (0.60)
32331	816	18.0 (10–30)	20.6 (14.4)	0.13 (0.03–0.25)	0.03 (0.98)
33232	823	14.0 (7–25)	16.2 (12.7)	0.10 (0.00–0.20)	0.01 (0.71)
33323	838	10.0 (5–20)	13.9 (12.4)	0.07 (–0.02–0.16)	–0.03 (1.04)
uncon	3,286	5.0 (0–10)	9.3 (15.1)	0.01 (–0.02–0.05)	–0.04 (0.52)
33333	3,278	2.0 (0–8)	5.6 (9.1)	0.00 (–0.08–0.05)	–0.13 (0.90)
death	3,288	0.0 (0–10)	8.5 (15.7)	0.00 (—)	0.00 (—)

relative to 11111 and death had no effect on the relative ordering of states. Median adjusted valuations were always higher than mean adjusted valuations. Again, no state had a median valuation below zero (although state 33333 was rated equal to death), while on mean valuations, three states were rated worse than death.

Logical inconsistency

There were no logical inconsistencies in the rank order of median (or mean) valuations, either raw or adjusted. At an individual level, 57.4% of respondents had no logical inconsistencies at all. The mean overall inconsistency rate was 2.5%.

The independent effect of different variables on adjusted valuations*

Social class. Social class was calculated according to the respondent's own current or most recent occupation, as used in the Census. For 34 of the 43 states (79%), social class had a significant effect on the valuation and in every case the median valuation for respondents in social classes III–V was higher than that for respondents in social classes I and II (Table 3). Twenty-two of the differences were significant at $p < 0.01$ or less. The trend was more noticeable as the severity of the state worsened.

Educational qualifications. For 31 of the 43 states (72%), education had a significant effect on the valuation and in every case the median valuation for respondents with intermediate or no educational qualifications was higher than that for respondents with degree qualifications (Table 4). All but seven of the differences were significant at $p < 0.01$ or less. The trend was more noticeable as the severity of the state worsened.

Home ownership. Home ownership had a significant effect on the valuation for 13 of 43 states (30%) and in every case the median valuation for respondents who rented their home (from local authority, housing association or privately) was higher than that for respondents who owned their home (outright or with a mortgage). All but three of the differences were significant at $p < 0.01$ or less.

* Distributions of adjusted valuations for all states were non-normal (Kolmogorov-Smirnov test, $p < 0.001$). Non-parametric tests were thus used in the analysis which follows.

Table 3. Adjusted VAS valuations by social class (medians and interquartile ranges). One hundred respondents were omitted due to missing data on social class.

