PERGAMON

Social Science & Medicine 54 (2002) 919–929

SOCIAL SCIENCE & MEDICINE

www.elsevier.com/locate/socscimed

To what extent can we explain time trade-off values from other information about respondents?

Paul Dolan[a,b,*], Jennifer Roberts[c]

[a] Sheffield Health Economics Group and Department of Economics, University of Sheffield, 30 Regent Street, Sheffield S1 4DA, UK
[b] Department of Economics, University of Oslo, Norway
[c] Sheffield Health Economics Group, University of Sheffield, 30 Regent Street, Sheffield S1 4DA, UK

Abstract

The time trade-off (TTO) is one of the most widely used health state valuation methods and was recently used to develop a set of values for the EQ-5D descriptive system from 3000 members of the UK general population. However, there is currently very little understanding of precisely what determines responses to TTO questions. The data that were used to generate this set of values are ideal for addressing this question since they contain a plethora of information relating to the respondents and their cognition during the TTO exercise. A particularly useful characteristic of this dataset is the existence of visual analogue scale (VAS) valuations on the same states for the same respondents. The results suggest that age, sex and marital status are the most important respondent characteristics determining health state valuations. The VAS valuations were found to add very little to the explanatory power of the models.

Keywords: Health state valuation; Time trade-off; Visual analogue scale; EQ-5D

Introduction

There can be little doubt that preference-based measures of health status will increasingly be used to evaluate the outcomes of health care interventions and to inform resource allocation decisions. A question that may have important implications for how resources are allocated is whether preferences over health states differ according to the characteristics of the respondent. If major differences exist between groups of raters, then this poses the problem for policy-makers of whose preferences are to be given the greatest weight when it comes to assessing the outcomes of health care.

It is common for the preferences of the whole population to be considered the most relevant when comparing interventions that affect different population sub-groups (Gold, Russell, Siegal, & Weinstein, 1996). However, where there exists a pre-determined budget for a particular population sub-group (for example, the elderly), then it might be appropriate to use only the values of those within that sub-group. Although the decision about whose values to use is ultimately a philosophical one, empirical evidence about the extent to which values differ according to the characteristics of the respondent will make clear the implications of using the values of different groups. The primary purpose of this paper is to provide precisely this kind of evidence.

In generating data to shed light on this issue, there are questions concerning how health status should be described and subsequently valued. To allow comparisons across different programmes which may impact upon different dimensions of health, the chosen health state descriptive system must allow for the different dimensions to be combined to form an overall single index. There now exist a number of descriptive systems which have been specifically designed for this purpose (Brazier, Deverill, & Green, 1999).

Valuations for states of health so described can be elicited in a number of different ways. Economists have

*Corresponding author. Sheffield Health Economics Group and Department of Economics, University of Sheffield, 30 Regent Street, Sheffield S1 4DA, UK.
E-mail address: p.dolan@sheffield.ac.uk (P. Dolan).

PII: S0277-9536(01)00066-1

tended to favour methods which ask respondents to value changes in health using parameters of the utility function, such as the risk of good and bad health in the case of the standard gamble (SG) and life years spent in good and bad health in the case of the time trade-off (TTO). By assuming that the value of a health improvement is linearly related to risk or life years, it is possible to provide information about the relative weight that respondents attach to different dimensions of health.

According to expected utility theory (EUT), a cardinal utility can be expressed as a linear function of the risk involved in a gamble if certain assumptions, such as transitivity and independence, hold (von Neumann & Morgenstern, 1953). This has led many to regard the SG as the "gold standard" for health status measurement (Gafni, 1994). However, considerable doubt has been cast on EUT both as a positive and as a normative theory (Camerer, 1993). Richardson (1994) has argued in favour of the TTO on the grounds that it collapses the relationship between a health state, its duration and its value into one single measure. However, for a response to a TTO question to provide a direct and unbiased estimate of health state value, it is necessary that there is no discounting of future utilities (Dolan & Jones-Lee, 1997).

It is difficult, then, to choose between SG and TTO on theoretical grounds alone since valuations from neither method can automatically be assumed to map directly onto utility. The few empirical studies that have directly compared the two methods have produced mixed results although Dolan, Gudex, Kind, and Williams (1996a, b) found that the TTO produced more complete and slightly more reliable data than the SG. It is valuations using the TTO method that are the focus of attention in this paper.

Health state valuations can also be elicited using the visual analogue scale (VAS) which simply requires respondents to give each health state a score, usually between 0 and 100. Although the VAS method does not require people to make trade-offs, it does have the practical advantage of being relatively simple to administer since it lends itself much more to self-completion than either of the other methods. As a result, the VAS is widely used in valuation studies, including the one reported here. The secondary purpose of this paper, then, is to assess whether the inclusion of VAS values adds to the explanatory power of the model used to explain TTO values.

