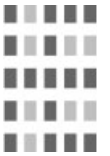
HEALTH ECONOMICS
Health Econ. **13**: 733–737 (2004)
Published online 15 December 2003 in Wiley InterScience (www.interscience.wiley.com). **DOI:**10.1002/hec.875



HEALTH ECONOMICS LETTERS

To what extent do people prefer health states with higher values? A note on evidence from the EQ-5D valuation set

Jennifer Roberts* and Paul Dolan
School of Health and Related Research, University of Sheffield, Sheffield, UK

Summary

The EQ-5D general population valuation set (or 'tariff') is increasingly being used in the evaluation of health care interventions and has been recommended by the National Institute for Clinical Excellence (NICE) for use in cost-utility analyses of health technologies. To be of use to decision-makers, the health gain implied by changes in health state values must reflect individual preferences. At the simplest level, if State A has a higher mean value than State B, then the majority of people should consider a move from B to A to be a good thing. In this paper, we examine the extent to which this is true by re-analysing data from the general population study used to derive the EQ-5D tariff. We show that, on average, the difference in value between two states has to be as large as 0.20 (on a scale where one represents full and zero represents death) for 70% of respondents to agree with the *sign* of that difference (never mind its size). Results such as these have important implications for the use of the EQ-5D tariff that has been generated from these data. Copyright © 2004 John Wiley & Sons, Ltd.

Keywords EQ-5D; health state valuations

Introduction

The EQ-5D is a generic health state classification system designed to allow quality adjustment weights to be attached to each of the health states it generates [1,2]. Values for these health states have been elicited from a representative sample of the UK population and a general population valuation set (or 'tariff') has been generated for the EQ-5D [3]. This tariff is increasingly being used in the evaluation of health care interventions and has been recommended by the National Institute for Clinical Excellence (NICE) for use in cost-utility analyses of health technologies.

The rationale behind these developments is to enable the preferences of the general population to inform resource allocation decisions in health care. If a health care intervention improves some dimensions of health but makes others worse, does it lead to an improvement or deterioration in the overall health state? This is a question that can only be answered by making value judgements; and if the EQ-5D tariff is employed, the assumption is that the relevant values are those of the general public.

The problem, as this short paper will demonstrate, is that the EQ-5D tariff masks a huge amount of individual variation in the values for EQ-5D health states. As would be expected, there are *cardinal* differences; that is, different people value health states differently. However, what may be more surprising is the extent of the *ordinal* differences; that is, there is substantial disagreement as to whether the move from one state to another represents an improvement or deterioration in health, regardless of the actual value given

*Correspondence to: School of Health and Related Research, University of Sheffield, Regent Court, 30 Regents Street, Sheffield S1 4DA, UK. E-mail: j.r.roberts@shef.ac.uk

Received 15 January 2003
Accepted 27 October 2003

to these states. As Sculpher and Gafni have pointed out in this journal [4], the failure to recognise heterogeneity in preferences in economic evaluation is in distinct contrast to the way in which variation in clinical benefits is systematically incorporated into the evaluation of health technologies using sub-group analyses.

Using the health state valuation data originally collected to generate the tariff, which has been reported in this journal [5], this paper argues that ignoring variation in individual values can, despite the use of the EQ-5D tariff, result is decision making that does not adequately reflect people's preferences. In particular, taking an individual from a state with one tariff value to a state with a higher tariff value might actually make that individual worse off if s/he did not agree with ordering of those states.

Methods

The EQ-5D defines health on five dimensions: mobility, self-care, usual activities, pain/discomfort and anxiety/depression. Each dimension has three levels of severity which, broadly speaking, correspond to 'no problems', 'some problems' and 'extreme problems', and so 243 health states can be defined.

Three thousand three hundred and ninety five members of the UK general population took part in the original valuation survey, and the 2997 respondents with complete data are included in this analysis (which is the same data used to generate the tariff). Each respondent valued 12 EQ-5D states (drawn from a subset of 42) using the TTO method and the full details are given in Dolan *et al*. [5]. The states valued in the study were chosen to include as many combinations of levels across the five dimensions as possible, subject to the constraint that the states should be plausible to respondents. All respondents valued the most severe health state (33333) plus eleven others drawn from four groups defined according to their severity. Each respondent valued two 'very mild' states and three states from each of 'mild', 'moderate' and 'severe' categories. There are around 1200 values for each of the very mild states and 750 values for each of the other states.

