MEDICAL CARE
Volume 40, Number 5, pp 442–446

Brief Report

Modelling Valuations for Eq-5d Health States

An Alternative Model Using Differences in Valuations

PAUL DOLAN, DPHIL AND JENNIFER ROBERTS, PHD

OBJECTIVES. The EQ-5D is a preference-based measure of health and is increasingly being used in the evaluation of health technologies. A 'tariff' of values for all 243 EQ-5D health states has been generated using direct valuations on a subset of these states. The tariff is used to express the value of differences between health states, and so this paper explores whether a tariff with better predictive ability can be calculated using differences between values rather than using the values themselves.

METHODS. The original tariff (reported in this journal) was based on valuations for 42 EQ-5D states elicited from a representative sample of 2997 members of the UK general population using the time trade-off method. This same data are used to estimate a tariff based upon the differences in value between the worst possible state (33333) and all other states.

RESULTS. A simple model that fits the data well is one in which the differences in value between 33333 and all other states are explained in terms of the change in each dimension plus a term to pick up whether some dimensions change by the maximum amount whereas others do not change at all. The mean absolute difference between the actual values and those predicted by this model is 0.03 (compared with 0.039 in the original model).

CONCLUSION. The model presented in this paper predicts the difference between 33333 and all other states remarkably well and can be used to generate a tariff for all EQ-5D health states. In fact, this model more accurately predicts the values of states for which there are direct observations, and so we recommend its comparison with the original model in evaluative studies.

Key words: Health state measurement; time trade-off; EQ-5D. (Med Care 2002;40:442–446)

Preference-based measures of health are increasingly being used to evaluate health care interventions. To allow comparisons across different programs that may impact upon different dimensions of health, the chosen descriptive system must allow for the different dimensions to be combined form an overall single index. There now exist a number of descriptive systems that have been specifically designed for this purpose.[1] One such descriptive system is the EQ-5D,[2] which defines health in five dimensions (Table 1). For the EQ-5D to be of use in evaluative studies, it is necessary to attach a single index value to each of the 243 health states generated by its descriptive system. To facilitate this, valuations for subsets of EQ-5D states have been elicited from general

From the School of Health and Related Research, University of Sheffield, Western Bank, Sheffield.

Address correspondence and reprint requests to: Jennifer Roberts, University of Sheffield, Sheffield, United Kingdom S1 4DA. E-mail: J.R.Roberts@shef.ac.uk

Received July 23, 2001; initial review September 12, 2001; accepted December 3, 2001.

TABLE 1. The EQ-5D 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

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, 12223 means:

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

population samples in a number of countries and tariffs for all EQ-5D states have been estimated from these direct valuations.[3–6]

The purpose of a tariff for any descriptive system which generates more states than it is possible to elicit direct valuations for is to accurately predict the values of states for which valuations exist and to estimate the values of states for which valuations do not exist. In one of the largest studies of its kind, values were elicited from a representative sample of the UK general population using the time trade-off (TTO) method,[3] and a tariff of values for all EQ-5D has been generated from these data.[4] The original model used to estimate a tariff for the EQ-5D predicted the values for the 42 states valued in the study well; for example, the mean absolute difference (on a scale between −1–1) between the actual and estimated values was 0.039 (SD = 0.029), and for only three states was the difference in excess of 0.1.

Although we are confident that there is unlikely to be an equally parsimonious model that fits the valuation data and the original model, it is possible that a better predictive model could be generated using the differences between valuations rather than the valuations themselves. Such a model would have intuitive appeal because the values in any tariff are ultimately used to express the difference between health states. In this brief report, we estimate a tariff of values using exactly the same data that was used to generate the original tariff, but instead using differences in values between the worst state (33333) and all other states.