The data that are reported in this paper have been generated using the EQ-5D descriptive system (see Fig. 1), a single index measure which has been widely used in Europe (Brooks, 1996). In the largest valuation study of its kind, TTO valuations for a subset of EQ-5D health states were elicited from over 3000 members of the general public (Gudex, Dolan, Kind, & Williams, 1997). Because it is not feasible to elicit direct valuations for all of the 243 health states generated by the EQ-5D, a general population 'tariff' value for all EQ-5D states has been estimated from direct observations on 42 of these states (Dolan, 1997). The results from this study suggest that TTO valuations are primarily affected by the age of the respondent and, as a result, separate tariffs have been generated for the under and over 60s (Dolan, 2000).

This paper addresses some of these issues further. At the time when the general population tariff was estimated, it was not possible to simultaneously account for the fact that groups of valuations came from the same respondent and the truncated nature of the data. In addition, the dataset contains information on a wide range of background variables which has not been fully exploited and which can help to shed light on the importance of certain respondent characteristics. VAS values for the same health states are also added as explanatory variables.

In addition to modeling the full distribution of TTO values, this paper also gives consideration to whether valuations for states rated better than dead and worse than dead should be treated separately and whether respondents might differ according to whether or not they consider any states to be worse than dead. Firstly, the data are partitioned according to whether the TTO value is positive or negative. Since the procedures for valuing states rated as better and worse than dead are different from one another, it is possible that some background characteristics (as well as the VAS scores) might have a different effect on the two sets of values. Secondly, respondents are partitioned according to whether or not all their TTO values are positive. Using PROBIT models, it is possible to investigate whether any background characteristics can determine which group a respondent belongs to. Finally, separate models are estimated for these two groups of respondents.

The tariff reported in Dolan (1997) is increasingly being used in evaluative studies and is now being recommended by the UK's National Institute for Clinical Excellence for use in cost-utility studies. The analyses reported in this paper will highlight: (1) whether the general population tariff requires an alternative specification; (2) if tariffs in addition to those generated for age and sex are required; (3) the extent to which information regarding how respondents value health states using the VAS helps to explain the tariff values; (4) whether separate tariffs are required for states rated as better and worse than dead; and (5) whether respondent characteristics can explain why some people regard some states as worse than dead and others do not.

Study design

Piloting had shown that no one respondent could be expected to value more than about 12 EQ-5D states

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 11223 means:

1 No problems walking about
1 No problems with self-care
2 Some problems with performing usual activities
2 Moderate pain or discomfort
3 Extremely anxious or depressed

Fig. 1. The EuroQol descriptive system.

using the TTO method. Since this number was deemed to be too small to interpolate valuations for all 243 EQ-5D, a larger set of 42 states was chosen and each respondent was asked to value a subset of these. The most important consideration when choosing the states was that they should be widely spread over the valuation space so as to include as many combinations of levels across the five dimensions as possible. This was subject to the constraint that the states were likely to be considered plausible by respondents. Table 1 shows the set of states chosen for direct valuation and how a subset of these was chosen for each respondent.

Respondents were first asked to describe their own health using the EQ-5D. They were then asked to rank a predetermined set of health states in order from best to worst. It was explained that each state was to be regarded as lasting for 10 years without change, followed by death. Respondents were then asked to indicate where on a VAS with endpoints of 100 (best imaginable health state) and 0 (worst imaginable health state) they would rate each of the states. The health states were then valued by the TTO method using a double-sided board. One side was relevant for states that were regarded by the respondent as better than dead, and the other side for states that were regarded as worse than dead. In the former case, respondents were led by a process of "bracketing" to select a length of time, x, in the 11111 state that they regarded as equivalent to 10

Table 1
Health states valued in the study

Each respondent valued 33333
plus
2 from 5 "very mild" states:
11112 11121 11211 12111 21111
plus
3 from 12 "mild" states:
11122 11131 11113 21133 21222 21312 12211 11133 22121 12121 22112 11312
plus
3 from 12 "moderate" states:
13212 32331 13311 22122 12222 21323 32211 12223 22331 21232 32313 22222
plus
3 from 12 "severe" states:
33232 23232 23321 13332 22233 22323 32223 32232 33321 33323 23313 33212

years in the target state. Respondents were given an opportunity to refuse to trade-off any length of life in order to improve its quality. In the case of states worse than dead, the choice was between dying immediately and spending a length of time $(10 - x)$ in the target state followed by x years in the 11111 state.

It appears that a 0.05 difference between health state values would be considered meaningful in most decision contexts (O'Brien & Drummond, 1994). Based on previous knowledge about the standard deviation around valuations, a sample of about 3000 would be required to detect such a difference between states. To achieve this, Social and Community Planning Research (SCPR) selected 6080 addresses from the postcode address file. The fieldwork was carried out by 92 trained interviewers between August and December 1993 (Gudex et al., 1997).