State	Social class	
	I,II ($n = 947$)	IIIN/M,IV,V ($n = 2241$)
11111	1.00 (—)	1.00 (—)
11112	0.85 (0.75–0.93)	0.87 (0.75–0.94)
11121	0.85 (0.75–0.92)	0.87 (0.75–0.94)*
11211	0.84 (0.75–0.92)	0.87 (0.75–0.92)
21111	0.84 (0.72–0.90)	0.85 (0.75–0.92)*
12111	0.84 (0.74–0.90)	0.85 (0.75–0.90)
12211	0.72 (0.60–0.79)	0.73 (0.60–0.80)
11122	0.70 (0.51–0.79)	0.74 (0.60–0.81)**
12121	0.70 (0.52–0.80)	0.71 (0.56–0.80)
22121	0.60 (0.44–0.73)	0.65 (0.50–0.75)**
22112	0.60 (0.45–0.74)	0.64 (0.50–0.75)*
21222	0.53 (0.39–0.69)	0.58 (0.44–0.73)*
12222	0.51 (0.40–0.67)	0.56 (0.44–0.70)*
22122	0.53 (0.44–0.69)	0.53 (0.40–0.70)
11312	0.49 (0.34–0.67)	0.55 (0.40–0.70)**
11113	0.46 (0.28–0.67)	0.54 (0.35–0.75)***
21312	0.46 (0.30–0.59)	0.50 (0.35–0.65)**
22222	0.48 (0.31–0.55)	0.50 (0.38–0.62)**
13212	0.45 (0.25–0.58)	0.45 (0.30–0.60)
11131	0.38 (0.17–0.55)	0.48 (0.25–0.67)***
13311	0.36 (0.24–0.50)	0.40 (0.24–0.56)
12223	0.31 (0.17–0.47)	0.40 (0.25–0.56)***
21232	0.30 (0.15–0.47)	0.35 (0.20–0.50)*
32211	0.27 (0.11–0.44)	0.30 (0.15–0.48)*
11133	0.28 (0.10–0.45)	0.32 (0.15–0.50)*
21323	0.25 (0.09–0.38)	0.32 (0.18–0.48)***
23321	0.25 (0.12–0.39)	0.29 (0.16–0.42)**
22331	0.24 (0.07–0.35)	0.25 (0.14–0.40)*
21133	0.20 (0.05–0.36)	0.26 (0.11–0.45)***
22323	0.20 (0.06–0.33)	0.25 (0.13–0.40)***
33212	0.20 (0.07–0.30)	0.24 (0.10–0.36)*
23232	0.18 (0.05–0.35)	0.23 (0.11–0.37)***
23313	0.17 (0.05–0.30)	0.21 (0.10–0.33)*
22233	0.13 (0.03–0.25)	0.20 (0.10–0.35)***
32232	0.11 (0.00–0.24)	0.18 (0.06–0.29)***
13332	0.13 (0.03–0.25)	0.19 (0.05–0.31)***
32313	0.12 (0.02–0.25)	0.19 (0.07–0.30)***
32223	0.14 (0.03–0.25)	0.17 (0.06–0.28)*
33321	0.15 (0.04–0.25)	0.15 (0.05–0.25)
32331	0.10 (−0.02–0.20)	0.15 (0.05–0.25)***
33232	0.05 (−0.11–0.11)	0.11 (0.03–0.21)***
33323	0.05 (−0.08–0.13)	0.09 (0.00–0.17)***
uncon	0.00 (−0.04–0.05)	0.01 (−0.01–0.05)***
33333	0.00 (−0.12–0.05)	0.01 (−0.06–0.06)***
death	0.0 (—)	0.0 (—)

* $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$

Table 4. Adjusted VAS valuations by education (medians and interquartile ranges). Four respondents were omitted due to missing education data

State	Educational qualifications	
	Degree ($n = 661$)	Intermediate/None ($n = 2,623$)
11111	1.00 (—)	1.00 (—)
11112	0.85 (0.70–0.92)	0.88 (0.77–0.94)**
11121	0.85 (0.74–0.91)	0.87 (0.75–0.94)**
11211	0.83 (0.75–0.90)	0.86 (0.75–0.93)*
21111	0.84 (0.70–0.90)	0.85 (0.74–0.92)*
12111	0.85 (0.75–0.90)	0.84 (0.75–0.90)
12211	0.74 (0.59–0.78)	0.73 (0.60–0.80)
11122	0.70 (0.51–0.79)	0.74 (0.59–0.81)**
12121	0.70 (0.51–0.80)	0.71 (0.57–0.80)
22121	0.59 (0.43–0.72)	0.65 (0.50–0.75)**
22112	0.60 (0.44–0.75)	0.64 (0.50–0.75)
21222	0.56 (0.39–0.69)	0.56 (0.44–0.72)
12222	0.51 (0.40–0.66)	0.56 (0.44–0.71)*
22122	0.50 (0.40–0.65)	0.55 (0.40–0.70)
11312	0.50 (0.34–0.67)	0.55 (0.40–0.70)*
11113	0.47 (0.28–0.65)	0.53 (0.35–0.74)*
21312	0.44 (0.25–0.55)	0.50 (0.36–0.65)***
22222	0.48 (0.30–0.55)	0.50 (0.37–0.61)*
13212	0.42 (0.24–0.55)	0.45 (0.30–0.60)
11131	0.41 (0.21–0.60)	0.46 (0.25–0.65)
13311	0.35 (0.21–0.50)	0.40 (0.25–0.55)
12223	0.30 (0.17–0.47)	0.39 (0.24–0.55)***
21232	0.25 (0.11–0.44)	0.35 (0.20–0.50)***
32211	0.31 (0.12–0.45)	0.30 (0.15–0.47)
11133	0.23 (0.06–0.42)	0.32 (0.15–0.50)***
21323	0.24 (0.11–0.38)	0.32 (0.17–0.46)***
23321	0.23 (0.09–0.35)	0.29 (0.17–0.42)***
22331	0.20 (0.06–0.33)	0.26 (0.14–0.40)***
21133	0.20 (0.06–0.33)	0.27 (0.11–0.45)***
22323	0.21 (0.06–0.32)	0.25 (0.12–0.40)**
33212	0.17 (0.07–0.30)	0.24 (0.10–0.35)*
23232	0.15 (0.07–0.32)	0.23 (0.10–0.37)***
23313	0.15 (0.02–0.30)	0.20 (0.10–0.33)**
22233	0.13 (0.02–0.25)	0.19 (0.09–0.34)***
32232	0.10 (–0.03–0.21)	0.19 (0.06–0.29)***
13332	0.13 (0.01–0.25)	0.18 (0.05–0.30)**
32313	0.13 (0.01–0.25)	0.17 (0.07–0.30)**
32223	0.15 (0.05–0.25)	0.16 (0.05–0.28)
33321	0.16 (0.03–0.26)	0.15 (0.05–0.25)
32331	0.10 (–0.05–0.19)	0.15 (0.04–0.25)***
33232	0.06 (–0.06–0.12)	0.10 (0.02–0.21)***
33323	0.03 (–0.06–0.11)	0.09 (0.00–0.17)***
uncon	0.00 (–0.03–0.05)	0.01 (–0.01–0.05)***
33333	0.00 (–0.12–0.05)	0.00 (–0.06–0.06)***
death	0.00 (—)	0.00 (—)