The original purpose of the survey was to estimate a model to predict the values of all 243 EQ-5D health states from direct values on a subset of 42 states [3]. However, given this study design, it is also possible to consider the differences in valuations given by respondents to particular pairs of states. This is useful because, while we expect considerable variation across individual valuations of states, it may be that respondents perceive the differences between particular pairs of states to be similar. Consider the simplified schematic in Figure 1. Respondent 1 values states A and B more highly than Respondent 2, but the difference in the valuations of states A and B is the same for both respondents. With 42 states valued in the survey, there are 861 possible pair wise comparisons (i.e. $(42 \times 41)/2$) with an average of 230 observations on each pair of states.

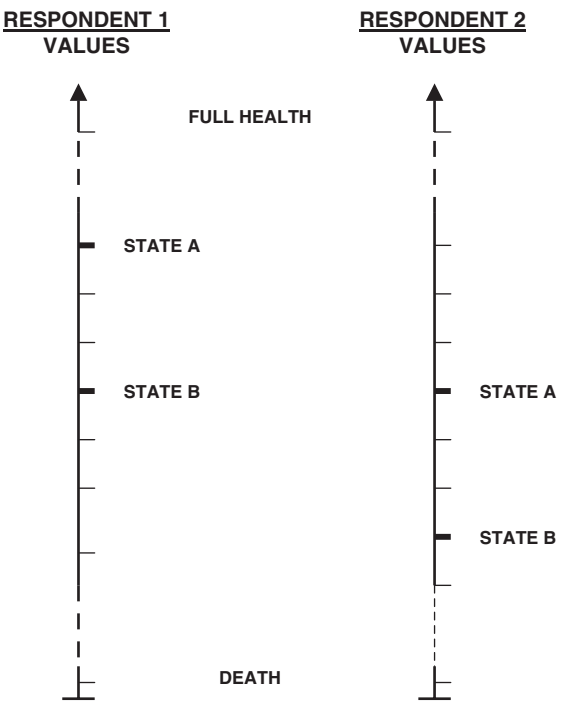


Figure 1. Differences in health state values across respondents. Respondents 1 and 2 give very different values to states A and B but the difference between the values is the same

Results

The original model used to estimate a tariff for the EQ-5D took account of individual heterogeneity by employing a random effects specification to

predict the mean values for each state [3]. This model predicted the mean values for the 42 states valued in the study pretty well. The mean absolute prediction error between the actual and estimated values was 0.039 (S.D. = 0.029), and for only three states was the error in excess of 0.1. However, as shown in Table 1, there is enormous variation across individual valuations of each state. Table 2 shows descriptive statistics for an illustrative sample of 9 of the 861 possible pairs of states. The standard deviations of the differences in valuations are as large as those surrounding the mean values themselves. None of these pairs have a logical ordering – some dimensions are better in State B than State A and some are worse – and so these states represent outcomes that reflect the kind of comparisons that heath care decision makers may face in practice.

Taking the third pair of states as an example, the mean difference in valuation is −0.157,