Methods

In total, 3395 respondents were interviewed but, so as to maintain comparability with the earlier analysis which generated the general population tariff, only the 2997 respondents with complete TTO and VAS data have been included in the analysis reported in this paper (Dolan, 1997). Table 2 shows that the full sample is highly representative of the general population and that excluding those respondents with incomplete data does not compromise the representativeness of the sample in terms of sex, age, education, social class, marital status and own health. In addition, the mean VAS and TTO valuations for all states, their associated standard deviations, and the percentage of the sample rating states as worse than dead on each method are all unaffected by whether those with missing data are included or not.

Table 2
Characteristics of the sample (figures are percentages)

Characteristic	Full sample ($n = 3395$)	After exclusions ($n = 2997$)	GHS[a]
Sex			
Male	43	43	47
Female	57	57	53
Age			
18–34	31	32	31
35–49	25	25	27
50–59	14	14	15
60+	31	30	28
Education			
Degree	9	9	8
Higher	11	11	10
A/O levels	40	41	45
None	37	37	35
Foreign/other	3	3	3
Social class			
I, II	29	30	30
III Non-manual	24	24	22
III Manual	20	21	21
IV, V	25	25	21
Other	1	1	3
Marital status			
Single	17	17	21
Married	60	60	64
Widowed	13	12	9
Divorced	10	11	6
Problems on			
Mobility	18	18	—
Self-care	4	4	—
Usual activities	16	16	—
Pain/discomfort	33	33	—
Anxiety/depression	21	21	—

[a] UK General Household Survey in 1993.

If full health and dead are assigned scores of 1 and 0 respectively, then for states that are rated as better than dead on the TTO, scores are given by the formula $x/10$ where x is the number of years spent in full health. For states that are rated as worse than dead, the score is given by the formula $-x/(10 - x)$. Thus, negative scores lie on a ratio (not an interval) scale and are theoretically bounded by $-\infty$ (though in this study, given the response categories available to respondents, they are bounded by -39). Of course, this asymmetry between positive and negative values gives great weight to negative scores when the mean values for each health state are calculated. In addition, values generated using a ratio scale can lead to biases in observers' judgements (Eyman, 1967).

As a result, values for states rated as worse than dead have been calculated using the formula $(x/10) - 1$, so that the values are bounded by -1, just as states which are better than dead are limited by a value of 1 for full health. Whilst it may well be appropriate to assume that negative values, like positive ones, lie on a linear scale (Poulton, 1989), it should be recognised that this transformation is not supported by EUT and is only one of many possible transformations. However, this was the transformation used in the earlier analysis to generate the tariff and is one that has been used elsewhere in the literature (Patrick, Starks, Cain, Uhlmann, & Pearlman, 1994; Dolan & Sutton, 1997).

Because raw VAS scores have different endpoints (i.e. best and worst imaginable health state), it is necessary to transform them onto the same scale as the TTO scores. This is done in the following way:

$$h_t = (h_r - \text{dead})/(11111 - \text{dead}),$$

where h_t and h_r are the transformed and raw VAS values, respectively, and 11111 and dead are the raw VAS values for those states. Hence, VAS scores lie within the same range as TTO scores (i.e. between -1 and 1). If $h_r > 11111$, then $h_t = 1$ and if $h_t < -1$, then $h_t = -1$.

Analysis at the individual level is complicated by the fact that each respondent valued 12 health states and thus it is reasonable to assume that those 12 scores are related to one another. This means that the variance of the error term is likely to be partly determined by the individuals who value the health states and is therefore unlikely to be constant. To address this, the random effects (RE) model is used, in which there is an overall intercept and an error term with two components; $e_{it} + u_i$. The e_{it} is the traditional error term unique to each observation and u_i is an error term representing the extent to which the intercept of the ith respondent differs from the overall intercept. This model assumes that the individual-specific error term is normally independently distributed which, given the size of the sample, seems a reasonable assumption.

The full set of TTO values is shown in Fig. 2, from which it can be seen that the scores are not normally distributed across the valuation range (for mean values for each of the health states valued in the study see Dolan et al., 1996a, b). Three classes of alternative functional forms are considered in an attempt to account for the highly skewed and truncated nature of the distribution. The first class represents power ladder and Box-Cox transformations which attempt to adjust for the skewness in the distribution. The second class is taken from the ad hoc transformations suggested by Abdalla and Russell (1995) in order to map the data from the range $(-1, 1)$ to the range $(-\infty, \infty)$ via the unit range $(0, 1)$. Finally, Tobit estimation is undertaken with censoring at both the top and bottom ends in order to take account of the fact that TTO valuations are bounded within the range -1 to 1. All of these transformations allow for an RE specification. In further steps, the data are partitioned according to whether the TTO value is positive or negative and whether respondents value all states as better than dead or at least one state as worse than dead. PROBIT models are estimated to distinguish between these two groups of respondents and separate TTO models are estimated for each group.