* $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$

Illness in others. Seventy-two per cent of all respondents reported that somebody close to them had had a serious illness at some time. Most referred to close family members and in half the cases the illness was either still continuing or had ended within the last 5 years. Experience of illness in others had a significant effect on the valuation for eight of 43 states (19%) at the $p < 0.01$ level or less. For all states, the median valuation for respondents without experience of illness in others was higher than that for respondents with this experience.

Other factors. Current self-reported health status had a significant effect on the valuation for only six of 43 states (14%) at $p < 0.01$. In five of these cases, the median valuation for respondents who reported some problem with any EuroQol dimension, or whose VAS rating of their own health was < 75 (lower interquartile range), was higher than that for respondents in full health or with a self-rating > 95 (upper quartile range). Other factors including age, sex, smoking, marital status, economic activity, work experience of looking after ill people, reported experience of serious illness in self, or geographical location (defined either by standard economic region or by regional health authority) had no significant effect on VAS valuations.

Controlling for related variables

Assessment of the independent effect of the respondent's personal characteristics on health state values was complicated by the fact that age, social class, education and home ownership were all significantly related with each other (χ^2 tests). Respondents aged 60 years and over were more likely to be in social classes III–V ($p < 0.001$), while those in social classes III–V were themselves less likely to have degree qualifications ($p < 0.001$) or to be home owners ($p < 0.001$). In order to control for this interaction, two similar approaches to analysis were taken. Firstly, the Mann–Whitney tests were simply repeated after controlling for social class, education and age in turn. Controlling for social class noticeably reduced the effect of education while controlling for education only slightly reduced the effect of social class. Controlling for age did not remove any of the effect of either social class or education. All significant differences were in the same direction where higher median valuations were generated by respondents in social classes (SC) III–V and respondents without degree level education.

Secondly, Kruskal–Wallis tests were conducted comparing valuations for four groups of respondents: SC I–II with degree education, SC I–II with inter-

Table 5. Adjusted test and retest VAS valuations. Numbers given are medians (and interquartile ranges).