Materials and Methods

Because the valuation study has been detailed elsewhere,[3] including in this journal,[4] only a brief summary is provided here. A total of 3395 members of the UK general population were interviewed in their own homes by trained interviewers. To enable modeling of the data at the individual level only the 2997 respondents with complete data are included in the analysis. Each respondent valued 12 EQ-5D states (drawn from a subset of 42) using the TTO method. For a state regarded as better than dead, the respondent was asked to select a length of time (x) in full health that she considered to be equivalent to 10 years in that state. For a state regarded as worse than dead, the respondent was asked to select a length of time $(10-x)$ in that state followed by x years in full health that she considered to be equivalent to dying immediately. If full health and dead are assigned scores of 1 and 0 respectively, then TTO values for states rated as better than dead are given by the formula $x/10$. Values for states rated as worse than dead are given as $-x/10$, so that all valuations lie in the range $[-1,1]$.

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. Therefore, level 1 on usual activities was not combined with level on mobility or level 3 on self-care. All respondents valued the most severe health state (33333) plus eleven others drawn from four groups defined a priori according to 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. Given this study design, it is possible to consider each respondent's value of each state as the difference from the value they give to 33333. This 'difference' data can then be used to estimate a tariff for all EQ-5D health states.

The regression equation is as follows:

$$y_{ij} = c + \beta_1'\overline{x1}_{ij} + \beta_2'\overline{x2}_{ij} + \gamma ANY13_{ij} + \epsilon_{ij}$$

where:

i = 1, 2, . . ., n represents individual health state values.

j = 1,2, . . ., m represents respondents.

c is the intercept.

y_{ij} is the difference between the TTO valuation of the comparator state (i) and state 33333, valued by respondent j.

$\overline{x1}_{ij}$ and $\overline{x2}_{ij}$ are vectors of dummy explanatory variables that describe the difference between the comparator state and state 33333. There are five dummy variables in each vector, one for each dimension of the EQ-5D. $\overline{x1}$ takes the value of one if the difference between a particular dimension is one, ie, the comparator state is at level 2 on that dimension. $\overline{x2}$ takes the value of one if the difference between a particular dimension is two, ie, the comparator state is at level 1 on that dimension. There are 10 of these terms in total, with a difference of zero (ie, level 3 in the comparator state) acting as a baseline for each dimension. $ANY13_{ij}$ is a dummy variable that takes a value of one if there is a change of two levels on at least one dimension and no change in at least one other dimension.

The estimation methods used here are similar to those used to generate the original model.[4] The model is a generalized least squares regression using individual level data, with eleven (j = 11) observations for each of i = 2997 respondents (n = 32,967). A random effects model is used to account for the hierarchical nature of the data.[7] The error term is split into two components, $\epsilon_{ij} = e_{ij} + u_j$; e_{ij} is a traditional error term unique to each observation, whereas u_j is a respondent level random effect. This specification was confirmed using Breusch-Pagan[8] and Hausman[9] tests. The rejection of the null hypothesis for both of these tests (shown in Table 1) shows that individual effects are necessary and that random effects are the correct specification.

The final model[1] was derived from a set of alternative models, which always included the $x1$ and $x2$ vectors, but differed in the way they dealt with the additional term. Alternatives to $ANY13$ were terms that reflected the presence of level 3 in any dimension, the presence of level 1 in any dimension, a difference of 2 on any dimension and a difference of zero on any dimension. In addition, a simple model with no additional term was also estimated. These models were analogous to those evaluated to generate the original tariff.

The best model was chosen on the basis of a battery of diagnostic tests. These included the significance and expected sign and size of the coefficient estimates; explanatory power; the adherence of the residuals to the assumptions of least squares estimation (independence and constancy of variance) and a general Ramsey reset test for misspecification.[10] In addition the models were compared in terms of their ability to predict observed mean health state values. This included the size of the prediction errors and the presence of any systematic variation (Ljung-Box test[11]) and the normality of the error distribution (Jarque-Bera test[12]). To investigate the robustness of the parameter estimates, the model was also estimated on an internal sub-sample of a randomly chosen two-thirds of respondents, and also by dropping the difference between 33333 and each state in turn from the model.