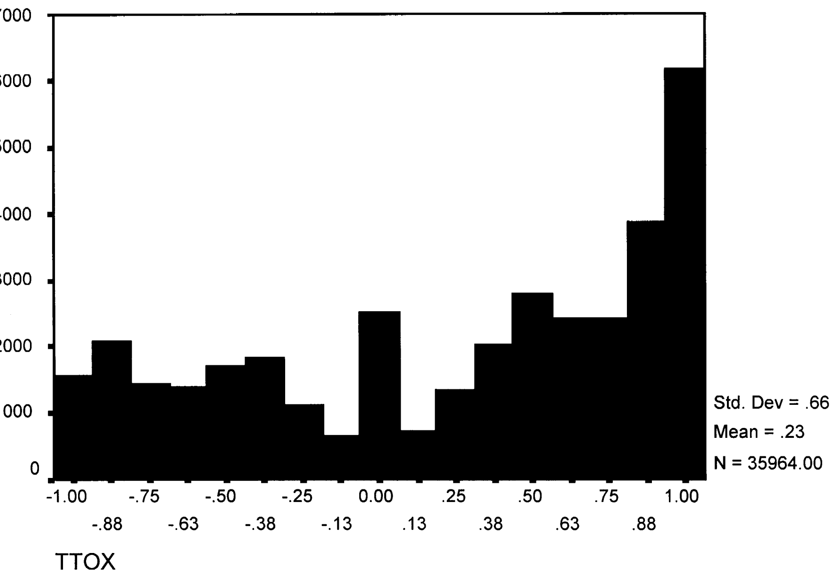

Fig. 2. Distribution of TTO value.

Table 3
Variable definitions

Variable	Definition
TTO	TTO valuation (in the range −1–1)
VAS	VAS valuation (in the range −1–1)
Health state dummies	
MOB2	Dummy = 1 when *mobility* is at level 2, 0 otherwise
MOB3	Dummy = 1 when *mobility* is at level 3, 0 otherwise
SC2	Dummy = 1 when *self-care* is at level 2, 0 otherwise
SC3	Dummy = 1 when *self-care* is at level 3, 0 otherwise
UA2	Dummy = 1 when *usual activity* is at level 2, 0 otherwise
UA3	Dummy = 1 when *usual activity* is at level 3, 0 otherwise
PAIN2	Dummy = 1 when *pain* is at level 2, 0 otherwise
PAIN3	Dummy = 1 when *pain* is at level 3, 0 otherwise
MOOD2	Dummy = 1 when *mood* is at level 2, 0 otherwise
MOOD3	Dummy = 1 when *mood* is at level 3, 0 otherwise
N3	Dummy = 1 when *any dimension* is at level 3, 0 otherwise
Other respondent characteristics	
AGE	Age of respondent in years
SEX	Sex of respondent: 0 = male, 1 = female
MARSTAT	Marital status of respondent: 0 = single, separated, divorced or widowed, 1 = married or cohabiting
XUA	Dummy for usual activities dimension of own health: 1 if level 2 or 3, 0 otherwise
AGEVAS	Interaction between AGE and VAS
SEXVAS	Interaction between SEX and VAS

There are three groups of explanatory variables (see Table 3 for a full set of definitions). First, there is a core group of variables that describes the health state being valued. This is made up of a set of binary dummies that represent the level of each EQ-5D dimension. For all dimensions, level 1 acts as the baseline, hence the expected coefficients are negative and should be greater (in absolute value) for level 3 dummies than level 2 ones. The use of binary dummies is slightly different from the method employed by Dolan (1997) but the two methods are formally equivalent. Following Dolan (1997), this group of variables also includes the 'N3' dummy, which equals 1 if any dimension is at level 3. The constant term represents the valuation of state 11111, and therefore its expected value is unity. However, extrapolation problems, arising from the fact that state 11111 is not part of the estimation, means that the interpretation of the constant is not straightforward. Here, as in the previous analysis, the difference between unity and the constant represents the dysfunction associated with any move away from full health.

Second, there are variables that represent respondent characteristics. Information was available on a wide range of background variables including age, sex, marital status, smoking behaviour, employment status, education level, socio-economic status and whether the respondent had children or not. The results from a number of earlier studies suggest that variation among population subgroups is not explained by the different demographic characteristics of respondents (Froberg & Kane, 1989). However, TTO valuations from this study have been shown to differ by age (Dolan, 2000) even if some of these differences might be an artefact of the TTO method rather than due to real differences in underlying preferences (Robinson, Dolan, & Williams, 1997). The effect of age is explored more fully here. For example, age has been transformed into life expectancy by using life table information for men and women. Since the TTO method requires people to trade-off future years of life, the different life expectancies of men and women of the same age, particularly at advanced ages, might affect valuations. More generally, the large sample in this study relative to earlier studies might reveal differences by other respondent characteristics and any important interactions between them.

