Published paper
McCabe, C., Brazier, J., Gilks, P., Tsuchiya, A., Roberts, J., O'Hagan, A. and Stevens, K. (2006) *Using rank data to estimate health state utility models.* Journal of Health Economics, 25 (3). pp. 418-431.

Using rank data to estimate health state utility models.

Christopher McCabe[1†‡], John Brazier[†‡], Peter Gilks[‡], Aki Tsuchiya[†], Jennifer Roberts[†], Anthony O'Hagan[‡ϕ] Katherine Stevens[†]

Health Economics and Decision Science[†]
Centre for Bayesian Statistics in Health Economics[‡]
Department of Probability and Statistics[ϕ]
University of Sheffield

JEL classification: I10

Keywords: Preference-based health measures; Ordinal preferences; SF-6D; HUI2; Modelling stated preference data

[1] Address for correspondence: Christopher McCabe, Senior Lecturer in Health Economics
Sheffield Health Economics Group, ScHARR, University of Sheffield, Regent Court
31 Regent Street, Sheffield, S1 4DA. Email: c.mccabe@sheffield.ac.uk
Tel: 00 (44) 114 222 5454; Fax: 00 (44) 114 272 4095

Abstract

In this paper we report the estimation of conditional logistic regression models for the Health Utilities Index Mark 2 and the SF-6D, using ordinal preference data. The results are compared to the conventional regression models estimated from standard gamble data, and to the observed mean standard gamble health state valuations.

For both the HUI2 and the SF-6D, the models estimated using ordinal data are broadly comparable to the models estimated on standard gamble data and the predictive performance of these models is close to that of the standard gamble models. Our research indicates that ordinal data have the potential to provide useful insights into community health state preferences. However important questions remain.

Acknowledgement

The authors wish to acknowledge the support and encouragement of their co-workers on the SF-6D and HUI2 valuation projects. In addition, George Torrance from McMaster University and Joshua Salomon from Harvard University provided valuable comments on the work as it progressed. We are grateful to the UK Medical Research Council, who funded the HUI2 valuation project and to GlaxoWellcome, who funded the SF-6D valuation project. All errors and omissions remain the authors' responsibility.

Introduction

As cost effectiveness analysis has become more important in health care decision making processes, the interest in how to value health outcomes has increased. There is a substantial body of research on the relative strengths and weaknesses of alternative methods (e.g. Torrance 1986; Brazier et al 1999). Such research has focused primarily on three valuation methods; Time Trade Off (TTO); Standard Gamble (SG); and Visual Analogue Scales (VAS), also called category scaling.

Work that has attempted to identify a preferred method has tended to support the use of TTO and/or SG (Brazier et al. 1999; NICE 2004). VAS has been criticised on a number of points, both theoretical (does VAS capture strength of preference?) and empirical (the data may be subject to end-point and context bias). (Torrance et al 2001) However, it is widely accepted that TTO and SG have significant limitations. (Brazier et al 1999) What is remarkable is the degree to which the role of ordinal data in health state valuation has been largely ignored; notable exceptions to this observation being the work by Kind (1982, 1996).

Ranking exercises are conventionally included in health state valuation interviews as a warm-up exercise, in order to familiarise the interviewee with the health state classification system being valued and with the task of considering preferences between hypothetical health states, (Furlong et al

1990). The use of the data from these ranking exercises has generally been limited to checking the degree of consistency between the valuations obtained from the SG or TTO valuation exercises and the ranking exercise.

Kind (1982) identified Thurstone's (1927) model of comparative judgement as a potential theoretical basis for deriving cardinal preferences from rank preference data. Thurstone's method considers the proportion of times that health state *A* is considered worse than health state *B*. The preferences over the health states represent a latent cardinal utility function. Individual's stated preferences draw upon this latent function but imperfectly; i.e. there are errors in individual's expression of the latent utility function. The closer two health states, A and B, lie on the latent utility function the greater the likelihood that an individual will incorrectly state that they prefer B to A, when in fact the utility they expect to gain from health state A is greater than the utility they expect to gain from health state B. Thus there is a relationship between observed ordinal preferences and the underlying cardinal latent utility function. McFadden (1974) proposed the conditional logistic regression model as a means of modelling this latent utility function from ordinal data. The assumptions underlying McFadden's choice model are clearly described by Saloman (2003).

Recently Salomon (2003) presented work that applied conditional logistic regression models to the rank data collected as part of the Measurement and Valuation of Health Study (MVH). Salomon estimated a model equivalent to that reported by Dolan (1997). This model did not produce utilities on the 0-1 scale necessary for use in estimating Quality Adjusted Life Years. Salomon rescaled the model coefficients on to the full health-death (1-0) scale, using the mean measured TTO value for the PITS state in the EQ-5D classification (3,3,3,3,3). In this paper we present an approach that avoids the need for external health state utility data, as in such rescaling, by directly estimating a parameter for the state death, as part of the model. This method is applied to rank data from two health state valuation surveys; a UK based valuation survey for the Health Utilities Index Mark 2, (McCabe et al 2004a) and the UK valuation survey for the SF-6D, (Brazier et al, 2002).