State	*n*	Test	Retest
11111*	212	1.00 (—)	1.00 (—)
11112	85	0.89 (0.75–0.94)	0.89 (0.78–0.94)
11121	84	0.88 (0.76–0.93)	0.85 (0.75–0.92)
11211	73	0.87 (0.75–0.94)	0.89 (0.80–0.94)
21111	91	0.85 (0.72–0.90)	0.88 (0.75–0.92)
12111	91	0.85 (0.75–0.91)	0.85 (0.75–0.94)
12211	52	0.75 (0.61–0.83)	0.80 (0.70–0.85)
11122	46	0.74 (0.60–0.79)	0.73 (0.55–0.83)
12121	50	0.73 (0.59–0.80)	0.75 (0.65–0.80)
22121	47	0.63 (0.50–0.75)	0.67 (0.56–0.75)
22112	63	0.60 (0.49–0.75)	0.69 (0.48–0.75)
21222	59	0.60 (0.47–0.70)	0.63 (0.51–0.75)
12222	57	0.60 (0.38–0.72)	0.60 (0.45–0.70)
22122	52	0.60 (0.45–0.75)	0.56 (0.50–0.72)
11312	54	0.55 (0.46–0.70)	0.55 (0.45–0.69)
11113	62	0.51 (0.44–0.67)	0.49 (0.30–0.60)
21312	59	0.51 (0.39–0.65)	0.45 (0.26–0.62)
22222	49	0.50 (0.40–0.60)	0.51 (0.46–0.65)
13212	52	0.50 (0.35–0.73)	0.50 (0.30–0.75)
11131	60	0.42 (0.15–0.67)	0.46 (0.26–0.61)
13311	48	0.41 (0.28–0.58)	0.45 (0.28–0.57)
12223	55	0.39 (0.18–0.50)	0.30 (0.14–0.44)
21232	48	0.35 (0.18–0.50)	0.34 (0.14–0.50)
32211	50	0.34 (0.22–0.55)	0.43 (0.24–0.60)
11133	58	0.34 (0.15–0.51)	0.33 (0.12–0.46)
21323	52	0.31 (0.20–0.46)	0.30 (0.10–0.43)
23321	51	0.30 (0.11–0.44)	0.32 (0.12–0.45)
22331	55	0.30 (0.14–0.44)	0.29 (0.20–0.44)
21133	51	0.30 (0.20–0.44)	0.28 (0.19–0.44)
22323	63	0.25 (0.11–0.35)	0.28 (0.10–0.44)
33212	48	0.25 (0.15–0.31)	0.23 (0.10–0.30)
23232	46	0.24 (0.11–0.35)	0.20 (0.12–0.35)
23313	53	0.23 (0.12–0.34)	0.25 (0.10–0.42)
22233	51	0.22 (0.13–0.38)	0.20 (0.06–0.31)
32232	62	0.22 (0.10–0.35)	0.18 (0.10–0.40)
13332	53	0.21 (0.11–0.28)	0.20 (0.07–0.26)
32313	53	0.19 (0.08–0.33)	0.21 (0.09–0.35)
32223	47	0.18 (0.03–0.27)	0.11 (0.00–0.23)
33321	47	0.17 (0.04–0.26)	0.20 (0.07–0.34)
32331	44	0.16 (0.10–0.29)	0.20 (0.05–0.30)
33232	58	0.10 (0.05–0.20)	0.10 (0.02–0.20)
33323	51	0.10 (–0.01–0.16)	0.09 (–0.06–0.15)
uncon	212	0.01 (–0.01–0.05)	0.01 (0.00–0.05)
33333	211	0.01 (–0.06–0.06)	0.00 (–0.07–0.05)
death	212	0.00 (—)	0.00 (—)

* Data from nine respondents excluded due to missing or unusable data.

mediate/no education, SC III–V with degree education, SC III–V with intermediate/no education. Controlling for either social class or education reduced, but did not remove, the effect of the other factor. While social class had a greater effect, it appeared that social class and education were both important factors and were acting together.

Test–retest reliability

The 221 respondents in the retest group were representative of the whole sample in all respects except educational level, where more (71.4%) retest respondents had qualifications compared with respondents not in the retest sample (63%, $\chi^2 = 6.26$, df = 1, $p < 0.05$). At test, there were no significant differences ($p < 0.01$) in valuations between those respondents who went on to re-interview and those who did not. Mean logical consistency at retest was 2.2%, slightly less than at test.