There is some, but by no means unambiguous, evidence to suggest that valuations are influenced by respondents' current health status. Kind and Dolan (1995) and Dolan (1996) found that those in poorer health generally gave higher values to the same health states. However, studies by Llewellyn-Thomas et al. (1984) and Hadorn and Uebersax (1995) found that respondents own health did not influence their valuations. In this study, there was data relating to how a respondent rates themselves on the EQ-5D dimensions and how they value their current health on the VAS. Because a respondent's own health might affect their valuations, own health was categorised in a number of different ways. These categories included: 'problems' or 'no problems' on each EQ-5D dimension separately and all dimensions combined; 'mild', 'moderate' or 'severe' health as defined according to the categorisation of states in Table 1; VAS rating of own health as a continuous and as a categorical variable; various combinations of the EQ-5D description and VAS valuation of own health. There was also data relating to whether respondents have worked in a job which involves looking after ill people, and whether they have experience of serious illness in themselves or those close to them.

In addition, there was information about whether the respondent thought of themselves or someone else when answering the questions and whether their valuations depended on something that was to happen to them within the next 10 years. There was also the interviewers rating of how difficult the respondent found the TTO task. Finally, there are the VAS values of the same health states as those valued in the TTO exercise. Squared and cubed transformations of the VAS score were also used as explanatory variables to allow for more flexible functional forms.

The significance or otherwise of the background characteristics was determined using a standard general-to-specific approach in which an unrestricted model (i.e. including all variables) was first estimated (Hendry, 1987). Variables were then omitted using a series of individual and joint tests for significance. The final restricted model is tested against the original unrestricted model. Given the large sample size and the possibility of drawing erroneous conclusions, $t(0.01)$ is judged to be an appropriate significance level. All models were estimated using STATA (v6.0).

Results

Model 1 in Table 4 is a random effects model for the entire sample of TTO valuations and includes only the binary dummies representing each health state and the N3 term. This model is analogous to the one presented in Dolan (1997). This relatively simple model performs well in that all coefficient estimates have the expected sign and size and explanatory power is good at 0.459. Predictive capacity is judged by the model's ability to predict the actual mean TTO values of each of the 42 states valued in the study. The mean absolute prediction error is 0.039, with only three errors greater than 0.10 and 12 greater than 0.05. The mean prediction error is not statistically different from zero (as shown by the t-test in the bottom row of the table). In addition, the prediction errors are normally distributed as indicated by the Jarque–Bera (JBRES) test and show no autocorrelation when the prediction errors are ordered in descending order of actual mean TTO score (Ljung–Box (LB) test). In other words, the prediction errors are random.

The various transformations of the TTO values were ultimately designed to improve predictive ability but every one resulted in worse predictive ability than the untransformed model. For example, the Tobit transformation resulted in large prediction errors: 28 greater than 0.05 and 13 greater than 0.10. The predictions from the Tobit model also showed signs of bias ($t = -4.726$, $p = 0.000$) and autocorrelation ($LB = 4.76$, $p = 0.039$). The other transformations also resulted in comparable prediction errors and showed similar levels of bias and autocorrelation. Given the failure of these techniques and the relatively good performance of the untransformed core model, transformations of the dependent variable are not pursued further in this paper.

Model 2 in Table 4 contains those respondent characteristics that were found to have a significant effect on TTO values at $t_{0.01}$. The resultant model contains relatively few variables: only age, age-squared, sex, marital status and the usual activities dimension of own health have significant coefficient estimates. The inclusion of life expectancy instead of age yields very similar results and has no effect on the coefficient on sex. The various manipulations of the respondents description and valuation of their own health did not have any significant effect on valuations. It can be seen that the coefficients on the core variables are robust to the inclusion of the significant respondent characteristics and that these characteristics have only a small effect on the explanatory power of the model. The coefficients on the respondent variables suggest a quadratic relationship between age and TTO values such that, ceteris paribus, values reach a peak at age 45 and the values given by an 18 year-old are the same as those given by a 72 year-old. The models also suggest that being male, married or having some problems on the usual activities dimension results in higher valuations.

The two other models in Table 4 contain the respondents VAS valuations of the health states. Model 3 is the core model with the addition of the respondents VAS scores and Model 4 contains the significant respondent characteristics as well as the significant interactions between VAS scores and the age and sex of the respondent. The coefficient estimates on the remaining variables are substantially reduced by the inclusion of VAS scores, which attract a large positive coefficient and increase the explanatory power of the model. The inclusion of VAS scores in Model 2 results in a slight improvement in predictive ability in that the mean absolute prediction error is 0.034, seven errors are greater than 0.05, and five are greater than 0.10. However, it is unclear how the VAS values should be entered into the model since squared and cubic terms also fitted the data equally well.