Methods

Data

Detailed descriptions of the HUI2 and SF-6D classification systems, and the valuation surveys have been reported in detail elsewhere, thus, we will only provide a brief summary of them here, (Brazier et al, 2002; McCabe et al 2004a). (See Appendices 1 and 2)

Health Utilities Index Mark 2

The Health Utilities Index Mark 2 is a six dimension health state classification (sensation, mobility, emotion, cognition, self care and pain) with either four or five levels for each dimension. It describes a total of 8,000 distinct health states. It was developed specifically for use with paediatric populations, (Torrance et al 1996). (See Appendix 1)

In the present study, one hundred and ninety eight respondents ranked 8 health states from the HUI2 classification plus Full Health and Immediate Death, (McCabe et al 2004). The health states valued were sampled from an orthogonal array for the HUI2 classification. The interviewees then valued the same 8 health states using the McMaster version of the SG question; i.e. the chance board prop was used to aid the respondent in understanding the question.

The chance board was designed to communicate probabilities to respondents who have little or no experience of the concepts. The probability of the two uncertain outcomes and the one certain outcome associated with the two choices are displayed at the same time. Furlong et al state that “The board uses diagrams of common gambling-type wheels with colour coded pie-shaped segments representing the probabilities.” (Furlong et al. 1990)

The risk of death was varied in a ping-pong manner until the respondent identified a risk of death at which they were indifferent between the impaired health state and the uncertain choice. Where health states were ranked as worse than immediate death, the worse than death version of the SG question was used, (Furlong et al 1990). In the worse than death version of the SG question; the options are death with certainty or the uncertain choice of full health or the impaired health state being valued.

The respondent was asked to imagine that they were a ten year old child who would live for another 60 years in the outcome health state.

SF-6D

The SF-6D has 6 dimensions: physical functioning, role Limitations, social functioning, pain, mental health and vitality. Each dimension has 4, 5 or 6 levels. The classification describes a total of 18,000 health states, (Brazier et al 2002). (See Appendix 2)

A representative sample of 611 members of the UK population provided standard gamble valuations for a sample of 249 health states defined by the SF-6D classification.

The interview consisted of an exercise to rank 5 health states that the respondent would then be asked to value, plus the best and worst states defined by the SF-6D and immediate death. This was followed by a series of SG questions. The SG question asked the respondent to value one of 5 certain SF-6D health states, in a lottery with the best and 'pits' health states as the alternative outcomes. All respondents were then asked to provide a SG valuation of the PITS state in relation to death. The form of the sixth SG valuation depended upon whether the respondent has ranked the PITS state as better or worse than death, in the ranking exercise. The result of the sixth SG exercise was then used to 'chain' the health state values in order to place them on to the 1-0, full health –death scale. The interviewers used the McMaster chance board prop and the ping-pong version of the SG question, (Furlong et al. 1990).

The respondent was asked to answer the question for him or herself, imagining that they would remain in the outcome health state for the rest of their lives, (Sturgis and Thomas, 1998).

Model specification

To model the predicted health state valuations using the ordinal preference data we used conditional logistic regression as outlined by McFadden (1974). To operationalise this model we assumed that the ranking exercise is equivalent to the respondent making a series of individual selections from smaller and smaller sets of states. Thus, in ranking 10 health states we assume that the respondent first chooses the most preferred health state from all 10, before choosing the most preferred health state from the remaining 9 and so on, until all the health states have been assigned a rank between 1 and 10. To characterise this as equivalent to pair wise choice we must rely on the hypothesis of the Independence of Irrelevant Alternatives; i.e. the ranking of the pair is not affected by the other states that are ranked in the same exercise.

The conditional logistic regression model assumes that respondent i has a latent utility value for state j, U_{ij}, and that given the choice of two states j and k, the respondent will choose state j over state k if $U_{ij} > U_{ik}$. Hence given the task of choosing the preferred state from a finite group of different states, respondent i will choose state j if $U_{ij} > U_{ik}$ for all $j \neq k$.

Each individual's cardinal utility function for state j is $U_{ij} = \mu_j + \varepsilon_{ij}$ where μ_j is representative of the tastes of the population and ε_{ij} represents the particular taste of the individual. If the error term ε has an extreme value distribution, then the odds of choosing state j over state k are $exp\{\mu_j - \mu_k\}$.

