PLEASE DO NOT CITE OR QUOTE WITHOUT PERMISSION

Valuing health using VAS and rank data: does the VAS contain cardinal information?

Hulme, C. and Edlin, R.P.
Academic Unit of Health Economics, University of Leeds

Introduction

The role of visual analogue scales (VAS) in valuation studies has been extensively discussed. Valuation studies have favoured standard gamble (SG) and time trade off techniques (TTO) over VAS with proponents of the former techniques arguing that the lack of theoretical foundation means that valuations using VAS cannot relate to the underlying theory of QALYs. The lack of observable trade off properties precludes preferences being measured on a cardinal scale. VAS data have also been criticised for the absence of uncertainty and problems with bias due to framing effects including context bias and end state aversion (Parkin and Devlin, 2006).

The inferences, and indeed the premise, of many of these perceived deficiencies have been much debated (Torrance et al, 2001; Parkin and Devlin, 2006; Brazier and McCabe, 2007) and, as part of this on-going debate, recently Brazier and McCabe (2007) asked whether VAS data added anything to rank data; echoing Torrence and colleagues' suggestion that VAS functioned primarily as a prop for ranking exercises (2001).

Previous studies suggest that analyses of ordinal (rank data) can provide valuation functions broadly equivalent to cardinal health state data models (Salomon, 2003; McCabe et al, 2006).

Using a conditional logit model, Salomon proposes estimation of cardinal values be carried out using aggregate data on ordinal rankings. The model uses TTO data from a UK general population survey (Dolan et al, 1994). Within the survey health states were described using the EQ-5D descriptive system (five dimensions: mobility, self care, usual activities, pain/discomfort, anxiety/depression) and respondents asked to describe their own health state before ranking 13 hypothetical states (taken from

42) plus immediate death and unconscious. Rankings were followed by ratings of the same states using a VAS. Finally a series of TTO questions were used for the 13 states. Respondents initially indicated whether each state was preferred to death.

The model assumes each respondent, i, has a latent utility for state, j, U_{ij} that includes a systematic component and an error term $U_{ij} = \mu_j + \varepsilon_{ij}$ and that a given health state has the same expected latent utility value across all respondents. In addition respondent i will choose j over k if $U_{ij} > U_{ik}$ for all $j \neq k$. The rank data are treated as a series of choices. Initially one state, with the best rank, is chosen over all others, then the state with the second rank over all others with the exception of the first and so on. This is equivalent to a ranking of n states and provides n(n-1)/2 discrete preferences. The model assumes that the expected value of the latent utility of each health state is a linear function of the ratings of the EQ-5D domains $\mu_j = x'_j\,\theta$. Before running the regression the data are rescaled in three distinct ways:

1. normalisation to match the scale of observed TTO values in the data
2. normalisation to produce a utility of 0 for the 33333 state
3. normalisation to produce a utility of 0 for death

Additional analyses run the rankings in reverse order to take account of a right skewed distribution in order to consider whether this produces important differences.

Salomon found all three rescaling options predictions were strongly correlated with the observed TTO values (Pearson's r 0.985 for options 1 and 2 and 0.984 for 3). Using ICC, normalisation to match the scale of observed TTO values in the data was the best fit and similar to that of the TTO (0.974 vs. 0.993). Comparison of modelled and observed TTO values by state show that the four states with the largest discrepancies between predicted values from the rank model and mean observed TTO values, all included level 3 on the dimension of pain, and all were states with the largest differences in rank positions between the direct ordering exercise and the TTO. Comparison of inverted and non-inverted rank orderings showed comparable results and when compared to the observed TTO values were almost identical to the main model. These results led Salomon to the conclusion that the information content of aggregate rank data is similar to that of the TTO. However, Salomon highlights a number of important limitations or areas of further research including whether utilities are correlated across health states at the level of the individual (the assumption of independence from irrelevant alternatives) and choice of scale anchors which is complicated by differences in the relative ranking of death between the two methods (unlike the SG and TTO being dead is not a health state for VAS which causes confusion for respondents (Brazier and McCabe, 2007).

Building on Salomon's model, McCabe et al (2006) estimate conditional logistic regression models for HUI2 and SF-6D using ordinal preference data to explore whether the results are comparable to those estimated by SG. Both the HUI2 and SF-6D have six dimensions (HUI2 sensation, mobility, emotion, cognition, self-care and pain; SF-6D physical functioning, role limitations, social functioning, pain, mental health and vitality). HUI2 respondents ranked eight health states from its classification plus full health and immediate death (McCabe et al 2005 a,b) and then valued the same states using SG.

