Valuing health states from different instruments using rank data.*

Mónica Hernández
Department of Economics, University of Sheffield
John Brazier, Donna Rowen, Aki Tsuchiya
Health Economics & Decision Science, University of Sheffield

Paper Number: P24

ABSTRACT: Health policy is increasingly being informed by economic evaluation that measures outcomes using the Quality Adjusted Life Year (QALY). The QALY combines quantity and quality of life into a single measure of health outcome using weights to adjust for the quality of life. These weights are estimated using preference-based measures of health-related quality of life. However, it has been shown that different preference-based measures can generate different weights. We analyse primary data collected for the purpose of this study in which respondents are asked to rank a mixture of states described using different preference-based measures. We develop a rank ordered mixed logit model and obtain estimated values for all the states included in the dataset on a common metric so that direct comparisons across the preference-based measures included in the study are possible. Mapping functions across the preference-based measures are also estimated jointly by allowing for measurement error and clustering around the preference-based measures.

KEYWORDS: Rank ordered mixed logit model. Normal error component logit-mixture. Preference-based measures of health. Quality of life.

JEL CLASSIFICATION: C25, C51

ADDRESS FOR CORRESPONDENCE: Mónica Hernández. Department of Economics. University of Sheffield. 9 Mappin Street. Sheffield S1 4DT.
E-mail: monica.hernandez@sheffield.ac.uk

* We are grateful to the Medical Research Council for financial support of this research. We would like to thank Professor Coast and Professor Netten for the use of ICECAP and OPUS respectively. We would also like to thank all the interviewees who took part in the survey. All responsibility for the analysis and interpretation of the data presented here lies with the authors.

1 Introduction

Health policy is increasingly being informed by economic evaluation that measures outcomes using the Quality Adjusted Life Year (QALY). The QALY combines quantity and quality of life into a single measure of health outcome using preference weights to adjust for the quality of life. These weights are estimated using preference-based measures of health-related quality of life, generating a single index score which can be compared across different health care interventions or programmes. One advantage of preference-based measures is that they reflect peoples' strength of preference for different outcomes and produce a valuation rather than simply a measurement of health. All preference-based measures are valued on an interval scale where full health is the upper anchor with an assigned value of 1 and 0 is the lower anchor, usually assumed equivalent to dead.

Many different preference-based measures are commonly used since there is no common agreement on the use of a single measure for all patient groups and conditions. Indeed, in recent years a large number of new preference-based measures of health have emerged. The quality adjustment weights generated by generic preference-based measures are different for the same patients (see for example Longworth and Bryan, 2003, O'Brien et al, 2003, Brazier et al, 2004, Barton et al, 2004 and Espallargues et al, 2005). A key reason for these differences is that these measures were valued using different valuation techniques, at different points in time and using different samples of respondents (often from different countries). In other words, they have not been obtained using the same yardstick, so it is unlikely they are going to be the same.

Comparisons across studies using different preference-based measures would be inaccurate if they assume that a QALY calculation is unaffected by the preference-based measure used to generate the quality adjustment weight. What is lacking is a way of relating the responses on one preference-based measure to another by using a common metric whilst preserving the advantages of the descriptive system of each preference-based measure.

In this paper we analyse primary data collected by interview in which respondents are asked directly to rank health states described using different preference-based measures. This means that the relationship between different instruments is determined directly by people's preferences for different hypothetical states. Ranked data is often analysed using a rank ordered logit model. However, there are two limitations of this type of logit model that are specially relevant in our case: the independence of irrelevant alternatives and the inability to accommodate repeated choices unless they are assumed to be independent. If these restrictions are rejected by the data, the rank ordered logit model will give inconsistent estimates of the parameter values and inferences based on this model could be misleading. For this reason we develop a rank ordered mixed logit model tailored to our dataset which relaxes these two assumptions. The aims of this paper are to estimate values for all the health states included in the dataset on a common metric so that direct comparisons across the preference-based measures included in the study are possible and to use this common metric to map across preference-based measures. For this purpose, the estimated parameters of the rank ordered mixed logit model and their covariance matrix are used to estimate joint mapping functions by maximum likelihood allowing for measurement error in the dependent variable and clustering in the error term around the preference-based measures.

2 Data

We analyze primary data collected by interview in geographical areas of the North of England during June to October 2007. We followed a thorough process of data checking with the purpose of detecting any errors. The process highlighted a few potential errors which were corrected only in cases where we had sufficient evidence to believe that a recording error had been made. Full details of the data collection process, the study design and characteristics of the respondents can be found in Rowen et al, 2009. Here we present a brief summary of the data relevant to this paper.

The study involves six preference-based measures of health and quality of life: EQ-5D (generic), SF-6D (generic), HUI2 (generic for children), AQL-5D (asthma specific), OPUS (social care specific), ICECAP (capabilities). The choice of preference-based measures reflects a range of different types of measures that are currently in use or have been recently developed for use in the UK. These are summarised in Table 1. Each respondent was asked to perform three ranking and valuing tasks. In each task, respondents were shown 8 cards with descriptions of 8 states and were first asked to rank them. The survey design contains 20 variations of the ranking and valuing tasks. Ties in the ranking are possible if states are considered equal by a respondent. Once respondents had completed the ranking, they were also asked to value the same set of states using a visual analogue scale (VAS). This entails the respondent rating each state on a vertical line on a scale from 0 ('worst imaginable state') to 100 ('best imaginable state'). Respondents were not required to maintain the same ordering as in the ranking. Each interview involves hypothetical states from three of the six preference-based measures indicated above. The eight states in each task always include two generic states 'best state' and 'dead'. The remaining six states comprise three states from two different preference-based measures. For each of the two preference-based measures, the three states always consisted of the worst state and one mild and one moderate state within the preference-based measure. Although the worst states for all the preference-based measures are included in the study, there are only two preference-based measures for which the best state defined by the classification system of the preference-based measure is included, EQ-5D and OPUS. These states measure the absence of health and social care problems respectively. The published state value for these two states is 1 since they are used as the top anchors in their respective preference-based measures. One of the issues that we aim to determine in this study is whether these two states are regarded equally according to people's preferences.