For the analyses reported here, the expected value of each unobserved utility was assumed to be a linear function of the categorical levels on the dimensions of each dataset respectively. The general model specification is:

$$\mu_{ij} = g(\boldsymbol{\beta}'\mathbf{x}_{\mathbf{ij}} + \theta D + u_{ij}) \qquad (1)$$

where μ = utility; i = 1, 2, …, n represents respondents and j = 1,2, …, m represents health states. g is a function specifying the appropriate functional form, which is assumed here to be linear. u_{ij} is an error term whose autocorrelation structure and distributional properties depend on the assumptions underlying the particular model used.

x is a vector of dummy explanatory variables ($x_{\lambda\delta}$) for each level λ of dimension δ of the instrument in question. For example for the SF-6D, x_{23} denotes dimension δ = 3 (social functioning), level λ = 2 (health limits social activities <u>a little of the time</u>). For any given health state, $x_{\lambda\delta}$ will be defined as

$x_{\lambda\delta}$ = 1 if, for this state, dimension δ is at level λ

$x_{\lambda\delta}$ = 0 if, for this state, dimension δ is not at level λ

Level 1 is the baseline for each dimension.

D is a dummy variable for the state ‘Death’, which takes the value 1 for this health state. For all other health states the variable Death is always set at 0.

The value of the full health state is constrained to equal 1. The value of any other health state is calculated as 1 minus the sum of coefficients for each of the dimension level dummies in the state.

Rescaling model coefficients on to the death-full health (1-0) scale

The latent variable μ is not estimated on the zero-one (death-full health) scale required for calculating QALYs. Therefore, we rescaled the coefficients using the formula $\beta_{r\lambda\delta} = \beta_{\lambda\delta} / \theta_D$; where $\beta_{r\lambda\delta}$ is the rescaled coefficient on dimension level $\lambda\delta$ and θ is the coefficient on death. These rescaled coefficients provide predictions for health state values on the same scale as SG or TTO valuations, although not necessarily the same values. This method of rescaling anchors death at zero, and full health at 1, whilst retaining the possibility of a health state having a value of <0; i.e. worse than death.

Model Assessment

Models are assessed in a number of stages. The first stage checks that the estimated model coefficients have the expected negative sign and that they are statistically significant. The second stage checks for logical inconsistencies; i.e, that lower levels of functioning are associated with greater decrements in health state value.

The rescaled coefficients are then compared to the coefficients from the preferred models estimated on the SG data from the same valuation interviews, (Brazier et al 2002, McCabe et al 2004). We assessed the predictive performance of the models using the following battery of measures:

- Root Mean Square Error (RMSE),
- Mean Absolute Error (MAE),
- Intra-class correlation co-efficient (ICC)
- Proportion of health state values predicted to within 0.05 of the observed mean of the standard gamble valuations
- Ljung-Box test for autocorrelation in the errors, (Ljung G, Box G 1979).

The RMSE and MAE are both summary measures of the prediction error compared to the observed mean SG value. We report both for comparability with other health state valuation model literature, some of which report the MAE, whilst others report the RMSE. (Salomon, 2003; Brazier et al, 2002; Dolan 1997; McCabe et al, 2004a)

In addition we plot the health state values predicted by the models against the observed mean SG values and the values predicted by the original SG models. We also plot the errors against the observed mean values.

We use the Hausman test to test the validity of the Independence of Irrelevant Alternatives assumption (IIA), (Weesie J, 2004). Hausman's test compares the maximum-likelihood estimator of beta based on the full dataset with maximum likelihood estimators of beta based on data in which one alternative is dropped. Under IIA, beta(restricted) and beta(overall) should be approximately equal. IIA is violated if the two estimates of beta are significantly different.

We report model coefficients, significance levels, diagnostic plots and tests of predictive performance for both the HUI2 and the SF6D models.

Results

Health Utilities Index Mark 2

Table 1 reports the original and rescaled coefficients for the rank health state utility models for the HUI2. It also gives the results for each of the diagnostic tests. For comparative purposes the same information is provided for the SG health state valuation model (McCabe et al 2004).

The similarity of the rank and SG data models is quite striking. The rank model has one more inconsistency than the SG model, and does not distinguish as clearly between the different levels on the mobility dimension. However, this dimension is one of the weaker dimensions in the SG model. With the exception of the sensation and mobility dimensions, the utility decrement for the impaired level of functioning on each dimension are larger in the SG than the rank model. The predictive performance of the two models is closer than we would have expected given the difference in the level of information the two models were estimated from. This said, the SG model does perform better than the rank model on all tests.

Figure 1 plots the observed health state values and the prediction errors for both the SG and the rank health state models. The plots confirm the similarity of the predictive performance of the rank and SG models.

SF-6D

Table 2 reports the same information for the SF-6D models.

The rank data model is quite different from the SG model. It is notable that the number of inconsistencies is lower in the rank data model than the SG model. Whilst there are inconsistencies in the coefficients for role physical, in both models, there are fewer in the rank model than the SG model. The vitality dimension in the SG model has a number of inconsistencies, the rank model by contrast has none. The predictive performance of the rank model is slightly worse than the SG model, for most tests. However, this may not be surprising as the SG model is being used to estimate the data on which it was estimated, whilst the rank model is being used to estimate a different dataset, although the data was obtained from the same sample of respondents. The LB test results suggest that the relationship between prediction error and observed health state utility is less strong for the rank model than the SG model.