The mean ICC was 0.78 (SD = 0.19) and only 13 respondents had an ICC < 0.6. ICCs of the respondents who reported new illness in self or others since the first interview (4% and 13% respectively) were not statistically significantly different from the ICCs of other respondents.

Adjusted median retest valuations are shown in Table 5. There was one logical inconsistency, where state 21133 had a slightly higher median value than state 11133 but should logically have had a lower valuation. Social class and education again influenced health state valuations although the number of significant differences were fewer. Respondents in SC III–V had higher median values than those in SC I–II for seven states ($p < 0.01$), while respondents with intermediate/no education had higher median values than those with degrees for five states ($p < 0.01$). For those states with non-significant differences (*i.e.* $p > 0.01$), these same trends held for all but five states for social class, and all but two states for education.

Discussion

The study was successful in generating VAS valuations for 45 health states from a large representative sample of the British general population. The methods used generated an almost complete and logically consistent VAS data set, and the mean ICC of 0.78 for test–retest reliability is of the same order of magnitude as that reported for TTO by Churchill *et al.*[27] The VAS method proved to be acceptable to a wide range of respondents with no marked difficulties for older

respondents, contrary to previous findings, albeit among patients.[28,29] Although respondents with unusable data tended to be older, these exclusions were due principally to their high valuations for death, rather than to an inability to understand the VAS method.

There is considerable evidence from this study that, although respondents rank health states in roughly the same order using a similar range of values, social class and education have a significant effect on the actual valuations. Higher median valuations are given by respondents in SC III–V and those with intermediate or no educational qualifications. These findings are quite robust in that the trend is the same in all cases and most of the differences reach a significance level of $p < 0.01$. The trends remain after controlling for age and are stronger for the more severe states. It appears from the analysis that social class has a dominant effect. Differing views about the nature of health would not seem to explain this finding — results from two large-scale studies of lay concepts in the general population[30,31] suggest that there is not a class stereotype in which the higher social classes hold more esoteric notions of health such as 'health as well-being' or a 'striving for self-realisation'. The rationale then might be that people in lower social classes have lower expectations of health and thus accept poor health more equably than respondents in higher classes. Lower expectations of health might arise from greater experience of poor health and it was found that significantly more of SC III–V respondents than SC I–II respondents reported problems with mobility (20% and 13% respectively, $p < 0.01$) and usual activities (17% *cf.* 13%, $p < 0.01$) and there was even more difference on pain (36% *cf.* 25%, $p < 0.001$) and mood (23% *cf.* 15%, $p < 0.001$). Psychological factors may play the most important part — Buckingham and Drummond[32] concluded, from a study involving patients with asthma, that the VAS valuations were most affected by variables that reflect mood, such as anxiety and depression. In this study there is tentative evidence that level three on pain/discomfort or anxiety/depression is influential in generating the social class effect but, as noted earlier, self-reported current health status did not have a significant effect on valuations.

So far as is known, neither socio-economic status nor education have previously been found to have a significant effect on valuations for health states[12,33–36] although most of these studies contained only small numbers of respondents. Valuations have been reported to show a striking increase and then decrease with advancing age.[36] As in this study, experience of illness may be a factor of some relevance[34,36,37] but these findings are not consistent.[12,38,39]

If the effect of social class identified here is a real one, it has important implications in the use of health state valuations. Can it be accepted that a representative data set such as this one represents all social groups proportionate to their presence in the general population, and is therefore appropriate for use in resource allocation? Or would this be seen as unwarranted aggregation of data ignoring important individual differences? Perhaps separate tariffs should be produced for different social class groupings. If so, how should these be used — only when the treatments under study are influenced by social class differences? Only when severe states are involved? Ultimately such questions can be addressed through sensitivity analysis. If the evaluation of a treatment based on social tariffs derived from different population subgroups does not differ, then it may be safe to assume that the choice of tariff is immaterial. This may in fact turn out to be the usual case, in view of the overall similarity of scores despite statistically significant differences state by state. Should that evaluation vary according to the choice of subgroup tariff however, then decisions regarding the value of that treatment may have to be based on other criteria. Researchers in this field have an obligation to report any findings which indicate significant differences between population subgroups. Any conflict of evidence may ultimately require a political judgement.