The dataset contains data for 501 individuals. Two respondents were dropped from the sample since they had no ranking or rating data. In addition, not all respondents completed all three ranking tasks; four completed only one of the three tasks and one failed to complete one of the tasks. One respondent only ranked one card ('best imaginable state') in a task and another respondent only ranked two cards ('best imaginable state' and 'dead') and therefore, these two tasks were dropped from the sample. There are also seven instances of a state not being ranked in a task. In these cases we exclude the state description from the choice set of the respondent and keep the remaining observations in the task. After these deletions, the final sample size of pseudo-observations is 11,884.

There are several features of the data which are as expected but will nevertheless become important in the model specification and are thus described here. The generic 'best state' is ranked first 99.73% of the times: 98.38% of the times it is ranked first on its own and the remaining 1.35% of the times it is ranked first but tied with another state. It is only on four occasions that another state is ranked higher than 'best state', three of which are observations for the same individual. It is interesting to note that out of the 20 times the generic 'best state' is tied with other health instruments, 12 times it is tied with EQ5D 11111 and 4 times it is tied with OPUS 1111.

The bottom of the ranking however presents more variation than the top. Table 2 shows the frequencies with which instruments are ranked last. This table only includes the cases with no ties at the end of the ranking which amounts to 93% of the bottom rankings. The highest percentage corresponds to 'dead' being ranked last in 1,120 occasions. This increases to 1,197 when ties are included making the frequency 80.55% of the total sample. There are 11 states under the heading 'Other' in Table 2 that are ranked last as well but each one only on one occasion. The remaining states that are ranked last are the worst states of the six preference-based measures included in the study.

3 Model specification

Ranked data is often analysed using a logit model. It is assumed that individual i faces J different alternatives in each of the T choice situations or time periods. In our case, the sets of alternatives and the number of choice situations differ across individuals and therefore we should use J_{it} and T_i but for simplicity of exposition we will use J and T. The utility that individual i gets from alternative j in choice situation t can be decomposed into two parts: a deterministic part, ν_{ijt}, which typically is a linear function of some fixed parameters β and an unknown stochastic part, ε_{ijt}

$$U_{ijt} = \nu_{ijt} + \varepsilon_{ijt} \tag{1}$$

In each choice situation, the individual chooses the alternative with the highest utility; individual i chooses alternative j in choice situation t, if and only if $U_{ijt} > U_{ikt}\ \forall k \neq j$. The logit model is obtained under the assumption that each ε_{ijt} is independent and identically distributed (IID) type I extreme value. The logit model for ranked data (Luce, 1959, Plackett, 1975) can be easily obtained from the standard logit model. Let r_{it}^{l} be the alternative ranked in lth position and $\mathbf{R}_{it} = \left\{r_{it}^{1}, r_{it}^{2}, \ldots, r_{it}^{J}\right\}$ be the ranking of the J alternatives from best to worse. The probability of this ranking can be written as the product of the logit probabilities of choosing one alternative at a time from successively smaller subsets of alternatives

$$\Pr\left(\mathbf{R}_{it}\right) = \prod_{l=1}^{J-1} \frac{\exp\left(\nu_{ir_{it}^{l}t}\right)}{\sum_{s=l}^{J} \exp\left(\nu_{ir_{it}^{s}t}\right)} \tag{2}$$

Effectively, each ranking is expressed as $J-1$ independent choices or pseudo-observations by the individual and for this reason this model is often called the exploded logit (see for example Chapman and Staelin, 1982).The probability of observing a set of rankings by an individual can be written as

$$\mathbf{P}_i = \prod_{t=1}^{T} \prod_{l=1}^{J-1} \frac{\exp\left(\nu_{ir_{it}^{l}t}\right)}{\sum_{s=l}^{J} \exp\left(\nu_{ir_{it}^{s}t}\right)} \tag{3}$$

There are two limitations of the simple logit model that are specially relevant for modelling purposes. The first limitation is the property of Independence of Irrelevant Alternatives (IIA) which implies that the relative odds of choosing one alternative over another depend neither on the rest of the alternatives in the set, nor on the alternatives already chosen in the ranking. In many situations, this property might represent behaviour correctly, but in cases where some alternatives are close substitutes, this property will be too restrictive.[1] In our case, it is essential to allow for correlations across alternatives due to the nature of the dataset. Respondents are ranking across preference-based measures and alternatives (states) across these preference-based measures could be viewed as very similar.

A second limitation of the logit model relates to the way it handles repeated choices. In the present case, each individual performs three different ranking tasks and in addition each ranking task is exploded into a number of pseudo observations. If there are unobserved factors affecting each decision and these factors are correlated over choices, the logit model will be misspecified since the error terms for any set of choices for a given individual are assumed to be independent. This second limitation could be handled using

[1] This property is often only discussed in relation to the logit model and sometimes, implicitly assumed to be a consequence of the shape of the distribution. However, this restrictive property stems from the assumption of independent errors and applies to many more models including a probit model with independent errors (Hausman and Wise, 1978).

clustering but the first limitation would remain. Misspecification of the model will lead to inconsistent estimates and makes inferences unreliable.