Figure 2 plots the observed mean values and the prediction errors for both the SG and rank data models. It is clear that there is greater variability in the errors for the SF-6D compared with the HUI2.

Independence of Irrelevant Alternatives

Table 3 reports the test of the assumption of Independence of Irrelevant Alternatives for both the HUI2 and SF-6D rank models. The results are not consistent across all the rank groups, but for both the HUI2 and the SF-6D models, there is evidence that this assumption does not hold. The models appear to be most sensitive to the exclusion of those states ranked highly or lowly.

Discussion

In this paper we have reported the estimation of population cardinal health state valuation models for the HUI2 and the SF-6D, from individual ordinal preference data. In both cases the models bare comparison to the health state valuation models estimated from SG (cardinal) data provided by the same respondents.

The impetus for this research was an analysis of rank data for the EQ-5D, presented by Salomon (2003). The predictive performance of the rank EQ-5D model, in relation to the observed mean TTO value, (MAE=0.062), is slightly superior to the SF-6D rank model (MAE 0.11) but identical to the HUI2 rank model.

Our apparent success in estimating cardinal health state valuation models from ordinal data raises many questions. In describing our results as a success, we are assuming that the SG data are the appropriate ‘gold standard’ by which to judge these models. It is arguable that our results say as much about the limitations of SG data as they do about the existence or otherwise of a latent utility function. Research is required to examine whether respondents expressed preferences are consistent with the models that are derived from the SG (and TTO) values they provide. Such work is likely to require qualitative as well as quantitative methods.

Our analysis of the performance of the rank models has assumed that the relationship between the observed SG values and the predictions of the rank models is linear. There is no reason why this should be so. The ranking exercise does not involve risk, whilst the SG explicitly incorporates risk into the valuation process. Standard models of risk attitude would suggest that a linear model would not be the best functional form (Dyer and Sarin 1982). Future work should look at the performance of alternative functional forms. Theoretical perspectives on the relationship between rank and SG data should inform such research.

Similarly, the use of a linear additive function for the HUI2 model is at odds with the research of its developers and others, who report that a multiplicative multi attribute utility function fits the MAUF data best. (Torrance et al 1996; Feeny et al. 2002; McCabe et al, 2004b). In contrast, McCabe (2003) reports that a linear additive model for the HUI2 had better predictive performance, in a validation dataset, than the multiplicative multi-attribute utility function. Currim and Sarin (1984) also found that a linear additive model had better predictive performance than a multiplicative multi-attribute utility function, although this was not in a health state preference context. More research is needed on the choice of functional form for health state preference models.

The application of the conditional logistic regression model requires that the rank data exercise be characterised as a sequential choice process. Whilst we believe that this assumption is defensible, we accept that other models of the ranking process are equally plausible. The results of the Hausman test results suggest that this assumption may not be robust and therefore our results must be treated with some caution. There is an increasing body of research suggesting that respondents apply decision heuristics to complex choice scenarios, and that lexicographic preferences are common in contingent valuation studies. (Lloyd, 2003; Cairns et al, 2002) Research on the thought processes of individual's undertaking ranking exercises would be a valuable contribution to this field.

A potential solution to this problem would be to design the ranking exercise to ensure consistency with the underlying assumptions of the model. Thus the respondent would be presented with all the health states to be ranked and asked to identify the highest ranked health state. This would be recorded and then the respondent would be presented with the remaining health states and again asked to identify the highest ranked health state from that set. This process would be repeated until all the states had been ranked. Work to establish the feasibility of undertaking this type of valuation exercise and to compare the results with those from the ranking exercises presented here would be of significant value.

Our analyses assume that the rank data are preference data. The literature on health state preference elicitation has generally argued that VAS data are not preferences because the valuation process does not require the respondent to trade. This same observation can be applied to ranking exercises. If rank data are reflecting an underlying utility function the utility functions may reflect Broome's (1994) concept of the relative 'goodness' of different health states, rather than the conventional expected utility, that the SG is designed to measure.

The analyses assume that the information content of the rank is unaffected by the order of the rank or indeed the number of states to be ranked. Hausman and Ruud (1987) have hypothesised that respondents may take more care with the initial ranking exercises than the later ones. Thus the risk of a ranking being incorrect would be systematically related to a health state's position in the rank; i.e. the assumption of independence of irrelevant alternatives would not hold. Koop and Poirier (1994) report that a limited relaxation of this assumption in a model of voter preferences did not have a significant impact upon the results. Our results suggest that the assumption does not hold for either model, and that the models are sensitive to both the highly ranked and lower ranked health states, but relatively insensitive to those states ranked in the middle.