For the SF-6D a representative sample of the UK population were asked to value a sample of SF-6D health states using SG. Respondents were asked to rank five health states plus the best and worse states defined by the SF-6D. Using SG techniques respondents valued each one of five health states with best and PITS health states as alternative outcomes and then valued PITS in relation to death. A sixth SG valuation was dependent on whether respondents had value PITS higher or lower than death.

McCabe and et al's model, like Salomon's, assumes independence of irrelevant alternative (ranking of the pair not affected by other states ranked in the same exercise); again the respondent has a latent utility value for state j, U_{ij} and respondent i will choose j over k if $U_{ij} > U_{ik}$ for all $j \neq k$. In this model the expected value of the unobserved utility is a linear function of categorical level on the dimensions of each dataset: $U_{ij} = \beta x_{ij} + \theta D + \varepsilon_{ij}$ where x vector of dummy variables ($x_{\lambda\delta}$) for each level λ of dimension δ.

In this model co-efficients are rescaled using $\beta_{r\lambda\delta} = \beta_{\lambda\delta}/\theta_D$ where $\beta_{r\lambda\delta}$ is rescaled co-efficient on dimension level $\lambda\delta$ and θ the co-efficient on death. Thus the rescaled co-efficients provide predictions on the same scale as SG or TTO and this anchors death at 0 and full health at 1 and retains possibility of a health state <0.

In the case of the HUI2 the results show similarity between rank and SG models; although the former has one more inconsistency and doesn't distinguish as clearly between levels of mobility. For the SF-6D model the rank model differs to that of the SG. Despite the rank model having a lower number of inconsistencies the predictive performance is better in the SG model. McCabe et al also note that there is evidence in the models that the assumption of independence doesn't hold – that models are sensitive to exclusion of states ranked highly or lowly. They conclude that the rank models performed better than expected given the different informational content and that the results are consistent with the existence of a latent utility function.

The two studies outlined show promise for the use of rank data in valuing health. Whilst previous papers use TTO and SG, it is arguably inefficient to use intensive survey methods in order to obtain rank data only. If rank data are sufficient to create a valuation algorithm, then VAS values may be an appropriate way to obtain this data. Parkin and Devlin (2006) argue that VAS may provide some advantages over more complex methods in areas of feasibility, reliability and practicality (in particular, by allowing postal surveys).

This paper uses the same dataset as Salomon but concentrates on VAS data. We consider whether the information contained in the VAS rank data provides similar results to the cardinal VAS data. That is, do we lose valuable information by using ordinal (rank) data rather than cardinal VAS scores? The ordinal data infers a preference between two states, which provides different information than the VAS score – the ordinal data considers the difference between states but does not consider the intensity of preference between them, whilst the cardinal score does not (in itself) consider differences between states but provides a level of utility. More useful comparisons may be between ordinal preferences (between states) and cardinal differences (between states), and cardinal differences (between states) and the VAS scores (level only).

These cardinal differences have been considered previously. Dolan and Roberts (2002) investigated the cardinal preferences between states and explored whether a tariff with better predictive ability can be calculated using differences between EQ5D states and 33333 rather than the levels themselves. Dolan and Roberts (2004) also consider the degree to which individual ordinal preferences (inferred from TTO responses) correspond to differences in the mean TTO levels across individuals.

They suggest a large degree of heterogeneity in preferences between individuals; this necessitates the use of random effects within this paper.

This paper considers whether ordinal preferences, cardinal differences, and cardinal scores from VAS data provide substantively different valuation algorithms. In the case where a cardinal difference model can provide a substantively better fit than an ordinal preference model, then we can say that the VAS contains useful cardinal information that cannot be incorporated into an ordinal model. A further aim is to assess whether VAS -based ordinal preferences (rather than TTO- or SG-based ordinal preferences) are likely to be sufficient reliably to inform policy.

Data

Whilst the data from the UK general population survey (Dolan et al, 1994) is outlined in brief above, in more detail 3395 interviews were conducted; the sample was representative of the general population. Each interview consisted of five components. The first was self reported health using the EQ-5D descriptive system and VAS; secondly a ranking exercise consisting of 15 pre-determined states (including full health and immediate death) in which each state last 10 years followed by death. This was followed by VAS rating of the same states. The next component was a TTO exercise in which full health and immediate death removed (leaving just 13 states). Collection of personal background data completed the interview.