One way of relaxing these two limitations is by using a mixed logit model. Early applications of this model can be found in Boyd and Mellman, 1980 and Cardell and Dunbar, 1980. McFadden and Train, 2000, contributed significantly to this literature by showing that a mixed logit model can approximate any random utility model. The model is usually derived from a random coefficients perspective by replacing the fixed parameters β in the deterministic part ν_{ijt} by random parameters that vary across individuals. Although the interpretation of the model is different, a mixed logit model can also be derived from an error components perspective. Both derivations generate an additional stochastic term, ξ_{ijt}, in the utility in equation (1) to give

$$U_{ijt} = \nu_{ijt} + \xi_{ijt} + \varepsilon_{ijt}$$

This second component can be correlated between alternatives and time and can be heteroskedastic. The presence of this second stochastic term generates a pattern of correlation in the utilities over alternatives and can also generate correlation between the utilities of an individual across choice situations. This additional random term is assumed to have zero mean and a distribution $f(\xi|\Psi)$ where Ψ are some fixed parameters that determine this distribution and need to be estimated together with β. Conditioning on ξ the probability of a given choice is logit and the probability of observing a certain set of rankings is analogous to that in equation (3)

$$\mathbf{P}_i(\xi) = \prod_{t=1}^{T} \prod_{l=1}^{J-1} \frac{\exp\left(\nu_{ir_{it}^{l}t} + \xi_{ir_{it}^{l}t}\right)}{\sum_{s=l}^{J} \exp\left(\nu_{ir_{it}^{s}t} + \xi_{ir_{it}^{s}t}\right)}$$

Since ξ is not known, $\mathbf{P}_i(\xi)$ needs to be integrated over the density of ξ to obtain the unconditional probability of the sequence of choices for person i

$$\mathbb{P}_i = \int \mathbf{P}_i(\xi) f(\xi|\Psi) \, d\xi$$

Thus, the choice probability of a mixed logit model is a mixture of logits and $f(\xi|\Psi)$ is the mixing distribution. The model is usually estimated by simulated maximum likelihood since the loglikelihood of the model involves a multiple integral. In the empirical application Halton draws are used which have been shown to have good coverage.

3.1 Ties in the data.

To our knowledge, this is the first application of a mixed logit model to rank data. Rank data brings an additional complication to the model. Theoretically, the probability of tied alternatives is zero but in practice ties in the data are often observed in rankings.

Hausman and Ruud, 1987, proposed a generalisation of the likelihood of the logit model for tied alternatives based on the marginal likelihood principle taking advantage of the duality between the logistic model for rankings and the partial likelihood of Cox regression. It is assumed that the individual has a preferred order of the alternatives but the econometrician does not observe it. The contribution of the tied alternatives to the likelihood is obtained by adding the probabilities of all possible permutations of the ranked alternatives. If there are ties in the ranking, individual i will assign only L different ranks to the J different alternatives ($L < J$). Use K_l to denote the number of of tied alternatives in rank l. Let $p = (p_1, \ldots, p_{K_l})$ be an element of Q_l, the set of permutations of the numbers $1, \ldots, K_l$ so that out of all the alternatives with rank l, $r_{it}^l [p_k]$ denotes the one that appears on the p_kth position in a permutation p. The probability of a ranking $\mathbf{R}_{it}$ in equation (2) can be generalised to

$$\Pr(\mathbf{R}_{it}) = \prod_{l=1}^{L} \sum_{p \in Q_l} \prod_{k=1}^{K_l} \frac{\exp\left(\nu_{ir_{it}^l[p_k]t}\right)}{\sum_{s=k}^{K_l} \exp\left(\nu_{ir_{it}^l[p_s]t}\right) + \sum_{s>l} \sum_{m=1}^{K_s} \exp\left(\nu_{ir_{it}^s[p_m]t}\right)} \tag{4}$$

The loglikelihood for the rank ordered mixed logit model allowing for ties is found by using equation (4) instead of (2). The expression in equation (4) is complex and computationally demanding and approximations have been suggested in the literature (Breslow, 1974, Efron, 1977). However, Farewell and Prentice, 1980, showed that these approximations can be inaccurate. In our dataset almost 31% (154) of the respondents were found to have at least one tie in one of the tasks. Given the large proportion of respondents with ties it seems sensible to avoid using approximations. To gauge the difference it could make, the rank ordered logit model was estimated using both Efron's approximation and using the marginal likelihood principle. The use of Efron's approximation generates differences in the underlying estimated parameters of up to 1.81. These differences translate into large differences in the scaled parameters; up to 0.12 if we include 'best state' or up to 0.08 if we exclude it. Although the computational cost is great, it seems prudent to avoid approximations to the likelihood function in the present case.

3.2 Specification of the model.

There are no characteristics of the alternatives which vary over observations, only alternative specific constants are used. Although each respondent only sees a maximum of 17 different alternatives, the number of total alternatives across all respondents is 83. Allowing all alternatives to have random coefficients generates correlations in repeated choices for alternatives that have been seen before, however it does not relax the IIA assumption unless the random effects are allowed to be correlated being necessary in this case to estimate a full covariance matrix. Having such a large number of alternatives makes it impractical to follow this route. In this paper we take an error components approach and try to allow for a flexible enough structure in the covariance matrix of the utilities so that we obtain consistent estimates while keeping a parsimonious specification. Using alternative specific constants, the utility in equation (1) for the rank ordered logit becomes

$$U_{ijt} = \beta_j + \varepsilon_{ijt} \tag{5}$$

Our rank ordered mixed logit specification adds a number of error components to this utility. First, an individual latent factor, ξ_{1i}, is added to the utility in equation (5). It represents a characteristic of the respondent that affects his/her choices but it is not observed. This characteristic enters all utilities but with different factor loadings, τ_j, so that its impact differs by alternative. The utility in equation (5) becomes

$$U_{ijt} = \beta_j + \tau_j \xi_{1i} + \varepsilon_{ijt} \tag{6}$$

In addition, given the discussion in the data section, we allow for a nested structure in the mixed logit model. The alternative 'best state' appears to be sufficiently different to the rest of alternatives to warrant its own nest. The two top states EQ-5D 11111 and OPUS 1111 are placed in a second nest. A third nest includes the alternative 'dead' and the worst states of all six instruments and the final nest encompasses all the remaining alternatives. These nests are allowed to have different variances, $\sigma^2_{s_j}$ $s_j = 2, \ldots, 5$, unlike the nested logit model. The utility in equation (6) augmented with the nested structure becomes

$$\begin{aligned} U_{ijt} &= \beta_j + \tau_j \xi_{1i} + \sigma_{s_j} \xi_{s_j i} + \varepsilon_{ijt} \qquad s_j = 2, \ldots, 5 \\ s_j &= \begin{cases} 2 & \text{if } j = \text{ 'Best State'} \\ 3 & \text{if } j = \text{ 'EQ-5D 11111' or 'OPUS 1111'} \\ 4 & \text{if } j = \text{ 'Dead' or any of the worst states} \\ 5 & \text{otherwise} \end{cases} \end{aligned} \tag{7}$$