Should future research confirm the promise of ordinal data to support the modelling of cardinal health state preferences, it is by no means clear what the implications for future health state valuation work would be. It may be that ranking data may make it possible to incorporate the views of populations for whom the TTO and SG procedures are felt to be too arduous e.g. younger children (Saigal et al 1996). However, the ranking tasks themselves are not simple and no research to date has examined children's ability to understand them.

An alternative benefit may be that the future valuation surveys may require fewer resources. In addition, ranking exercises may be more feasible in postal interviews than TTO and SG, again allowing more efficient implementation of health state valuation surveys. It might be that rank data offers the convenience of the VAS without the problems of context and end-point bias, (Torrance et al. 2001).

These results raise questions about the relationship between discrete choice experiments and the conventional methods of obtaining health state preferences for calculating QALYs. The format of the discrete choice question fits more immediately within the comparative judgement framework than the ranking exercises described above. It seems reasonable to expect that discrete choice scenarios that included a dimension for mortality (or risk of mortality) might be suitable data sources for a similar modelling strategy to that described in this paper. This said, it may also prove to be the case that ranking is a more efficient means to collect these data than discrete choice experiments. Comparative research is required to examine the pros and cons of these two alternative approaches.

Summary

In this paper we have presented two models of population cardinal health state preferences based upon individual ordinal health state preference data; one for the SF-6D health state classification, the other for the HUI2 health state classification. We have compared these to models estimated on SG valuation data, in terms of the degree of accuracy and bias in predicting mean observed SG health state valuations in the estimation samples.

The ordinal rank models perform much better than might have been expected given the difference in the informational content between the SG and ranking exercises.

The results are consistent with Thurstone's law of comparative judgement (1927), and the existence of a latent utility function. The results also suggest that there is potential for discrete choice experiments to provide health state preference data on the full health-death scale. Further research on the potential for ordinal health state valuation data to reflect cardinal population preferences is required.

References

Brazier, J.E., Deverill, M., Green, C., Harper, R., Booth, A., 1999. A review of the use of health status measures in economic evaluation. Health Technology Assessment 3 (9).

Brazier, J.E., Roberts, J., Deverill, M., 2002. The estimation of a preference based measure of health from the SF-36. Journal of Health Economics 21 (2), 271- -292.

Broome, J., 1994. QALYs. Journal of Public Economics 50(2), 149-167.

Cairns, J., van der Pol, M., Lloyd, A. 2002. Decision making heuristics and the elicitation of preferences. Health Economics 11(7):655-658

Currim, IS., Sarin, RK., 1984. A comparative evaluation of multi-attribute consumer preference models. Management Science 30(5):543-61

Dolan, P., 1997. Modelling valuations for EuroQol health states. Medical Care 35 (11), 1095- -1108.

Dolan, P., Sutton, M., 1997. Mapping visual analogue scale scores on to time trade off and standard gamble utilities. Social Science and Medicine 44 (10), 1519- -1530.

Dyer, J.S., Sarin, R.K., 1982. Relative risk aversion. Management Science 28(8), 857- -886.

Furlong, W., Feeny, D., Torrance, G.W., Barr, R., Horsman, J., 1990. Guide to design and development of health state utility instrumentation. CHEPA Working Paper, 90-9. McMaster University Hamilton, Ontario.

Hausman, J.A., Ruud, P.A., 1987. Specifying and testing econometric models for rank-ordered data. Journal of Econometrics 34, 83- -104.

Kind, P., 1982. A comparison of two models for scaling health indicators International Journal of Epidemiology 11(3):271-5

Kind, P., 1996. Deriving cardinal scales from ordinal preference data: the analysis of time trade-off data using pairwise judgement models. Paper presented to HESG Brunel University.

Koop, G., Poirier, D.J., 1994. Rank-ordered logit models: an empirical analysis of Ontario voter preferences. Journal of Applied Econometrics 9, 369- -388.

Ljung, G., Box, G., 1979. On a measure of lack of fit in time series models. Biometrika 66, 265- -270.

Lloyd, A. 2003. Threats to the estimation of benefit: are preference elicitation methods accurate? Health Economics;12(5):393-402

McCabe, CJ., 2003. Estimating preference weights for a paediatric health state classification (HUI2) and a comparison of methods. University of Sheffield, Sheffield.

McCabe, C., Stevens, K., Roberts, J., Brazier, J.E., 2004. Health state values for the HUI2 descriptive system: results from a UK survey. Health Economics. (*forthcoming*)

McCabe, C., Stevens, K., Brazier, JE., 2004 Utility Values for the Health Utility Index Mark 2: An Empirical Assessment of Alternative Mapping Functions. Medical Care. (*forthcoming)*

McFadden, D., 1974. Conditional Logit analysis of qualitative choice behaviour, in Zarembka, P., (ed), Frontiers in Econometrics. New York Academic Press, USA, pp. 105- -142.