Methods

The existing literature uses a variety of methods, and some methods may be more or less appropriate to particular types of data than others. As such, the general aim is to identify an appropriate valuation algorithm for each type of data (ordinal preference, cardinal difference, VAS score) and compare their respective predictive abilities against mean EQ-5D VAS scores.

Seven separate random effects logit regressions were carried out for different datasets representing ordinal preferences over dead, full health and the 13 states valued by each individual. Seven similar regressions were run for the cardinal differences between states using a standard random effects model. In both cases the random effect identifies individual respondents.

All datasets were first prepared by scaling individual VAS responses against the values provided for dead (to 0) and full health (to 1). Values worse than death were restricted to fall on the interval (0, -1] using the lower anchor of the VAS scale.

The following groups were excluded:

- Those who failed to give a value for dead, full health, or the pits state (33333)
- Valued fewer than three other states in addition to these three
- Rated death as better than full health, own health as worse than pits, or pits as better than full health.
- Rated fewer than three states better than dead.

Of the 3395 individuals surveyed, 162 were excluded from the full MVH dataset, in comparison to the 84 exclusions in the main study. The seven datasets used in the regressions were constructed as follows.

<u>Pairwise combinations of all states: *"Exhaustive"*</u>

Following Salomon's methodology, our "exhaustive" dataset assumes independence of irrelevant alternatives between any two states. Here, the most preferred state in the ordinal ranking of n states is compared against the (n-1) states it is preferred to, the next state in our ordered list is compared against the (n-2) states (i.e. excluding the most preferred state), the next against (n-3) states and so on. Here the ranking of n states provides n(n-1)/2 preferences; for an individual providing all 15 valuations, this leads to 105 distinct ordinal preferences and cardinal differences.

<u>Extractions from a rank order: *"Proximate"*, *"Random 1"*, *"Random 2"*</u>

Three datasets were prepared from a ranking between states. The "proximate" dataset uses the method from McCabe et al, and utilises an individual orderings between the states. Here, the most preferred state was compared to the next most favoured state; this latter state to the one following, and so on. Here, each state (other than the most and least preferred) appears on two occasions for an individual

– once against the next most preferred state, and once against the next least preferred state. The rank order of n states provides n-1 preferences; for an individual providing all 15 valuations, this leads to 14 distinct ordinal preferences and cardinal differences.

In the "Random 1" and "Random 2" datasets a similar process was used but the states were not ordered by preference. Instead, a random order was generated between the states and the methodology above repeated to again provide n-1 distinct ordinal preferences and cardinal difference where n states were valued.

<u>Comparisons against a defined state: *"vs Full"*, *"vs Dead"*, *"vs Pits"*</u>

When constructing a regression based on cardinal differences Dolan and Roberts (2002) computed preferences against the EQ-5D state 33333. In theory, any state could be used for this purpose, and we consider full health, dead and pits here. In each case, preferences for n states will provide (n-1) ordinal preferences and cardinal differences.

For each of these datasets, dummies were created that reflect the differences between two states in a comparison. Dead is coded as for full health with the exception of the values on the variable "D" (where death is present as a first or second state, see Table 1 below).

For example, suppose that the first state considered is full health, EQ-5D state 11111 and the second state considered is 11311 in which there are no problems in any domain with the exception of usual activities where the respondent indicates 'I am unable to perform my usual activities'. Since the states are identical in all dimensions except usual activities, MO21, SC21, PD21, AD21, MO31, SC32, PD32, and AD31 all take the value 0. UA21 and UA32 both take the value 1, reflecting the higher health in the first state. D = 0 because neither state involves death and N3 = -1 because the second state involves a level 3.

The ORD variable takes the value 1 if the first state is preferred, 0 if the second state is preferred, and 0.5 given indifference. The CARD variable presents the difference in (scaled) cardinal values between the states.

For the final regression predicting VAS scores rather than cardinal differences, all states are coded as above, as if they were compared against state 33333. So here, for 33233 (versus 33333), UA32 = 1 and all other variables take the value 0. The LEVEL variable gives the disutility from full health. (In reporting the coefficients, we multiply all by -1 in order to provide a measure of utility and hence greater comparability to the other results.)