We assume that all the $\xi_{s_j i}$ have a standard normal distributions and are independent of each other.

3.3 Identification of the model.

Not all the parameters of the rank ordered logit or the rank ordered mixed logit are identified theoretically but identification of the simple rank ordered logit is well established and straightforward; the model needs to be normalised for level and scale. This is usually accomplished by setting one of the alternative specific constants to zero and the variance of the error term to $\pi^2/6$. Alternatively, the same normalisation can be achieved by setting two of the alternative specific constants to two different values (for example 0 and 1) and allowing the variance of the error term to be estimated freely as $(\sigma_1 \pi)^2/6$. To normalise the level of the ranked ordered logit, the constant for the alternative 'dead' is set to zero, so the other alternative specific constants are measured relative to 'dead'. We can set the scale of the model by setting to one the constant of one or both of the top states (either OPUS 1111 or EQ-5D 11111) and directly estimating the scale parameter, or by setting the scale parameter to one. We use the latter since it is straightforward to calculate the scaled parameters and their standard errors using the delta method.

The rank ordered mixed logit model also needs the same identification restrictions but additional restrictions might be needed to identify the covariance structure of the error components. When additional restrictions are needed Walker et al, 2007, showed that an equality condition needs to be checked to ensure that the proposed normalisation does not change the structure of the model. The rank ordered mixed logit model can be written using a factor analytic form as follows

$$M_i U_i = M_i \beta + M_i F V \xi_i + M_i \varepsilon_i$$

where U_i is a $(J_{it}T_i \times 1)$ vector of utilities, M_i is a respondent specific identity matrix with the rows corresponding to alternatives not seen by the ith respondent deleted, β is a $(J_{it}T_i \times 1)$ vector of unknown time-invariant alternative specific constants, F is a $(J_{it}T_i \times 5)$ matrix of fixed factor loadings, V is a (5×5) diagonal matrix containing all σ_{s_j} $(s_j = 2, \ldots, 5)$ and 1 as the diagonal elements, ξ_i is a (5×1) vector of IID standard normal random variables and ε_i is a $(J_{it}T_i \times 1)$ vector of IID type I extreme value random error. The unknown parameters to be estimated are found in the three matrices β, F and V.

Theoretical identification of all the parameters of the model requires that the rank of the Jacobian of the covariance matrix of utility differences equals the number of parameters to be estimated minus one (rank condition). In our case, the covariance matrix of utility differences can be written as:

$$cov\left(\Delta U_i\right) = \Delta M_i F V V' F' M_i' \Delta' + \Delta \frac{\left(\sigma_1 \pi\right)^2}{6} I_{J_{it}T_i} \Delta'$$

Checking theoretical identification requires checking the rank condition for all 20 sets of different ranking tasks. Intuitively, it is straightforward to see that one of the factor loadings,τ_j, in equation (7) will need to be normalised to zero for identification purposes. The factors enter the utility function in the same way as any observable characteristic and as a result an analogous identification restriction is needed. It can be shown that the rank of the Jacobian of the covariance matrix of utility differences for the rank ordered mixed logit model presented here is such that only one of the factor loadings needs to be normalised to achieve theoretical identification of the model and that all σ_{s_j}'s are theoretically identified. The normalisation needed in the factor loadings have been shown to be arbitrary (Walker et al, 2007) and since no additional identification restrictions are needed, the equality condition will hold. We set the factor loading of the alternative 'dead' to zero so that the factor loadings reflect relative preferences.

Although our model can be shown to be identified theoretically, empirical identification cannot be shown until after estimation of the model. However, this issue which is sometimes overlooked, is particularly important because estimation by simulation can conceal identification issues if the number of replications is not large enough (Chiou and Walker, 2007). In the empirical section we use a large number of replications to estimate the model and check empirical identification by re-estimating the model with an increased number of replications.

3.4 Mapping functions using a common scale.

After consistent estimation of the parameters of the model using a rank ordered mixed logit model, we need to estimate the relationship between the original value set for each of the preference-based measures used in this study and the true parameter values β of the preference-based measures in our common metric. We want to estimate a relationship of the following type:

$$\beta = X\alpha + \epsilon \tag{8}$$

where β is the $((J-1) \times 1)$ vector of true parameters, α is a $k \times 1$ vector $(k < J)$ of parameters of interest, X is a block diagonal matrix containing the independent variables used for the mapping regressions of each of the six preference-based measures:

$$X = \begin{pmatrix} X_1 & 0 & \cdots & 0 \\ 0 & X_2 & \cdots & 0 \\ & & \ddots & \\ 0 & 0 & \cdots & X_6 \end{pmatrix}$$

and the ϵ's are iid normally distributed, mean zero and heteroskedastic with covariance matrix

$$\Omega = \begin{pmatrix} \theta_1 I & 0 & \cdots & 0 \\ 0 & \theta_2 I & \cdots & 0 \\ & & \ddots & \\ 0 & 0 & \cdots & \theta_6 I \end{pmatrix}$$

We do not have the true parameter vector β, instead we have an estimate $\hat{\beta}$ so that $\hat{\beta} = \beta + \zeta$. Substituting this expression in (8), we get the following regression

$$\begin{aligned} \hat{\beta} &= X\alpha + \epsilon + \zeta \\ &= X\alpha + \omega \end{aligned}$$

where ω is a composite error term. It is assumed that ϵ and ζ are independent and normally distributed so that the errors in the original estimation of the published values are not correlated with the errors in our rank ordered mixed logit model. We cannot use OLS to estimate the parameter vector α since the composite error term ω does not follow the Gauss-Markov assumptions. We can, however, use maximum likelihood to estimate α together with the six θ's consistently using $\Omega + Cov\left(\hat{\beta}\right)$ as an estimate of the covariance matrix of ω, where $Cov\left(\hat{\beta}\right)$ is calculated by using the inverse of the Hessian matrix of the rank ordered mixed logit model.