National Institute for Clinical Excellence, 2004. Guide to the methods Technology Appraisal. NICE England

[11] Saigal, S., Feeny, D., Rosenbaum, P., Furlong, W., Burrows, E., Stoskopf, B., 1996. Self-perceived health status and health related quality of life of extremely low birth weight infants at adolescence. JAMA 276 (6), 453- -459.

Salomon, J., A., 2003 Reconsidering the use of rankings in the valuation of health states: a model for estimating cardinal values from ordinal data. Population Health Metrics 1(12)

Sturgis, P., Thomas, R., 1998. Deriving a preference based utility score for the SF-6D: technical report. Survey Methods Centre at SCPR. England

Thurstone, LL 1927. A law of comparative judgement. Psychological Review, 34, 273-286

Torrance, G.W., 1986. Measurement of health state utilities for economic appraisal: a review. Journal of Health Economics 5, 1- -30.

Torrance, G.W., Feeny, D.H., Furlong, W.J., Barr, R.D., Zhang, Y., Wang, Q., 1996. A multi-attribute utility function for a comprehensive health status classification system: Health Utilities Mark 2. Medical Care 34 (7), 702- -722.

Torrance, G.W., Feeny, D., Furlong, W., 2001. Visual Analogue Scales: Do they have a role in the measurement of preferences for health states?. Medical Decision Making 21, 329- -334.

Weesie, J. Hausman's test for the assumption "Independence of Irrelevant Alternatives" http://www.fss.uu.nlsoc/iscore/stata/iia.hlp (accessed 13th December 2004

Table 1: Ordinal and Standard Gamble Health State Valuation Models for HUI2[2]

	RankCoeff	RescaledCoeff	SGCoeff
Sensation Level 2	-0.9933	-0.1156	-0.1151
Sensation Level 3	-0.9351	-0.1089	-0.1223
Sensation Level 4	-2.1167	-0.2464	-0.2253
Mobility Level 2	-0.7287	-0.0848	-0.0516
Mobility Level 3	-0.9887	-0.1151	-0.1224
Mobility Level 4	-0.8041	-0.0936	-0.1308
Mobiliyt Level 5	-1.0085	-0.1174	-0.1103
Emotion Level 2	-0.8122	-0.0946	-0.0945
Emotion Level 3	-1.0001	-0.1164	-0.1119
Emotion Level 4	-1.4291	-0.1664	-0.1801
Emotion Level 5	-1.4378	-0.1674	-0.1824
Cognition Level 2	-0.3223	-0.0375	-0.0567
Cognition Level 3	-0.5438	-0.0633	-0.0966
Cognition Level 4	-0.7732	-0.0900	-0.1676
Self Care Level 2	-0.4409	-0.0513	-0.0516
Self Care Level 3	-0.6924	-0.0806	-0.1138
Self Care Level 4	-0.7762	-0.0904	-0.1158
Pain Level 2	-0.8132	-0.0947	-0.1114
Pain Level 3	-0.9401	-0.1095	-0.1155
Pain Level 4	-1.2169	-0.1417	-0.1626
Pain Level 5	-1.7654	-0.2055	-0.2538
Death	-8.5895	-1	
n states		51	51
MAE		0.062	0.051
No.>0.05		23	18
No.>0.10		12	5
RMSE		0.0775	0.0657
LB		36.11	25.78
ICC		0.953	0.936
No. of Logical Inconsistencies		2	1

[2] All coefficients for both models were significant at the p<0.1.

Table 2: Ordinal and Standard Gamble Health State Valuation Models for SF-6D[3]

	RankCoeff	RescaledCoeff	SGCoeff
Physical Functioning 2	**-0.3636**	**-0.0566**	**-0.0600**
Physical Functioning 3	**-0.4313**	**-0.0671**	-0.0200
Physical Functioning 4	**-0.9856**	**-0.1534**	**-0.0600**
Physical Functioning 5	**-0.6340**	**-0.0987**	**-0.0630**
Physical Functioning 6	**-1.4475**	**-0.2253**	**-0.1310**
Role Limitations 2	**-0.3211**	**-0.0500**	**-0.0570**
Role Limitations 3	**-0.4069**	**-0.0633**	**-0.0680**
Role Limitations 4	**-0.4053**	**-0.0631**	**-0.0660**
Social Functioning 2	**-0.3627**	**-0.0565**	**-0.0710**
Social Functioning 3	**-0.4203**	**-0.0654**	**-0.0840**
Social Functioning 4	**-0.5737**	**-0.0893**	**-0.0930**
Social Functioning 5	**-0.8055**	**-0.1254**	**-0.1050**
Pain 2	**-0.3772**	**-0.0587**	**-0.0480**
Pain 3	**-0.3635**	**-0.0566**	-0.0340
Pain 4	**-0.6520**	**-0.1015**	**-0.0700**
Pain 5	**-0.8187**	**-0.1275**	**-0.1070**
Pain 6	**-1.1912**	**-0.1854**	**-0.1810**
Mental Health 2	**-0.2157**	**-0.0336**	**-0.0570**
Mental Health 3	**-0.3371**	**-0.0525**	**-0.0510**
Mental Health 4	**-0.7016**	**-0.1092**	**-0.1210**
Mental Health 5	**-0.8993**	**-0.1400**	**-0.1400**
Vitality 2	**-0.1740**	**-0.0271**	**-0.0940**
Vitality 3	**-0.2140**	**-0.0333**	**-0.0690**
Vitality 4	**-0.3226**	**-0.0502**	**-0.0690**
Vitality 5	**-0.5267**	**-0.0820**	**-0.1060**
Death	**-6.4240**	**-1.0000**	
n states		249	249
MAE		0.088	0.074
No.>0.05		169	118
No.>0.10		84	52
RMSE		0.110	0.098
LB		106.720	169.570
ICC		0.8300	0.8300
No. of logical inconsistencies		3	5