Prior to all regressions except that involving LEVEL, data was randomly "switched" to eliminate a constant term in the cardinal differences regression . A uniform distribution on (0, 1) was estimated and a variable defined as to whether the value was above or below 0.5. If below 0.5, all the variables in the above table and CARD were multiplied by -1; in essence, the first and second states were reversed. In these cases, ORD was recoded as 1 – ORD.

Where items were found to have the wrong sign, they were removed from the regression (in order of least significance to most significant).

Results

Table 2 displays the coefficients from the fifteen regressions. Bolded figures within the table indicate significance at 5%, with removed items left blank. The regressions differ in the number of insignificant and omitted items. In the regression on cardinal variable LEVEL all the EQ-5D variables in Table 1 above are significant with the exception of the N3 term; a similar pattern is observed in the cardinal *Random 1*, *Random 2*, and *vs Dead* models. In the two exhaustive logit models, all variables including the N3 term are significant. At the opposite extreme, the model comparing ordinal differences against the Dead state has only three EQ-5D significant variables with the correct sign: MO32, SC32, and PD32. In the vs Dead model, at least, the cardinal model performs better. A fairer reflection of performance, however, is to consider the accuracy of a range of methodologies.

In order to compare the regression results against mean EQ-5D VAS scores, we estimate predictions for the 42 states (Unconscious was removed as there is no classification for this state on the EQ-5D) valued in the MVH study and rescale these so that death is again valued at 0 and full health at 1.

Table 3 presents a summary of the cardinal difference and ordinal preference model fits, including their range and errors against the mean VAS scores. The range of the unscaled VAS predictions varies greatly, from 0.02-0.23 for the cardinal *Proximate* model to -22 to 27 in the ordinal *vs Full Health* model. Scaled, the ordinal models appear typically place dead as the worst or near the worst possible state (typically valued at around 0.00), whilst in the cardinal models the worst state is typically valued between -0.3 and -0.4.

Of the cardinal differences models, it appears that the *vs Full Health* model provides the lowest root mean squared error (RMSE) at 0.14 and also fewest absolute errors above 0.05, 0.10 and 0.20 (at 57%, 32% and 7% of cases). This model is taken to be our preferred cardinal difference model.

For the ordinal difference models, it is clear that they have a generally higher RMSE (around 0.2 to 0.3 rather than 0.14 – 0.16) and that three models appear to have a better performance than the others. Here, the *Exhaustive*, *Random 1* and *Random 2* models all have an RMSE around 0.22. In all three cases, almost all cases have an absolute error above 0.05 and approximately half of the cases have an absolute error above 0.20. We use the *Exhaustive* model as our preferred ordinal preference model but do so with caveats, as detailed in the discussion below.

The performance of both the cardinal difference and ordinal preference models can be compared to that of the cardinal regression predicting LEVEL. Here, in contrast to other models, a lower proportion of predictions had absolute errors below 0.05 (50%), 0.10 (23%) and 0.20 (2%). Figure 1 presents the cumulative proportion of states by absolute error, where the higher values indicate superior fit. It is clear here that the VAS levels regression generally dominates, with the preferred cardinal difference model outperforming the preferred ordinal preference model.

Discussion

Several features are of note within the models presented above, particularly in relation to which appear to perform poorly for certain types of data. For the cardinal data, the *Proximate* method appears to relatively poorly. This may be a feature of the data used here. In the MVH study, states were selected so that all individuals will consider two very mild states, three mild states, three intermediate states and three severe states. Here, the rankings between similarly severe states will typically involve very low cardinal differences between the states but potentially large differences in the descriptions of the states. As a result, the coefficients in this model are particularly small. This has the effect of reducing the range of unscaled health estimates, and exaggerating the possible range of health state values when anchored at 0 and 1. We might expect, and do observe, smaller unscaled coefficients and a similar lack of fit for the ordinal *Proximate* model. (In contrast the SG-based model used in McCabe et al considered states that were more evenly spread, so that the suppression of coefficient values observed here was not a factor.)

For the ordinal data, the *vs Full Health* model has a very high coefficient on dead that is non-significant. By the exclusion criteria, being in full health is always preferred to being dead, so there is unanimity and the logit regression is unable to estimate a standard error for the coefficient – hence, the lack of significance. The implication of the large coefficient on dead is that the other coefficients also increase in value.