4 Results

4.1 Rank ordered mixed logit

The analysis presented here uses 500 Halton draws. We drop the first 15 to avoid issues of correlation in the draws (Train, 2003). Figure 1 plots the value of the likelihood function at the estimated parameter values for different number of Halton draws. The value of the likelihood seems to stabilise by 500 draws although there are still some small but sharp changes which might signal the need for more replications. Currently the model is being re-estimated using a higher number of Halton draws as a check.

Figure 1: Value of the log-likelihood function at the estimated parameter values for different numbers of draws.

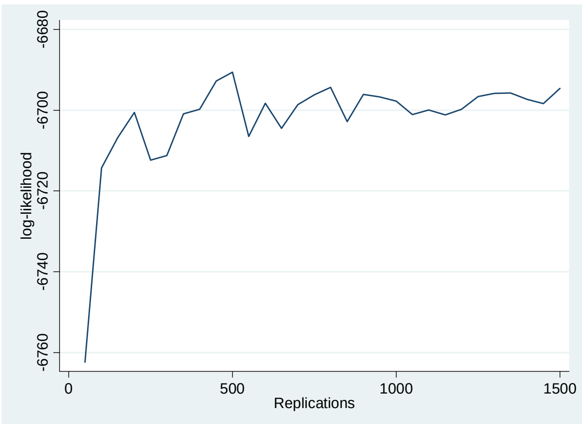


Table 3 presents the estimated parameters of the rank ordered mixed logit model rescaled so that the difference in expected utility between the most preferred state from a preference-based measure (OPUS 1111) and 'dead' equals one. Based on the significance of the estimated factor loadings, τ_i, it is clear that IIA will be rejected in support of our rank ordered mixed logit model making inferences based on the restrictive rank ordered logit model unreliable. A likelihood ratio test between the two models emphatically rejects the rank order logit with a test statistic of 1140.28 and zero p-value. Unfortunately, some of the parameters we are testing are on the boundary of the parameter space which distorts the distribution of the statistic. In these cases the test has been found to be conservative (Andrews, 2001) and therefore would not change the present conclusion.

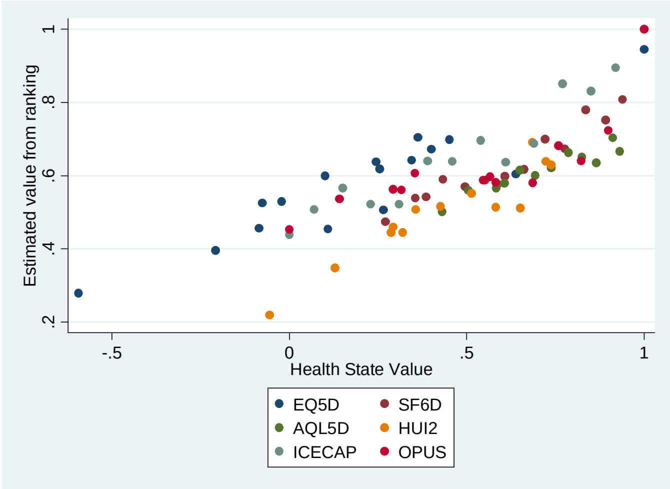
Although looking at the standard errors it seems that some of the error components which define the nests are not significant at standard levels, we need to interpret these standard errors with caution since we are testing on the boundary of the parameter space and therefore the statistics do not have the usual distributions. For this reason, we also estimated a rank ordered mixed logit model with only a latent factor (see equation (6)) and compared it with the full model. The likelihood ratio test just rejects the model with only the latent variable at 5% with a χ^2 test statistic of 11.4457 and p-value of 0.04. Thus, our rank ordered mixed logit model is a clear improvement on the basic model and even on the rank ordered mixed logit with only the latent factor.

There are substantial differences in the estimated states values between the rank ordered logit model and the rank ordered mixed logit model of up to 0.14 with both a median and a mean difference of 0.05.

There are several features of our model that are worth noting. First, statistically, there is no significant difference between the top two states, EQ-5D 11111 and OPUS 1111 and therefore we could impose this restriction and anchor them both at 1. Another feature is that the factor loadings of the latent factor are significant apart from the factor loading of the 'best state'. This is an important issue which indicates that 'best state' and 'dead' are so different that IIA would be a reasonable assumption between these two alternatives but not for the rest of states. To get a better feel for the model we can look at the correlations between utility differences. If IIA holds between alternatives, the correlation between utility differences should be 0.5 since preference of alternative A over alternative B would not imply a pattern of preference between alternative C and B. In other words half of the respondents who prefer A to B would be expected to prefer C to B and the other half would be expected to prefer B to C. Therefore, any departures from 0.5 in the correlations between utility differences points towards rejection of IIA. Table 4 shows the implied correlations of the utility differences between all the PITS of all the preference based measures and 'dead'. All these correlations are very high and clearly different from 0.5; a respondent who prefers one of the PITS to 'dead' is more likely to prefer all the rest of the PITS to 'dead'. This pattern of correlations is not unique to the worst states; the median correlation between all utility differences in our model equals 0.89, a clear departure from 0.5 for a large number of utility differences.