For all the models in Table 4, the Breusch-Pagan test confirms the importance of individual effects and the Hausman test suggests the random effects specification is

Table 4
Random effects models for entire sample of adjusted TTO values (TTO)[a]

	Core	Plus respondent characteristics	Plus VAS	Plus respondent characteristics and VAS
	(1)	(2)	(3)	(4)
C	0.919	0.769	0.643	0.538
	(0.005)	(0.046)	(0.012)	(0.047)
MOB2	−0.068	−0.069	−0.045	−0.046
	(0.005)	(0.005)	(0.005)	(0.005)
MOB3	−0.313	−0.313	−0.253	−0.254
	(0.008)	(0.008)	(0.007)	(0.007)
SC2	−0.104	−0.103	−0.072	−0.073
	(0.005)	(0.005)	(0.005)	(0.005)
SC3	−0.213	−0.213	−0.166	−0.166
	(0.007)	(0.007)	(0.007)	(0.007)
UA2	−0.036	−0.036	−0.026	−0.026
	(0.005)	(0.005)	(0.006)	(0.006)
UA3	−0.094	−0.095	−0.068	−0.068
	(0.007)	(0.007)	(0.007)	(0.007)
PAIN2	−0.122	−0.123	−0.095	−0.095
	(0.005)	(0.005)	(0.005)	(0.005)
PAIN3	−0.385	−0.385	−0.328	−0.329
	(0.007)	(0.007)	(0.006)	(0.006)
MOOD2	−0.071	−0.071	−0.05	−0.051
	(0.005)	(0.005)	(0.005)	(0.005)
MOOD3	−0.237	−0.236	−0.198	−0.196
	(0.007)	(0.007)	(0.006)	(0.006)
N3	−0.269	−0.269	−0.199	−0.199
	(0.009)	(0.009)	(0.007)	(0.007)
VAS			0.329	0.211
			(0.01)	(0.020)
AGE		0.009		0.009
		(0.002)		(0.002)
AGE^2		−0.0001		−0.0001
		(0.00002)		(0.00002)
SEX		−0.049		−0.063
		(0.012)		(0.013)
MARSTAT		0.056		0.053
		(0.013)		(0.013)
XUA		0.078		0.068
		(0.018)		(0.016)
AGEVAS				0.002
				(0.0003)
SEXVAS				0.036
				(0.012)
adj R^2	0.459	0.472	0.479	0.492
JBRES	0.005 [0.997]		0.191 [0.908]	
LB(7)	9.027 [0.251]		10.756 [0.150]	
MAE	0.039		0.034	
No > |0.05|	12		7	
No > |0.10|	3		3	
t(mean = 0)	−0.571 [0.570]		−0.502 [0.619]	

[a] White's heteroscedasticity consistent standard errors. Estimation via GEE (Liang & Zeger, 1987).

All coefficient estimates significant at $t_{0.01}$. Breusch-Pagan tests confirm the importance of individual effects and Hausman's test suggests that RE is appropriate. All models reveal problems on reset and heteroscedasticity tests.

the appropriate one. In all cases, just under one-half of the variation is within respondent (i.e. across health states) with the remainder being across respondents. The residuals appear to reveal heteroscedasticity and the reset test suggests some misspecification error, most likely the result of omitted variables since alternative functional forms did not improve this statistic.

Dividing TTO values into those greater than or equal to zero (i.e. better than dead) and those less than zero (i.e. worse than dead) produces little by way of useful results. For valuations greater than or equal to zero, the basic model structure is very similar and, as would be expected, the effect of all core variables is reduced. However, the explanatory power of the model is less and sex is no longer a significant variable. Whilst the prediction errors are broadly comparable to those for the full set of valuations, they exhibit evidence of both bias and autocorrelation. The misspecification and heteroscedasticity problems are still present. The models perform much worse for TTO valuations less than zero. The explanatory power of the models is very low and the predictive ability is very poor. The N3 variable and all of the level 2 binary dummies except the one on mobility no longer have a significant effect. The relationship with age appears to be linear (rather than quadratic) and having problems on the usual activities dimension is no longer significant. In both the better and worse than dead models, the effect of the VAS score appears to be broadly comparable.