The ordinal *vs Dead* and ordinal *vs Pits* models also perform relatively poorly in terms of the number of insignificant or omitted variables. Here, we can safely predict that most states will be preferred to being dead, that most states predicted by the EQ-5D will be preferred to the worst state describable by the instrument. It is probably unsurprising therefore that these models are unable to extract as much information from the dataset as others, and consequently have a worse performance.

One puzzling finding is the significance of the constant term in the ordinal preference models. Given that the states compared are randomly swapped, there should be no obvious preference for choosing the first or the second state in a comparison. However, the constant terms suggest that the first state is typically preferred. (This is the case even in *Random 1* and *Random 2*, where we look at preferences that start with a random ordering of states, and where we would not expect such a finding even given an error in coding syntax.)

In the constant difference models, the constant term is insignificant except in the case of the *Exhaustive* model. Rather than suggesting poor performance of the model (at 0.001 it is unlikely to have an impact), this may instead suggest a problem with the general method. Here, independence of irrelevant of alternatives is assumed to obtain almost eight times the number of data points as used in other models and this larger "sample size" allows more items to be considered and may exaggerate significance. We considered using a random ranking-based model like Random 1 or Random 2 as our preferred ordinal preference model for a related concern – that the *Exhaustive* model simply extracts too much data from limited preferences.

Overall, we found that the VAS Levels regression outperformed the cardinal differences regression *vs Full Health.* The VAS Levels regression considered disutility, as a "loss" against full health. However, the cardinal difference regression appears to do the same thing – considering a difference versus full health. The former outperformed the latter for two reasons – firstly, taking differences reduces the sample size – instead of having n observations from an individual we have only (n-1) observations, and so less information to estimate the underlying regression. The second reason is structural: the constant in the levels regression is informative because it functions as an "N2" term (in a similar way to the "N3" term). Here, any deviation from full health starts at 0.84 (i.e. 1.00 – 0.16) plus whatever additional decrement is suggested by the coefficients. In the cardinal difference regressions, this difference is lost and hence is not measured – as this role is significant in the levels regression, the fit of the model is poorer without it, and this is reflected in the results.

Conclusions

This paper has compared alternative analyses for different ordinal preferences over health states and cardinal differences between health states, alongside a regression estimating the levels of health. We have identified preferred models for the ordinal preference and cardinal difference states.

In contrast to previous studies we have found that ordinal preference models appear to give different results to cardinal data. The similarly severe states considered in the MVH study may have contributed to this. Unless similarly "close" states can be avoided (e.g. in experimental design), then "exploding" a ranked list of states for use in logit models appears unwise. Similarly, we have concerns about the reliance of an "exhaustive" logit on independence of irrelevant alternatives. McCabe et al found this assumption to be violated, and we have additional concerns over possibly exaggerated significance.

Given that the ordinal preference data performs worse than the cardinal difference and levels regression, it seems relatively clear that: (1) VAS contains at least some relevant and useful cardinal data and (2) ignoring such data worsens performance of the resulting measures. This is the case even when the cardinal data we use may contain range-frequency bias .Correcting this with the Parducci Weddel transformation (Torrence et al 2001) will not affect the ordinal data but could provide "cleaner" cardinal data. In such case, the discrepancy in performance between the ordinal and cardinal data may be even more pronounced. (We will re-run our analyses using PW adjusted data prior to HESG but time prevents this at present.)

A more general issue relates to whether ordinal preference data should be used in valuation tasks. There are a variety of different means by which ordinal preference data could be used – and this paper has not considered the full variety of methods available, with DCE and/or Best-Worst both alternative options (see for example, Ryan et al, 2001).

However, in standard valuation settings, it is likely that rank data will be elicited by post and based on tasks such as VAS rather than on more intensive SG or TTO data. (Previous research of this type in New Zealand has used VAS, with under one third of those contacted providing usable data (Devlin et al, 2000).) Both Salomon and McCabe et al used ranking data based on more intensive methods and both found that the ordinal data obtained were adequate. This is not a guarantee that ranking data per se is adequate, and the findings here are to the contrary.