[3] Coefficients in bold are significant at $p<0.1$

Table 3: Hausman's Test for Independence of Irrelevant Alternatives

Health Utilities Index Mark 2				SF-6D		
Category	Hausman	p		Category	Hausman	p
3	80.39	0.0000		8	126.03	0.0000
7	49.56	0.0007		6	35.2	0.1074
8	15.91	0.8202		.	24.03	0.5741
9	21.14	0.5119		7	30.62	0.2426
2	20.71	0.5388		4	32.32	0.1828
4	26.49	0.2311		5	75.53	0.0000
10	50.64	0.0005		3	110.45	0.0000
5	190.44	0.0000		2	221.1	0.0000
6	221.3	0.0000		-	-	-
1	-299.42	1.0000		-	-	-

Figure 1: Prediction Errors for SG and Rank Models: HUI2

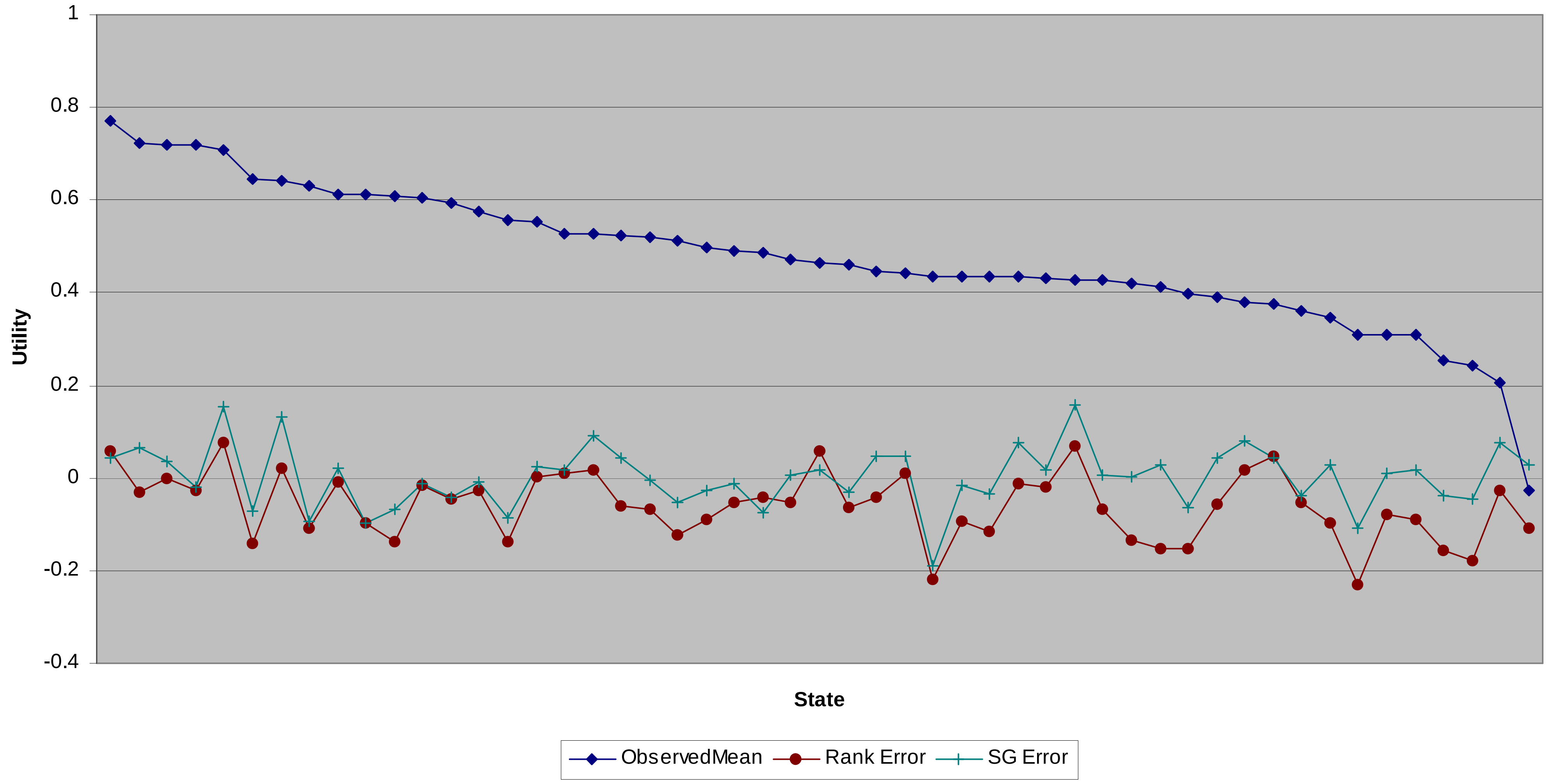

Figure 2: Prediction Errors for SG and Rank models: Sf-6D

Appendix 1: Health Utilities Index Mark 2 (Torrance et al. 1996)

Level	**Sensation**
1	Able to see, hear and speak normally for age
2	Requires equipment to see or hear or speak
3	Sees, hears, or speaks with limitations even with equipment
4	Blind, deaf, or mute

Level	**Self care**
1	Eats, bathes, dresses and uses the toilet normally for age
2	Eats, bathes, dresses or uses the toilet independently with difficulty
3	Requires mechanical equipment to eat, bathe, dress, or use the toilet independently
4	Requires the help of another person to eat, bathe, dress or use the toilet

	Mobility
1	Able to walk, bend, lift, jump and run normally for age
2	Walks, bends, lifts, jumps or runs with difficulty but does not require help
3	Requires mechanical equipment (such as canes, crutches, braces or a wheelchair) to walk or get around independently
4	Requires the help of another person to walk or get around and requires mechanical equipment
5	Unable to control or use arms or legs

	Cognition
1	Learns and remembers schoolwork normally for age
2	Learns and remembers schoolwork more slowly than classmates as judged by parents and/or teachers
3	Learns and remembers very slowly and usually requires special educational assistance
4	Unable to learn and remember

	Emotion
1	Generally happy and free from worry
2	Occasionally fretful, angry, irritable, anxious depressed or suffering from “night terrors”
3	Often fretful, angry, irritable, anxious depressed or suffering from “night terrors”
4	Almost always fretful, angry, irritable, anxious, depressed
5	Extremely fretful, angry, irritable, anxious or depressed usually requiring hospitalisation usually requiring hospitalisation or psychiatric institutional care

	Pain
1	Free of pain and discomfort
2	Occasional pain. Discomfort relieved by non-prescription drugs or self-control activity without disruption of normal activities
3	Frequent pain. Discomfort relieved by oral medicines with occasional disruption of normal activities
4	Frequent pain. Frequent disruption of normal activities. Discomfort requires prescription narcotics for relief