Table 1. Mean and standard deviation for all states valued in the study

State	Mean	S.D.	Interquartile Range	
21111	0.878	0.226	0.850	1.000
11211	0.869	0.222	0.825	1.000
11121	0.850	0.242	0.800	1.000
12111	0.834	0.287	0.800	1.000
11112	0.829	0.286	0.753	1.000
12211	0.767	0.321	0.625	1.000
12121	0.742	0.315	0.600	1.000
11122	0.722	0.365	0.600	1.000
22112	0.665	0.372	0.500	0.950
22121	0.642	0.421	0.500	0.925
21222	0.553	0.454	0.425	0.925
11312	0.552	0.466	0.400	0.925
12222	0.551	0.458	0.375	0.925
22122	0.540	0.467	0.400	0.925
21312	0.536	0.464	0.375	0.925
22222	0.500	0.478	0.325	0.875
11113	0.392	0.553	0.000	0.875
13212	0.389	0.532	0.175	0.800
13311	0.346	0.555	0.000	0.775
12223	0.216	0.560	−0.275	0.647
11131	0.200	0.604	−0.334	0.700
21323	0.160	0.588	−0.375	0.600
32211	0.152	0.593	−0.375	0.641
23321	0.147	0.607	−0.375	0.650
21232	0.064	0.602	−0.475	0.550
22323	0.042	0.583	−0.450	0.525
22331	−0.011	0.597	−0.525	0.500
33212	−0.022	0.593	−0.500	0.475
11133	−0.049	0.607	−0.575	0.475
21133	−0.063	0.594	−0.600	0.475
23313	−0.070	0.586	−0.550	0.400
23232	−0.084	0.583	−0.575	0.425
33321	−0.120	0.566	−0.625	0.375
22233	−0.142	0.568	−0.625	0.350
32313	−0.152	0.563	−0.625	0.300
32223	−0.174	0.563	−0.675	0.275
32232	−0.223	0.572	−0.725	0.250
13332	−0.228	0.551	−0.700	0.200
32331	−0.276	0.549	−0.775	0.025
33232	−0.332	0.509	−0.750	0.000
33323	−0.386	0.492	−0.825	−0.025
33333	−0.543	0.412	−0.925	−0.275

Table 2. Descriptive statistics for differences in valuations of EQ-5D health states

State A	State B	*N*	DIF (A–B)	S.D.	% NEG	99% CI	
22112	11211	321	−0.221	0.368	91.3	87.2	95.4
11312	11122	148	−0.188	0.463	81.1	72.8	89.4
22331	21323	165	−0.157	0.530	69.1	59.8	78.4
11122	21111	307	−0.128	0.323	85.3	80.1	90.5
33321	21133	190	−0.092	0.635	60.0	50.8	69.2
11211	21111	289	−0.020	0.220	74.4	67.8	81.0
12121	12211	139	0.014	0.235	70.5	60.5	80.5
23321	12223	198	−0.014	0.547	58.6	49.6	67.6
11131	12223	185	0.011	0.593	59.5	50.2	68.8

N = number of observations on pair of states (i.e. both State A and State B valued by same individual).
DIF (A–B) = mean value for State A minus mean value for State B (from those individuals who valued both states).
S.D. = standard deviation of DIF(A–B).
% NEG = percentage of all *N* differences that are negative i.e. where State B is given a higher value than State A.
99% CI = 99% confidence interval for %NEG.

suggesting that, on average, respondents valued state 21323 as higher than state 22331. This suggests that having extreme anxiety/depression is (on average) more than compensated for by being better off in terms of self-care and pain. However, the standard deviation of this difference is 0.53 and just over 30% of respondents valued state 22331 as higher than state 21323. So, for these people, the deterioration in anxiety/depression is more serious than the alleviation of the self-care and pain problems. Ninety nine percent confidence levels are calculated for the proportion of people agreeing with the ranking of states and in no case in Table 2 does the interval contain everybody. Overall, only 155 of the 861 comparisons have confidence intervals that include all respondents.

Figure 2 illustrates the differences in values of pairs of states for all 861 possible comparisons. The graph shows the mean difference in valuation against the percentage of people agreeing with the direction of the difference i.e. agreeing that State A is better (or worse) than State B. This figure shows the lack of consensus surrounding the ordering of the health states. For example, even with a mean difference in valuation of 0.20 on a 0 to 1 scale, only around two-thirds of people agree with the ordering of the states (slightly more if the states have a logical order and slightly less if they do not).

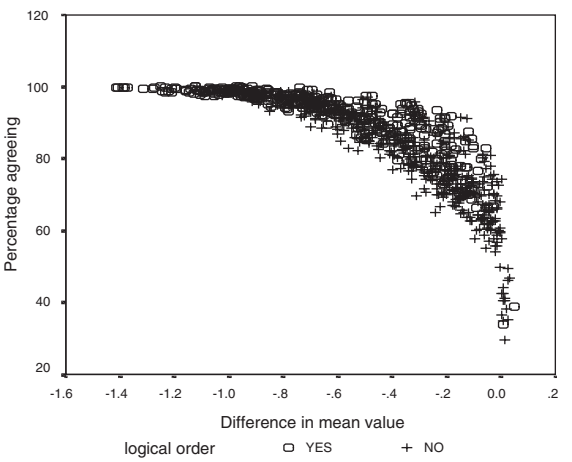


Figure 2. The difference in mean valuation and the percentage of respondents agreeing with the direction of the difference

Discussion

The tariff derived for the EQ-5D is assumed to represent the average values of the UK general population – but it is an aggregation of individual preferences that masks enormous variation. While the UK general population, on average, deems that state 21111 is better than 11122 by a distance of 0.13, 15% of respondents did not agree with this ranking of health states let alone the extent of the difference. And so the move from state 21111 to state 11122 is viewed as a deterioration in health by a significant minority of people. The results of our analysis suggest that the cardinal difference between states must be as much as 0.20 for two-thirds of the population to agree with the ordinal ranking.