This lack of success with separate models for positive and negative TTO values may not be surprising given that all states appear in each sample (i.e. every state is rated as better than dead or worse than dead by someone). An alternative approach is to distinguish between those respondents who rate all states as better than dead (comprising of 264 respondents) and those who rate at least one state as worse than dead (comprising of the remaining 2773 respondents). Separate TTO models were estimated for these two groups of respondents and, as would be expected, the models for the latter group were very similar to those for the overall sample. The models for those with no states rated as worse than dead generally had smaller coefficients but were similar in other respects. PROBIT models which sought to distinguish between the two groups of respondents contained few significant respondent characteristics and were unable to correctly predict those cases where all TTO values were positive.

Discussion

This paper has sought to explain TTO values for EQ-5D health states in terms of: (i) the different dimensions within the EQ-5D using the same core model as that reported in Dolan (1997); (ii) the background characteristics of the respondents; and (iii) VAS valuations given to the same set of health states. Random effects models were used to account for the fact that groups of valuations came from the same respondent. Various transformations were attempted in order to account for the skewed and truncated nature of the data but none of these improved on the linear random effects model. An important issue for future research, then, is whether there are any transformations of health state valuations which can better represent the complex distribution of those values.

Of the range of background variables that might explain TTO values, age, marital status and sex emerge as three of the most important (possible explanations can be found in Dolan et al., 1996a, b). Valuations increase slowly up to about the age of 45, fall slowly up to about 70 and then fall sharply in later years. The results suggest that, ceteris paribus, a 20 year-old will give about the same value to a health state as a 70 year-old. In terms of the magnitude of the effects, a 45 year-old will value a health state about 0.075 higher than a 70 year-old. The difference in valuations will be about 0.13 if the 45 year-old is married or co-habiting and the 70 year-old is single, separated, divorced or widowed. If the former is also a man whilst the latter is a woman, the difference will be 0.18. If the 70 year-old reports problems on the usual activities dimension of the EQ-5D, her value will be lower by a further 0.08. These differences in valuations would be likely to have a significant effect on the allocation of resources if differential weight is given to the preferences of different population sub-groups.

Of course, these differences relate to health states that have been valued using the TTO method for a 10-year duration. It would be interesting to see whether a similar study using the SG method, and with a comparable sample size to detect differences of a similar magnitude to those reported here, would produce similar results. The important and significant effect of age could partly be explained by the fact that each state was assumed to last for 10 years, followed by immediate death. For example, those respondents who did not believe that they actually had 10 years left to live might be more willing to give up these 'excess' life years (Dolan et al., 1996a, b). However, a qualitative follow-up of some of the respondents from this study suggests that similar results would have been obtained had each respondent's life expectancy been used rather than a fixed duration (Robinson et al., 1997).

It still remains the case, though, that differences in health state valuations cannot be explained by many of the demographic characteristics (such as occupation, education and social class) which have been shown to affect an individual's likelihood of reporting illness or using health care services. Moreover, it is surprising that differences in valuations cannot really be explained in terms of how respondents describe and value their own health. Although differences were found for the usual activities dimension, it seems likely that the significance of this variable is proxying for other important characteristics. Much of the debate in the literature on whose values to use revolves around whether the preferences of those with experience of illness should be given greatest weight (Dolan, 1999). If current health and experience of illness in self, family or friends does not affect valuations, then much of the heat can be taken out of this particular debate.

With regards to the inclusion of the VAS as an additional explanatory variable, Models 3 and 4 in Table 4 would appear to provide potentially useful results. However, squared and cubic VAS terms (and possibly even higher powers which were not tested) were also significant. Such relationships would mean that there are points in the valuation space where a marginal increase in the VAS score is associated with a marginal decrease in the corresponding TTO value. This suggests that the VAS terms are also likely to be proxying for effects not otherwise represented in the model. The failure of the model to pass the reset test is likely to be further evidence of this. In addition, the VAS does not appear to add very much to the explanatory power of the model. The method might, however, have an important role in familiarising respondents with health states and with the idea that they can be expressed using a scale.

By partitioning the data according to whether the TTO value is positive or negative, it was hoped that insights might be provided into the differential effect that certain background characteristics may have on the better and worse than dead scenarios. The models for positive scores are very similar to those for the full dataset. The main substantive difference is that sex ceases to be a significant variable, suggesting that the observed differences between men and women in the full model is the result of the lower values given by women to states rated as worse than dead. The models for negative TTO scores perform badly and it is difficult to say anything meaningful about those results. The fact that no model, not even the core one, performs very well when using only negative scores suggests that other transformations of the scores than the one used here might be more appropriate. Similarly, distinguishing between those respondents who rate all states as better than dead and those who rate at least one state as worse than dead provided few useful insights. Despite this, there is certainly the need for more theoretical and empirical investigation into the issue of how to interpret scores for states rated as worse than dead.

The analysis of the TTO data reported here suggests any attempt to explain health state valuations, no matter how sophisticated, is likely to tell only a small part of the story. It appears from this dataset, as from many others using different valuation techniques, that a considerable part of the variation in respondent valuations will remain unaccounted for. This may simply reflect the heterogeneity is people's preferences regarding health.