The state 'unconscious' appeared to be valued differently from the other states. Respondents in lower social classes or with low educational levels assigned higher median values to this state (both $p < 0.001$), and it was given a significantly higher median valuation by respondents who were aged under 60 ($p < 0.001$), were single ($p < 0.01$), were smokers ($p < 0.01$), had no experience of illness in others ($p < 0.001$) or who were in paid work (*cf.* retired, $p < 0.001$). Valuations of 'unconscious' may also be related to valuations of 'death', an area which has yet to be fully investigated in this study.

Several methodological issues warrant further investigation. In this paper the effects of background variables were analyzed on a state-by-state basis, requiring large numbers of tests and increasing the possibility that significant results would emerge by chance (hence the use of the rather stringent 1% significance level). It is difficult, given such analysis, to draw robust conclusions about the systematic effects of such variables. For example, the findings with respect to social class were significant for the vast majority but not all of the 45 states. An alternative analytical strategy would be to simultaneously consider respondents' values for the full range of health states, perhaps by estimating a 'fixed effects' model

for each individual with both intercept dummies and slope dummies, to test for any systematic relation with background variables.

The choice of the median or mean as the measure of central tendency has important consequences. The median adjusted valuations are higher than the mean valuations, and no EuroQol state has a negative median valuation at either test or retest. This is despite negative mean valuations for several of the more severe states and the assignment of negative valuations by some individuals, a finding noted elsewhere for states of dementia and coma.[40] The choice of the measure of central tendency has thus far been determined along statistical lines. However it should be recognized that the median explicitly excludes the influence of extreme values, which are filtered out with the implication that they are 'unacceptable'. The use of the mean allows each person's values to contribute equally and maintains this equality even for respondents with outlying values. Thus, any answer is 'acceptable' and extreme values can significantly influence the 'average'.

There is a further question as to whether or not CR methods such as the VAS generate interval data, and how this can be demonstrated. It is generally accepted that CR methods produce an ordinal ranking of states from each individual respondent. Some have argued that an interval scale is also produced,[9,41,42] although others disagree because of a tendency on the part of the respondent to use all the categories of a scale equally often.[43] Assuming that the VAS produces ordinal data only, it is possible to generate an interval scale based on respondents' aggregate data[44] but this is not wholly satisfactory if the effects of individual differences are to be studied. In this study, a process of fractionation was used to encourage respondents to think of the scale as having interval properties. Differing endpoints on the VAS scale may complicate the comparison of results from different studies. Some studies have used specific definitions such as hormone remission[33] or the WHO definition of good health[39] as the top anchor compared to death at the bottom, or a range from minimal to severe dysfunction.[37] Others have used less well defined concepts, such as 'most desirable' and 'least desirable'.[38] Further investigation is needed to determine the level of data that VAS produces and the influence of differing anchor points.

Finally, the value of a health state is likely to be influenced by the length of time spent in the state[36] although this relationship is apparently not a linear one.[45–47] Previous VAS studies have typically not specified a duration, with most requiring respondents to judge the desirability of different scenarios relative to each other at that particular time rather than over a stated period. In order to collect comparable data from the valuation methods used in this study, the same duration was used for ranking VAS and TTO. A duration of ten years was considered to be long enough to represent chronic illness, but not so long as to be unrealistic for older respondents who have relatively short life expectancy. Further work is needed to determine the effect of differing durations on VAS valuations and how such data should be compared.

Despite difficulties reported by others who have used VAS techniques to record valuations, this study has demonstrated a high degree of acceptance by respondents and suggests that the VAS method can perform very successfully in generating health state valuations from the general public. These are powerful arguments for the serious consideration of VAS in a survey context, possibly alongside a choice-based method. Given its ease of use and low resource cost, it is time to reconsider the place of VAS (and other category rating methods) in the study of health state valuations.