The confidence intervals cited in the Results section give some indication that the level of disagreement with the ordering of states is 'significant', given that very few of the intervals span 100% of respondents. However, the level of public agreement that is required by decision makers in order to have confidence that decisions are reflecting the population's values is open to debate. If we take a simple majority rule then this criteria is only violated for 17 of the 861 pairwise comparisons discussed here but if a two-thirds majority is required, then 116 pairs fail to satisfy this. And, unfortunately, the health states that need to be compared in real world decision making, for interventions with sometimes very small benefits, are much more likely to reflect those where there is less than two-thirds agreement on the ordering.

These results have important implications for the use of the EQ-5D tariff of values that has been generated from these data. Indeed, we suspect that this problem is not unique to the EQ-5D tariff and that others, such as the recently published SF-6D tariff [6], will mask similar levels of individual variation. Evaluations of health technologies that use the EQ-5D to generate QALYs will often produce health state improvements of 0.2 or less. The recent NICE technology appraisal of etanercept and infliximab for rheumatoid arthritis [7] was informed by evidence that estimated that the two drugs in question resulted in gains (per year) of 0.12 on the EQ-5D scale [8]. Similarly, evidence from the forthcoming HTA report on continuous subcutaneous insulin infusion (Colquitt JL, Green C, Sidhu MK, Hartwell D, Waugh N. Clinical and cost effectiveness of continuous subcutaneous insulin infusion for diabetes. *NHS HTA* (http://www.nice.org.uk/Docref.asp?*d*= 38421)) estimated that a severe hypoglycaemic event for people with diabetes corresponded to a decrement of 0.15 on the EQ-5D scale [9]. If the health states that people move between in these comparisons have no logical ordering on the descriptive system used, our results suggest that only around 60% of the general public might agree with the ordering of such changes in health.

The heterogeneity displayed in the pairwise comparisons reported in this paper is pervasive and we do not believe it is overcome by the preference sub-group approach suggested by Schulpher and Gafni [4]. We therefore advise that the results of cost-utility analyses using these tariffs should be treated with extreme caution at the individual patient and small group level.

References

1. Brooks R, with the EuroQol Group EuroQol: the current state of play. *Health Pol* 1996; **37**: 53–72.
2. Kind P, Dolan P, Gudex C, Williams A. Variations in population health status: results from a United Kingdom national questionnaire survey. *Br Med J* 1998; **316**: 736–741.
3. Dolan P. Modelling valuations for EuroQol health states. *Med Care* 1997; **35**: 1095–1108.
4. Sculpher M, Gafni A. Recognizing diversity in public preferences: the use of preference sub-groups in cost-effectiveness analysis. *Health Econ* 2001; **10**: 317–324.
5. Dolan P, Gudex C, Kind P, Williams A. The Time Trade-Off Method: Results from a General Population Study. *Health Econ* 1996; **5**(2): 141–154.
6. Brazier J, Roberts J, Deverill M. The estimation of a preference based measure of health from the SF-36. *J Health Econ* 2002; **21**: 271–292.
7. NICE Guidance on the use of etanercept and infliximab for the treatment of rheumatoid arthritis. *Technol Appr Guid* 2002; No. 36.
8. Jobanputra P, Barton P, Bryan S, Fry-Smith A, Burls A. The clinical effectiveness and cost effectiveness of new drug treatments for rheumatoid arthritis: etanercept and infliximab. *NHS HTA* 2001; **6**(21).
9. Nordfeldt S, Jonsson D. Short-term effects of severe hypoglycaemia in children and adolescents with Type 1 diabetes. A cost of illness study. *Acta Paediatr* 2001; **90**(2): 137–142.