References

Abdalla, M., & Russell, I. (1995). Tariffs for the EuroQol health states based on moelling the individual VAS and TTO data of the York survey. EuroQol Plenary Meeting, London, October 1994, Conference Proceedings, HERG Research Report No. 20, Brunel University.

Brazier, J. E., Deverill, M., & Green, C. (1999). A review of the use of health status measures in economic evaluation. *Journal of Health Services Research and Policy*, *4*, 174–184.

Brooks, R. (1996). EuroQol: The current state of play. *Health Policy*, *37*, 53–72.

Camerer, C. (1993). Individual decision-making. In J. Kagel, & A. Roth (Eds.), *Handbook of experimental economics*. Princeton, NJ: Princeton University Press.

Dolan, P. (1996). The effect of experience of illness on health state valuations. *Journal of Clinical Epidemiology*, *49*, 551–564.

Dolan, P. (1997). Modelling valuations for EuroQol health states. *Medical Care*, *11*, 1095–1108.

Dolan, P. (1999). Whose preferences count? *Medical Decision Making*, *19*, 482–486.

Dolan, P. (2000). The effect of age on health state valuations. *Journal of Health Services Research and Policy*, *5*, 17–21.

Dolan, P., Gudex, C., Kind, P., & Williams, A. (1996a). The time trade-off method: Results from a general population study. *Health Economics*, *5*, 141–154.

Dolan, P., Gudex, C., Kind, P., & Williams, A. (1996b). Valuing health states: A comparison of methods. *Journal of Health Economics*, *15*, 209–231.

Dolan, P., & Jones-Lee, M. (1997). The time trade-off: A note on the effect of lifetime reallocation of consumption and discounting. *Journal of Health Economics*, *16*, 731–739.

Dolan, P., & Sutton, M. (1997). Mapping visual analogue scale scores onto time trade-off and standard gamble utilities. *Social Science and Medicine*, *44*, 1519–1530.

Eyman, R. K. (1967). The effect of sophistication on ratio and discriminative scales. *American Journal of Psychology*, *80*, 52–540.

Froberg, D. G., & Kane, R. L. (1989). Methodology for measuring health state preferences III: Population and context effects. *Journal of Clinical Epidemiology*, *42*, 585–592.

Gafni, A. (1994). The standard gamble method: What is being measured and how is it interpreted? *Health Services Research*, *29*(2), 207–224.

Gold, M. E., Russell, L. B., Siegal, J. E., & Weinstein, M. C. (1996). *Cost-effectiveness in health and medicine*. New York: Oxford University Press.

Gudex, C., Dolan, P., Kind, P., & Williams, A. (1997). Valuing health states: Interviews with the general public. *European Journal of Public Health*, *7*, 441–448.

Hadorn, D. C., & Uebersax, J. (1995). Large-scale outcome evaluation: How should quality of life be measured?

Calibration of a brief questionnaire and a search for preference sub-groups. *Journal of Clinical Epidemiology*, *48*, 607–618.

Hendry, D. F. (1987). Econometric methodology: A personal perspective. In T. F. Bewley, *Advances in econometrics: 5th world congress*, vol. 2 (pp. 226–261). Cambridge: Cambridge University Press.

Kind, P., & Dolan, P. (1995). The effect of past and present illness experience on health state valuations. *Medical Care*, *33*(4), AS255–AS263.

Liang, K.-Y., & Zeger, S. L. (1987). Longitudinal data analysis using generalised linear models. *Biometrika*, *73*, 13–22.

Llewellyn-Thomas, H., Sutherland, H. J., Tibshirani, R., Ciampi, A., Till, J. E., & Boyd, N. F. (1984). Describing health states: Methodologic issues in obtaining values for health states. *Medical Care*, *22*, 543–552.

O'Brien, B. J., & Drummond, M. F. (1994). Statistical versus quantitative significance in the socioeconomic evaluation of medicines. *PharmacoEconomics*, *5*, 389–398.

Patrick, D. L., Starks, H. E., Cain, K. C., Uhlmann, R. F., & Pearlman, R. A. (1994). Measuring preferences for health states worse than death. *Medical Decision Making*, *14*, 9–18.

Poulton, E. C. (1989). *Bias in quantifying judgements*. Hove: Lawrence Erlbaum.

Richardson, J. (1994). Cost-utility analysis: What should be measured? *Social Science and Medicine*, *39*(1), 7–21.

Robinson, A., Dolan, P., & Williams, A. (1997). Valuing health states using the VAS and TTO methods: What lies behind the numbers? *Social Science and Medicine*, *8*, 1289–1297.

von Neumann, J., & Morgenstern, O. (1953). *Theory of games and economic behaviour*. New York: Wiley.