Results

The results for the best and most parsimonious model (see Equation 1) are shown in Table 1. All coefficients are significant and, as expected, the first ten explanatory variables all have a positive coefficient and the coefficients on variables that involve a move of two dimensions are always larger than on those that involve a move of one dimension. The coefficient on $ANY13$ suggests that when the comparator state includes a combination of both the best and the worst levels, the difference between the value for this state and state 33333 is reduced.

The explanatory power of the model compares favorably with that of the original, at 0.55 compared with 0.46 (although there is no precise counterpart to the coefficient of determination when using GLS). As with the original model, the Ramsey Reset test[10]suggests problems with misspecification. This is not surprising given the parsimonious model and the power of the test given the large number of observations. There are also problems with heteroscedasticity (nonconstant error variance) and so the coefficients are estimated

TABLE 2. Parameter Estimates (for Entire Sample)

Variable	Description	Estimate
C		−0.201 (0.012)
DIFMOB1	1 if the difference in the *mobility* level is 1; 0 otherwise.	0.320 (0.006)
DIFMOB2	1 if the difference in the *mobility* level is 2; 0 otherwise.	0.391 (0.007)
DIFSC1	1 if the difference in the *self-care* level is 1; 0 otherwise.	0.179 (0.006)
DIFSC2	1 if the difference in the *self-care* level is 2; 0 otherwise.	0.280 (0.007)
DIFUA1	1 if the difference in the *usual activities* level is 1; 0 otherwise.	0.084 (0.007)
DIFUA2	1 if the difference in the *usual activities* level is 2; 0 otherwise.	0.156 (0.007)
DIFPAIN1	1 if the difference in the *pain/discomfort* level is 1; 0 otherwise.	0.372 (0.006)
DIFPAIN2	1 if the difference in the *pain/discomfort* level is 2; 0 otherwise.	0.491 (0.006)
DIFMOOD1	1 if the difference in the *anxiety/depression* level is 1; 0 otherwise.	0.271 (0.006)
DIFMOOD2	1 if the difference in the *anxiety/depression* level is 2; 0 otherwise.	0.356 (0.006)
ANY13	1 if the differences include 0 and 2; 0 otherwise.	−0.125 (0.006)
R^2 (within)		0.554
Ramsey RESET(1)[10]		166.17 [.000]
Breusch-Pagan test for individual effects[8]		524.24 [.000]
Hausman test for random effects[9]		771.48 [.000]

All modelling carried out in Stata 7.0.
t statistics in parentheses.
Dependent variable: the difference between the TTO valuation of the comparator state and state 33333.

with White's heteroscedasticity-consistent standard errors.[13] Despite these problems, the parameter estimates are unbiased. Empirical analysis also reveals that the parameter estimates are robust to the exclusion of any one of variables, suggesting that collinearity is not a serious problem.

Using the observed mean value of −0.543 for state 33333, it is possible to use this model to predict the mean values for the other 41 states valued. For example, state 12223, when written as a difference from 33333, is 21110. Using the estimates reported in Table 2, the relevant dummies are therefore DIFMOB2, DIFSC1, DIFUA1, DIFPAIN1 and ANY13. The predicted value from the model is 0.201 + 0.391 + 0.179 + 0.084 + 0.372 to 0.125 = 0.700. Adding this to the observed mean value for state 33333 (−0.543) gives a predicted value for state 12223 of 0.157.

When predictions for all 41 states are generated in this way, the Jarque-Bera[12] test shows that the errors are normally distributed with a mean of zero, and the Ljung-Box[11] statistic shows that there is no autocorrelation in the errors (when ordered by actual mean valuation of each health state).

Figure 1 compares the actual mean values with the predicted values from the new model and with those from the original model. It can be seen that the new model compares favorably with the original model. For only 12 states is the prediction error from the new model larger than that from the original. The mean absolute error in the new model is 0.03 compared with 0.039 in the original. The SD around this number is also smaller −0.026 as compared with 0.029 in the original model.