References

Brazier, J. and McCabe, C. 2007. Is there a case for using visual analogue scale valuations in CUA by Parkin and Devlin a response: Yes there is a case, but what does it add to ordinal data? *Health Economics*. 16: 645-647 (2007)

Dolan, P., Gudex, C., Kind P., Williams, A. 1994. *The measurement and valuation of health: First report on the main survey.* Centre for Health Economics, University of York: York

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

McCabe, C., Stevens, K., Roberts, J., Brazier, J.E. 2005a. Health state valuations for the Health Utilties Index Mark 2 descriptive system: result from a UK valuable survey. *Health Economics.* 14: 231-244 (2005)

McCabe, C., Stevens, K., Brazier, J.E. 2005b. Utility values for Health Utilties Index Mark 2: An empirical assessment of alternative mapping functions. *Medical Care.* 43: 627-635 (2005)

Parkin, D. and Roberts, J. 2006. Is there a case for using visual analogue scale valuations in cost-utility analysis? *Health Economics*. 15: 653-664 (2006)

Roberts, J. and Dolan, P. 2004. To what extent do people prefer health states with higher values? A note on evidence from the EQ-5D valuation set. *Health Economics.* 13: 733-737 (2004)

Ryan, M., Scott, D.A., Reeves, C., Bate, A., van Teijlingen, E.R., Russell, E.M., Napper, M., Robb, C.M. 2001. *Eliciting public preferences for health care: a systematic review of techniques*. Health Technology Assessment; 5(5): 1-186

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 . http://www.pophealthmetrics.com/content/1/1/12 (accessed 10.11.08)

Torrance G.W., Feeny, D.H., Furlong. 2001. Visual analogue scales; do they have a role in the measurement of preferences for health states? *Medical Decision Making.* 21: 329-334

Table 1: Variable definitions

<table>
<tr><th>Name</th><th>Levels</th><th>Coding</th></tr>
<tr><td>MO21</td><td>Difference in mobility: levels 2 and 1</td><td rowspan="5">1 first state is level 1, second state level 2 or 3
-1 first state is level 2 or 3, second state is level 1
0 all other cases</td></tr>
<tr><td>SC21</td><td>Difference in self care: levels 2 and 1</td></tr>
<tr><td>UA21</td><td>Difference in usual activities: levels 2 and 1</td></tr>
<tr><td>PD21</td><td>Difference in pain/discomfort: levels 2 and 1</td></tr>
<tr><td>AD21</td><td>Difference in anxiety/depression: levels 2 and 1</td></tr>
</table>

<table>
<tr><td>MO32</td><td>Difference in mobility: levels 3 and 2</td><td rowspan="5">1 first state is level 1 or 2, second state level 3
-1 first state is level 3, second state is level 1 or 2
0 all other cases</td></tr>
<tr><td>SC32</td><td>Difference in self care: levels 3 and 2</td></tr>
<tr><td>UA32</td><td>Difference in usual activities: levels 3 and 2</td></tr>
<tr><td>PD32</td><td>Difference in pain/discomfort: levels 3 and 2</td></tr>
<tr><td>AD32</td><td>Difference in anxiety/depression: levels 3 and 2</td></tr>
</table>

<table>
<tr><td>D</td><td>Death present as first or second state</td><td>1 state 1 is dead
-1 state 2 is dead
0 death not present in either state</td></tr>
<tr><td>N3</td><td>Is an EQ-5D 3rd level present</td><td>1 first state has an EQ-5D level 3, second state has no EQ-5D level 3
-1 first state has no EQ-5D level 3, second state has an EQ-5D level 3
0 all other cases</td></tr>
</table>

Table 2: Regression coefficients from random effects and random effects logit models