Acknowledgements

This study was conducted jointly by the MVH Group at the Centre for Health Economics and Social and Community Research Planning, London. Financial support was received from the Department of Health and the ESRC. The authors would like to thank the interviewers and the participants from the general public who made this study possible, as well as John Cairns for his comments on an earlier draft of the manuscript.

References

1. Patrick DL, Erikson P. *Health status and health policy: quality of life in health care evaluation and resource allocation.* New York: Oxford University Press, 1992.
2. Torrance GW. Measurement of health state utilities for economic appraisal. *J Health Economics* 1986; **5**: 1–30.
3. Hornberger JC, Redelmeier DA, Petersen J. Variability among methods to assess patients' well-being and consequent effect on a cost-effectiveness analysis. *J Clin Epidemiol* 1992; **45**: 505–512.
4. Read JL, Quinn RJ, Bewick DM, Fineberg HV, Weinstein ML. Preferences for health outcomes: comparison of assessment methods. *Med Decis Making* 1984; **4**: 315–329.
5. Daly E, Gray A, Barlow D, McPherson K, Roche M, Vessey M. Measuring the impact of menopausal symptoms on quality of life. *BMJ* 1993; **307**: 836–840.
6. Nord E. The validity of a visual analogue scale in determining social utility weights for health states. *Int J Health Planning Management* 1991; **6**: 234–242.