5	Severe pain. Pain not relieved by drugs and constantly disrupts normal activities.

Appendix 2: The Short Form 6D (Brazier et al, 2002)

Level	Physical Functioning
1	Your health does not limit you in vigorous activities
2	Your health limits you a little in vigorous activities
3	Your health limits you a little in moderate activities
4	Your health limits you a lot in moderate activities
5	Your health limits you a little in bathing and dressing
6	Your health limits you a lot in bathing and dressing

	Role limitations
1	You have no problems with your work or other regular daily activities as a result of your physical health or any emotional problems
2	You are limited in the kind of work or other activities as a result of your physical health
3	You accomplish less than you would like as a result of emotional problems
4	You are limited in the kind of work or other activities as a result of your physical health and accomplish less than you would like as a result of emotional problems

	Social functioning
1	Your health limits your social activities none of the time
2	Your health limits your social activities a little of the time
3	Your health limits your social activities some of the time
4	Your health limits your social activities most of the time
5	Your health limits your social activities all of the time

Level	Pain
1	You have no pain
2	You have pain but it does not interfere with your normal work (both outside the home and housework)
3	You have pain that interferes with your normal work (both outside the home and housework) a little bit
4	You have pain that interferes with your normal work (both outside the home and housework) moderately
5	You have pain that interferes with your normal work (both outside the home and housework) quite a bit
6	You have pain that interferes with your normal work (both outside the home and housework) extremely

	Mental health
1	You feel tense or downhearted and low none of the time
2	You feel tense or downhearted and low a little of the time
3	You feel tense or downhearted and low some of the time
4	You feel tense or downhearted and low most of the time
5	You feel tense or downhearted and low all of the time

	Vitality
1	You have a lot of energy all of the time
2	You have a lot of energy most of the time
3	You have a lot of energy some of the time
4	You have a lot of energy a little of the time
5	You have a lot of energy none of the time

Footnote: The SF-36 items used to construct the SF-6D are as follows: physical functioning items1, 2 and 10; role limitation due to physical problems item 3; role limitation due to emotional problems item 2; social functioning item 2; both bodily pain items; mental health items 1 (alternate version) and 4; and vitality item 2.