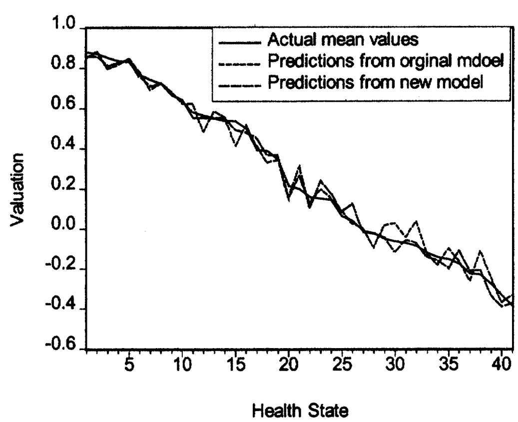

FIG. 1. Actual mean values and predictions from the original and new models.

There are eight states for which the error exceeds 0.05 and only one state (11131) for which the difference between the predicted and actual mean value exceeds 0.10. This compares with 12 and 3 states, respectively, in the original model. The relative superiority of the model based on differences in values is also borne out when the robustness of its parameters estimates are tested. Estimation using the values of two-thirds of respondents gives coefficient estimates that are similar to those reported for the full sample in Table 2. Using this model to predict the values of the remaining one-third of respondents gives a mean error of 0.032 as compared with 0.046 in the original model. Also, when each state is excluded in turn from the estimation, and the model is used to predict the value of the excluded state, the mean errors in the new and original models are 0.041 and 0.048, respectively.

Discussion

In this brief report, we have presented an alternative model for estimating a tariff for the EQ-5D, which is based on differences in valuations between the worst state and all other states. The parameters in the new model are similar to the original, explaining the differences in values (rather than the values themselves) in terms of the movements across each level of each dimension. As with the original algorithm the pain dimension has the largest effect with differences on the usual activities dimension being perceived as less important. The models differ only in their additional interaction term. The N3 term (to pick up whether any dimension is at level 3) used in the original model does not increase the explanatory power of the new model and the ANY13 term used here does not increase the explanatory power of the original model.

The two models produce some differences that might be considered important in evaluative studies. For example, depending on the pair of states chosen, the difference in the implied change in health given by the two tariffs could be as large as 0.25. Given such differences, we recommend that the new model be used alongside the original one to see what practical differences it might make. Where differences are found, we would recommend that the new tariff be used because it has better predictive ability and is based on differences in valuations rather than the valuations themselves, which might be an attractive property in itself.

References

1. **Brazier JE, Deverill M, Green C.** A review of the use of health status measures in economic evaluation, Journal of Health Serv Res and Policy 1999;4:174–184.

2. **Brooks R. EuroQol:** The current state of play. Health Policy 1996;37:53–72.

3. **Dolan P, Gudex C, Kind P, et al.** The time trade-off method: results from a general population study. Health Economics 1996;5:141–154.

4. **Dolan P.** Modelling valuations for EuroQol health states. Med Care 1997;11:1095–1108.

5. **Badia X, Roset M, Herdman M, et al.** A comparison of UK and Spanish general population time trade-off values for EQ-5D health states. Med Decis Making 2001;21:7–16.

6. **Tsuchiya A, Ikeda S, Ikegami N, et al.** Estimating an EQ-5D population value set: The case of Japan, Health Economics, forthcoming 2001.

7. **Greene WH.** Econometric Analysis. New York, NY: Macmillan; 2000:567–570.

8. **Breusch T, Pagan A.** The Lagrange multiplier test and its applications to model specification in econometrics. Rev Econ Stat 1980;47:239–253.

9. **Hausman JA.** Specification tests in econometrics. Econometrica 1978;46:1251–1271.

10. **Ramsey JB.** Tests for specification errors in classical linear least squares regression analysis. Journal of the Royal Statistical Society, Series B 1969;31:350–371.

11. **Ljung G.** Box On a Measure of Lack of Fit in Time Series Models. Biometrika 1979;66:265–270.

12. **Jarque CM, Bera AK.** A test for normality of observations and regression residuals, International Statistical Review, 1987;55:163–172.

13. **White H.** A Heteroskedasticity-Consistent Covariance. Matrix and a Direct Test for Heteroskedasticity. Econometrica 1980;48:817–838.