Figure 2 plots the current published state values against our estimated values on a common metric for all states included in the study. The valuation of the worst state of EQ-5D is not as low in value when compared to the worst states of other preference-based measures. At 5% it is significantly lower than the worst states of SF-6D, AQL-5D, ICECAP and OPUS but not significantly different to the worst state of HUI2.

Figure 2: Scatter plot of estimated β_j versus current published state values.



4.2 Mapping functions

Estimates of the parameter values and their covariance matrix are used in this section to estimate mapping functions across preference-based measures using a common metric. The mapping functions in this section are very preliminary since they are based on the estimated rank ordered mixed logit model in section 4.1. There are still some restrictions to be imposed in the model, for example that the coefficients of EQ5D 11111 and OPUS 1111 are the same. Imposing restrictions in the rank ordered mixed logit model will probably change the mapping functions and therefore we need to exercise caution when interpreting the results in this section.

Visual inspection of scatter plots suggests that linear relationships are probably adequate for all preference-based measures apart from OPUS. For this preference-based measure the scatter plot indicates that a cubic relationship might fit the data better. This is confirmed by the smaller Akaike Information Criteria of the model (98.5314) compared to the model with only linear terms (119.6511). A more thorough specification search is needed once the final rank ordered mixed logit model is estimated since we have found the mappings to be somewhat sensitive to the specification of the model. There is also a big gap between the published health state values of the highest two EQ-5D states. EQ-5D 11111 has a published value of one whereas the next highest EQ-5D state included in our data has a value of 0.639. The lack of points in this range indicates that care should be taken when interpreting results in this range since the mapping function might not be very well defined. Table 5 shows the estimated coefficients and standard errors of the mapping functions. Using these mapping functions we can calculate how each preference-based measure in our study maps onto EQ-5D. Table 6 shows how the ranges of the preference-based measures map onto EQ-5D. A few issues deserve attention. The lowest published value of HUI2 maps onto a value of EQ-5D which is lower than the lowest published value for EQ-5D and the highest published value of OPUS maps onto 1.01. These two problems seem to be a reflection of the estimated coefficients of the ranked ordered mixed logit model. However, we have already pointed out that although the point estimates of the pairs EQ-5D 11111 and OPUS 1111 and EQ-5D 33333 and HUI2 455445 are different, they are not statistically significantly different and therefore imposing these two restrictions in the rank ordered logit model would tend to solve these two problems. Leaving these two issues aside, a published state value of 1 for ICECAP would map onto a value of EQ-5D higher than 1. This issue seems to stem from the specification of the mapping function. We shall investigate in the final model if a better specified mapping function for ICECAP exists.

5 Conclusions

In this paper we analysed rank data in which respondents are asked directly to rank states described using different preference-based measures. This means that the relationship between different instruments is determined directly by people's preferences for different hypothetical states. Ranked data is often analysed using a rank ordered logit model. However, we have demonstrated that two of the assumptions embedded in the rank ordered logit, the independence of irrelevant alternatives and independent repeated choices are untenable in our dataset. Since these restrictions are rejected by the data, the rank ordered logit model will give inconsistent estimates of the parameter values and inferences based on this model could be misleading.

We have developed a rank ordered mixed logit model tailored to our dataset which relaxes these two assumptions and we have shown that there are considerable differences in the estimated scale values of the preference-based measures. Estimation of the rank ordered mixed logit model provides consistent estimates of the values for all the states included in the dataset on a common metric so that direct comparisons across the preference-based measures included in the study are possible. In addition to this, we have used this common metric to map across preference-based measures. For this purpose, the estimated parameters of the rank ordered mixed logit model and their covariance matrix were used to estimate mapping functions jointly by maximum likelihood allowing for measurement error in the dependent variable and clustering in the error term around the preference-based measures.

The main aim of this paper is to discuss the econometric issues involved in analysing these data. However, the results presented do raise interesting issues in their own right that deserve discussion. The coefficients shown in Table 3 have been 'normalised' by setting OPUS 1111 equal to 1.00. This results in the 'best state' getting a value significantly above 1. What is not clear is whether this reflects a genuine strength of preference of the best state over all other states, which seems unlikely or whether there continues to be a problem in estimating values for a state that so clearly dominates over all other states. Apart from this anomaly, the numbers seem to broadly reflect the logical ordering of each state within preference-based measures. The numbers provide another means of comparing across preference-based measures using a common scale that can be used to provide a new means of mapping between them.

The main finding of this paper is that to be able to draw meaningful inferences in situations in which the assumptions of the rank ordered logit model are unlikely to hold, the use of a rank ordered mixed logit model is essential.

Table 1: Measures of health and quality of life.

Instrument	Summary (Unique states)	Dimensions	Levels
EQ-5D	Generic (243)	5 dimensions: Mobility, self-care, usual activity,pain/discomfort and anxiety/depression	3 levels: no problems to extreme problems
SF-6D	Generic (18,000)	6 dimensions: Physical functioning, role limitations,social functioning, pain, mental health, vitality	Between 4 and 6 levels in each dimension
HUI2	Generic for children (8,000)	7 dimensions: Sensation, mobility, emotion, cognition, self care, pain, fertility	Between 4 and 5 levels in each dimension
AQL-5D	Condition specific for asthma (3,125)	5 dimensions: Concern about asthma, shortness of breath, weather and pollution stimuli, sleep impact and activity limitations	5 levels: no problems to extreme problems
ICECAP	Capability measure for older people in IK (1,024)	5 dimensions: Attachment, security, role, enjoyment, control	4 levels: all, a lot, a little, none
OPUS	Social care outcome measure for older people (243)	5 dimensions: Food and nutrition, personal care, safety, social participation, control over daily living	3 levels: no unmet needs, low unmet needs, high unmet needs

Table 2: Number of times an instrument is ranked last excluding ties.