7. Fernandez E, Turk DC. Sensory and affective components of pain: separation and synthesis. *Psychol Bull* 1992; **112**: 205–217.
8. Fahndrich E, Linden M. Zur Reliabilitat und Validitat der Stimmungsmessung mit der Visuellen Analog-Skala (VAS). *Pharmacopsychiat* 1982; **15**: 90–94.
9. Kaplan RM, Ernst JA. Do category rating scales produce biased preference weights for a health index? *Med Care* 1983; **21**: 193–207.
10. Kaplan RM, Bush JW, Berry CC. Category rating versus magnitude estimation for measuring levels of well-being. *Med Care* 1979; **17**: 501–521.
11. Torrance GW, Boyle HH, Horwood SP. Application of multi-attribute utility theory to measure social preferences for health states. *Operations Res* 1982; **30**: 1043–1069.
12. Wolfson AD, Sinclair AJ, Bombardier C, McGeer A. Preference measurements for functional status in stroke patients: interrater and intertechnique comparisons. In: Kane RL, Kane RA, eds. *Values and Long Term Care*. Lexington, MA: Lexington Books, 1982: 191–213.
13. O'Boyle CA, McGee H, Hickey A, O'Maller K, Joyce CRB. Individual quality of life in patients undergoing hip replacement. *Lancet* 1992; **339**: 1088–1091.
14. Sculpher MJ, Bryan S, Dwyer N, Hutton J, Stirrat GM. An economic evaluation of transcervical endometrial resection versus abdominal hysterectomy for the treatment of menorrhagia. *Br J Obstet Gynaecol* 1993; **100**: 244–252.
15. Llewellyn-Thomas HA, Sutherland HJ, Thiel EC. Do patients' evaluations of a future health state change when they actually enter that state? *Med Care* 1993; **31**: 1002–1012.
16. EuroQol Group. EuroQol: A new facility for the measurement of health-related quality of life. *Health Policy* 1990; **16**: 199–208.
17. Humphreys W, Evans F, Williams T. Quality of life: Is it a practical tool in patients with vascular disease? *J Cardiovascular Pharmacol* 1994; **23(Suppl 3)**: S34–S36.
18. Essink-Bot ML, Bonsel G, van der Maas PJ. Valuation of health states by the general public: feasibility of a standardised measurement procedure. *Soc Sci Med* 1990; **31**: 1201–1206.
19. Brazier J, Jones N, Kind P. Testing the validity of the EuroQol and comparing it with the SF–36 health survey questionnaire. *Qual Life Res* 1993; **2**: 169–180.
20. Nord E. EuroQol: health-related quality of life measurement: valuations of health states by the general public in Norway. *Health Policy* 1991; **18**: 25–36.
21. Brooks RG, Jendteg S, Lindgren B, Persson U, Bjork S. EuroQol: health-related quality of life measurement: results of a Swedish questionnaire exercise. *Health Policy* 1991; **18**: 37–48.
22. Dolan P, Gudex C, Kind P, Williams A. Valuing health states: a comparison of methods. (submitted for publication)
23. Stevens SS. Issues in psychophysical measurement. *Psychol Rev* 1971; **78**: 426–450.
24. Streiner DL, Norman GR. *Health measurement scales: a practical guide to their development and use*. Oxford: Oxford Medical Publications, 1994.
25. OPCS. *General Household Survey*. Series GHS No.23, London: HMSO, 1992.
26. OPCS. *Key population and vital statistics*. Series VS No.17, London: HMSO, 1990.
27. Churchill DN, Torrance GW, Taylor CC, *et al*. Measurement of quality of life in end-stage renal disease: the time trade-off approach. *Clin Invest Med* 1987; **10**: 14–20.
28. Ganz PA, Haskell CM, Figlin RA, la Soto N, Siau J. Estimating the quality of life in a clinical trial of patients with metastatic lung cancer using the Karnofsky Performance Status and the Functional Living Index — Cancer. *Cancer* 1988; **61**: 849–856.
29. Wiklund I, Swedberg K. Some methodological problems in analysing quality of life data in severe congestive heart failure patients. *J Clin Res Pharmacoepidemiol* 1991; **5**: 265–273.
30. van Dalen H, Williams A, Gudex C. Lay people's evaluations of health: are there variations between different subgroups? *J Epidemiol Comm Health* 1994; **48**: 248–253.
31. Blaxter M. Lay concepts of health and survey measures. *Survey Methods Newsletter* Winter 1988/1989.
32. Buckingham K, Drummond N. A theoretical and empirical classification of health valuation techniques. Paper to Health Economists' Study Group meeting, Strathclyde, 1993.
33. O'Connor A, Boyd N, Till J. Influence of elicitation technique, position order and test–retest error on preferences for alternative cancer drug therapy. *Nursing Research: Science for Quality Care, Proc 10th National Nursing Research Conference*. Toronto: University of Toronto, 1985.
34. Rosser R, Kind P. A scale of valuations of states of illness: is there a social consensus? *Int J Epidemiol* 1978; **7**: 347–358.
35. Kaplan RM, Bush JW, Berry CC. The reliability, stability and generalizability of a health status index. *Proc Social Stat Section*. American Statistical Association, 1978; 704–709.
36. Sackett DL, Torrance GW. The utility of different health states as perceived by the general public. *J Chronic Dis* 1978; **31**: 697–704.
37. Carter WB, Bobbitt RA, Bergner M, Gilson B. Validation of an interval scaling: the Sickness Impact profile. *Health Serv Res* 1976; **Winter**: 516–528.
38. Patrick DL, Bush JW, Chen MM. Methods for measuring levels of well-being for a health status index. *Health Serv Res* 1973; **8**: 228–245.
39. Llewellyn-Thomas H, Sutherland HJ, Tibshirani R, Ciampi A, Till J, Boyd NF. Methodologic issues in obtaining values for health states. *Med Care* 1984; **22**: 543–552.
40. 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.
41. Anderson NH. How functional measurement can yield validated interval scales of mental quantities. *J Appl Psychol* 1976; **61**: 677–692.
42. Kaplan RM, Bush JW, Berry CC. Health status: types of validity and the index of well-being. *Health Serv Res* 1976; **Winter**: 478–507.
43. Parducci A. Contextual effects: a range-frequency analysis. In: Carterette EC, Friedman MP, eds. *Handbook of Perception: II. Psychophysical Judgement and Measurement* London: Academic Press, 1974.
44. Torgerson WS. *Theory and Methods of Scaling*. New York: John Wiley & Sons, 1958.
45. Sutherland HJ, Llewellyn-Thomas H, Boyd NF, Till JE. Attitudes towards quality of survival: the concept of maximal endurable time. *Med Decis Making* 1982; **2**: 299–309.
46. Gudex C, Dolan P. Valuing health states: the effect of duration. (submitted for publication)
47. Dolan P, Gudex C. Time preference, duration and health state valuations. *Health Economics* (in press)

(Received 25 September 1995; accepted 12 March 1996)