		Exhaustive	Ranking:			Preferences vs:		
		Logit	Proximate	Random 1	Random 2	vs Full health	vs Dead	vs Pits
CARDINAL	LEVEL	CARD	CARD	CARD	CARD	CARD	CARD	CARD
MO21	**0.01**	**0.01**		**0.01**	**0.01**	**0.04**	**0.01**	0.01
MO32	**0.24**	**0.24**	**0.05**	**0.25**	**0.24**	**0.21**	**0.24**	**0.24**
SC21	**0.06**	**0.06**	**0.01**	**0.06**	**0.06**	**0.13**	**0.06**	**0.05**
SC32	**0.04**	**0.04**	**0.02**	**0.03**	**0.04**	0.01	**0.04**	**0.04**
UA21	**0.18**	**0.18**	**0.03**	**0.17**	**0.18**	**0.19**	**0.18**	**0.18**
UA32	**0.05**	**0.05**	**0.02**	**0.05**	**0.05**	**0.05**	**0.05**	**0.06**
PD21	**0.02**	**0.02**	0.00	**0.02**	**0.01**	**0.05**	**0.01**	**0.02**
PD32	**0.26**	**0.26**	**0.05**	**0.26**	**0.25**	**0.24**	**0.26**	**0.26**
AD21	**0.10**	**0.10**	**0.01**	**0.10**	**0.11**	**0.16**	**0.10**	**0.10**
AD32	**0.19**	**0.19**	**0.04**	**0.18**	**0.18**	**0.16**	**0.19**	**0.19**
D	**-0.85**	**-0.85**	**-0.07**	**-0.84**	**-0.85**	**-1.01**	**-0.84**	**-0.84**
N3	0.00	0.00	**0.02**	0.00	-0.01	**-0.06**	0.00	0.01
Constant	**-0.16**	**0.00**	0.00	0.00	0.00	0.00	0.00	0.00
ORDINAL		ORD	ORD	ORD	ORD	ORD	ORD	ORD
MO21		**0.05**	0.02		0.01	**3.77**		
MO32		**1.54**	**0.50**	**1.56**	**1.58**		**0.90**	**1.28**
SC21		**0.39**	**0.27**	**0.36**	**0.44**	**3.80**		
SC32		**0.11**	0.03	**0.11**	**0.08**		**0.38**	**0.57**
UA21		**1.09**	**0.34**	**1.07**	**1.11**	**3.40**	0.66	**0.97**
UA32		**0.29**	**0.12**	**0.30**	**0.31**	**1.56**	0.09	**0.44**
PD21		**0.10**	**0.14**	**0.14**	**0.10**	**2.65**	0.02	0.02
PD32		**1.51**	**0.42**	**1.50**	**1.48**	**5.00**	**0.92**	**1.35**
AD21		**0.64**	**0.26**	**0.64**	**0.70**	**3.27**	0.01	
AD32		**1.04**	**0.28**	**0.96**	**1.02**	**3.61**	0.83	**1.18**
D		**-6.33**	**-2.90**	**-6.26**	**-6.49**	-49.24	-4.88	**-5.83**
N3		**-0.03**	**-0.09**	-0.07	-0.06	**2.86**	-0.86	0.22
Constant		**0.09**	**0.23**	**0.09**	**0.07**	0.12	**0.11**	**0.26**

Bolded figures are significant at 5%.

Table 3: Fit of the cardinal difference and ordinal preference models

	Exhaustive	Ranking:			Preferences vs:		
	Logit	Proximate	Random 1	Random 2	vs Full health	vs Dead	vs Pits
CARDINAL DIFFERENCES							
MAX	1.00	1.00	1.00	1.00	1.00	1.00	1.00
MIN	-0.35	-1.90	-0.36	-0.35	-0.28	-0.35	-0.36
UNSCALED ESTIMATES							
MAX	1.14	0.23	1.14	1.14	1.23	1.14	1.15
MIN	0.00	0.02	0.00	-0.01	-0.06	0.00	0.00
ERRORS							
MAX	0.54	1.59	0.55	0.54	0.57	0.54	0.54
RMSE by state	0.16	0.54	0.16	0.16	0.14	0.16	0.16
%errors above 0.05	73%	84%	73%	73%	57%	75%	73%
%errors above 0.10	48%	73%	45%	52%	32%	50%	48%
%errors above 0.20	18%	59%	18%	18%	7%	18%	18%
ORDINAL PREFERENCES							
MAX	1.000	1.000	1.000	1.000	1.000	1.000	1.000
MIN	-0.073	0.000	-0.072	-0.059	0.000	0.000	0.000
UNSCALED ESTIMATES							
MAX	6.844	2.623	6.729	6.888	27.169	3.923	6.084
MIN	0.055	-0.273	0.020	0.014	-22.073	-0.956	0.252
ERRORS							
MAX	0.580	0.637	0.582	0.583	0.813	0.617	0.586
RMSE by state	0.217	0.268	0.217	0.218	0.424	0.266	0.323
%errors above 0.05	93%	95%	93%	95%	95%	95%	95%
%errors above 0.1	84%	91%	86%	84%	93%	95%	95%
%errors above 0.2	48%	66%	48%	48%	64%	68%	84%

Figure 1: Fit of the preferred models