Health state	Frequency	Percentage
Dead	1,120	81.16
HUI2 455445	102	7.39
EQ-5D 33333	88	6.38
ICECAP 44444	22	1.59
OPUS 3333	20	1.45
SF-6D 645655	13	0.94
AQL-5D 55555	4	0.29
Other	11	0.80
Total	1,380	100

Table 3: Scaled parameter estimates.

		RO logit		RO mixed logit			
Health State		β_j	s.e.	β_j	s.e.	τ_j	s.e.
	σ_2			0.7218	(0.2910)[1]		
	σ_3			0.1020	(0.0721)[1]		
	σ_4			0.0115	(0.0351)[1]		
	σ_5			0.0330	(0.0146)[1]		
EQ-5D	11111	0.8320	0.0512	0.9442	0.0675	-0.1320	0.0406
	11322	0.7218	0.0441	0.7043	0.0356	-0.2860	0.0354
	12311	0.6493	0.0377	0.6984	0.0339	-0.2000	0.0313
	13211	0.6747	0.0356	0.6721	0.0329	-0.2713	0.0301
	21113	0.6314	0.0359	0.6419	0.0318	-0.2391	0.0290
	23121	0.6206	0.0335	0.6372	0.0324	-0.2647	0.0299
	11223	0.5976	0.0346	0.6189	0.0319	-0.2529	0.0296
	22212	0.5600	0.0310	0.6044	0.0304	-0.2217	0.0289
	21331	0.5602	0.0378	0.5999	0.0340	-0.2559	0.0334
	13132	0.4711	0.0305	0.5292	0.0323	-0.2911	0.0292
	12133	0.4507	0.0331	0.5248	0.0339	-0.2919	0.0320
	31112	0.4236	0.0332	0.5064	0.0335	-0.2599	0.0300
	31231	0.3649	0.0301	0.4559	0.0323	-0.2653	0.0322
	32121	0.3446	0.0305	0.4543	0.0340	-0.2952	0.0313
	33313	0.2786	0.0247	0.3958	0.0322	-0.3292	0.0298
	33333	0.1412	0.0121	0.2790	0.0287	-0.3458	0.0282
SF-6D	211111	0.8060	0.0425	0.8077	0.0381	-0.1360	0.0393
	112221	0.7495	0.0417	0.7796	0.0357	-0.1614	0.0326
	211211	0.7645	0.0412	0.7527	0.0344	-0.1625	0.0377
	111453	0.7005	0.0365	0.7003	0.0313	-0.2076	0.0343
	214411	0.6429	0.0365	0.6729	0.0315	-0.1847	0.0348
	424421	0.5926	0.0335	0.6176	0.0322	-0.2860	0.0289
	623133	0.5416	0.0323	0.5985	0.0313	-0.2228	0.0335
	545622	0.5337	0.0317	0.5895	0.0316	-0.2617	0.0280
	311655	0.5458	0.0302	0.5874	0.0308	-0.2654	0.0291
	624343	0.5153	0.0307	0.5691	0.0318	-0.2758	0.0284
	422655	0.4800	0.0292	0.5416	0.0321	-0.2981	0.0311
	535645	0.4744	0.0274	0.5379	0.0310	-0.2839	0.0294
	645655	0.3759	0.0191	0.4733	0.0293	-0.2973	0.0262
AQL-5D	21223	0.6915	0.0364	0.7029	0.0310	-0.1903	0.0292
	13321	0.6551	0.0361	0.6660	0.0317	-0.2032	0.0317
	12543	0.6428	0.0328	0.6627	0.0305	-0.1960	0.0289
	53411	0.6274	0.0358	0.6504	0.0326	-0.2107	0.0290
	32441	0.6145	0.0343	0.6353	0.0311	-0.2154	0.0286
	45143	0.5864	0.0334	0.6208	0.0314	-0.2140	0.0295
	23534	0.5761	0.0334	0.6150	0.0315	-0.2118	0.0294
	52314	0.5580	0.0296	0.6010	0.0305	-0.2330	0.0283
	34254	0.5315	0.0314	0.5792	0.0315	-0.2300	0.0284
	55424	0.5211	0.0291	0.5763	0.0301	-0.2144	0.0276
	15355	0.5143	0.0306	0.5664	0.0312	-0.2398	0.0280
	34554	0.5068	0.0300	0.5595	0.0311	-0.2249	0.0285
	55555	0.4174	0.0204	0.5012	0.0285	-0.2418	0.0247

Table 3 (cont): Scaled parameter estimates.

		RO logit		RO mixed logit			
Health State		β_j	s.e.	β_j	s.e.	τ_j	s.e.
HUI2	112222	0.6958	0.0392	0.6911	0.0331	-0.2628	0.0299
	121132	0.5986	0.0346	0.6383	0.0317	-0.2335	0.0319
	112123	0.5745	0.0326	0.6299	0.0313	-0.2104	0.0290
	323331	0.4887	0.0306	0.5515	0.0334	-0.3299	0.0338
	314431	0.4486	0.0310	0.5162	0.0337	-0.3047	0.0312
	234111	0.4290	0.0286	0.5143	0.0322	-0.2808	0.0286
	331131	0.4563	0.0279	0.5112	0.0314	-0.2962	0.0283
	344222	0.4208	0.0289	0.5078	0.0342	-0.3564	0.0346
	125425	0.3598	0.0275	0.4589	0.0327	-0.2991	0.0302
	133444	0.3569	0.0251	0.4454	0.0319	-0.3326	0.0295
	144325	0.3478	0.0269	0.4445	0.0332	-0.3275	0.0310
	445234	0.2266	0.0221	0.3480	0.0314	-0.3199	0.0296
	455445	0.0974	0.0109	0.2190	0.0287	-0.3681	0.0291
ICECAP	21131	0.8597	0.0551	0.8943	0.0611	-0.1519	0.0426
	31212	0.9124	0.0514	0.8508	0.0362	-0.1830	0.0331
	12321	0.8438	0.0474	0.8312	0.0393	-0.2162	0.0344
	23324	0.7034	0.0372	0.6967	0.0317	-0.2587	0.0306
	22242	0.6795	0.0380	0.6875	0.0339	-0.3101	0.0419
	14344	0.6233	0.0336	0.6402	0.0321	-0.3057	0.0300
	33333	0.6164	0.0339	0.6390	0.0324	-0.2625	0.0320
	43111	0.6046	0.0339	0.6360	0.0325	-0.2471	0.0305
	43443	0.5177	0.0312	0.5668	0.0315	-0.2523	0.0287
	44143	0.4584	0.0297	0.5221	0.0316	-0.2715	0.0282
	43334	0.4425	0.0288	0.5216	0.0318	-0.2834	0.0317
	42444	0.4401	0.0268	0.5070	0.0312	-0.3047	0.0279
	44444	0.3274	0.0175	0.4378	0.0292	-0.3047	0.0257
OPUS	1111	1.0000	-	1.0000	-	-0.1838	0.0385
	2121	0.7277	0.0404	0.7232	0.0322	-0.1865	0.0312
	3121	0.6368	0.0377	0.6820	0.0334	-0.1938	0.0302
	2212	0.5933	0.0325	0.6394	0.0318	-0.2127	0.0285
	2331	0.5613	0.0302	0.6064	0.0319	-0.2759	0.0283
	3132	0.5319	0.0333	0.5977	0.0326	-0.2252	0.0309
	1322	0.5407	0.0295	0.5871	0.0309	-0.2552	0.0291
	3221	0.5340	0.0291	0.5820	0.0306	-0.2341	0.0277
	2123	0.5340	0.0314	0.5812	0.0318	-0.2718	0.0282
	3313	0.5050	0.0311	0.5629	0.0315	-0.2457	0.0285
	1233	0.4937	0.0283	0.5605	0.0310	-0.2552	0.0280
	1333	0.4515	0.0286	0.5366	0.0322	-0.3012	0.0286
	3333	0.3419	0.0178	0.4530	0.0291	-0.2870	0.0249
Best	State	1.4901	0.0745	2.9137	0.8614	-0.0484	0.1417
Dead		0.0000	-	0.0000	-		

[1] Standard errors are provided for illustration purposes only.

Table 4: Correlations between utility differences between the PITS and 'dead'.

		EQ-5D 33333	SF-6D 645655	AQL-5D 55555	HUI2 455445	ICECAP 44444	OPUS 3333
EQ-5D	33333	1.0000					
SF-6D	645655	0.9256	1.0000				
AQL-5D	55555	0.9048	0.8987	1.0000			
HUI2	455445	0.9391	0.9282	0.9064	1.0000		
ICECAP	44444	0.9276	0.9187	0.8999	0.9303	1.0000	
OPUS	3333	0.9227	0.9145	0.8968	0.9251	0.9161	1.0000

Table 5: Estimated coefficients of the mapping functions.

Variable		**Estimate**	**s.e.**
EQ-5D	Constant	7.0194	0.4379
	Health State Value	4.6266	0.5789
SF-6D	Constant	5.1000	0.4693
	Health State Value	5.8098	0.4084
AQL-5D	Constant	5.2038	0.4739
	Health State Value	4.4494	0.3928
HUI2	Constant	4.1696	0.5090
	Health State Value	6.2582	0.6874
ICECAP	Constant	6.2375	0.4940
	Health State Value	5.6608	0.6105
OPUS	Constant	6.3096	0.3936
	Health State Value	12.0763	1.1315
	$(\text{Health State Value})^2$	-28.2135	3.5464
	$(\text{Health State Value})^3$	21.5223	2.8023
	θ_1	0.7271	0.1439
	θ_2	0.1879	0.0945
	θ_3	0.1560	0.0577
	θ_4	0.5389	0.1205
	θ_5	0.5393	0.1317
	θ_6	0.0000	0.1923

Table 6: Estimated EQ-5D mapping ranges.

	Published value range		**EQ-5D mapping range**	
EQ-5D	-0.594	to 1	-0.594	to 1
SF-6D	0.271	to 1	-0.075	to 0.841
AQL-5D	0.431	to 1	0.022	to 0.569
HUI2	-0.0552	to 1	-0.691	to 0.737
ICECAP	0	to 1	-0.169	to 1.055
OPUS	0	to 1	-0.153	to 1.011

References

Allison, P. D. and Christakis, N. A. 1994. Logit models for sets of ranked items. Sociological Methodology 24,199-228.

Alwin, D. F. and Krosnick, J. A. 1985. The measurement of values in surveys: A comparison of ratings and rankings. Public Opinion Quarterly 49, 535-552.

Andrews, D. W. K. 2001. Testing when a parameter is on the boundary of the maintained hypothesis. Econometrica 69, 683-734.

Barton, G. R., Bankart, J., Davis, A. C.and Summerfield, Q. A. 2004. Comparing utility scores before and after hearing aid provision: Results According to the EQ-5D, HUI3 and SF-6D. Applied Health Economics and Health Policy 3,103-105.

Beggs, S., Cardell, S. and Hausman, J. 1981. Assessing the potential demand for electric cars. Journal of Econometrics 16, 1-19.

Bhat, C. 2000. Incorporating observed and unobserved heterogeneity in urban work mode choice modelling. Transportation Science 34, 228-238.

Böckenholt, U. 1992. Thurstonian representation for partial ranking data. British Journal of Mathematical and Statistical Psychology 45, 31-39.

Boyd, J. H. and Mellman, R. E. 1980. The effect of fuel economy standards on the U.S. automotive market: an hedonic demand analysis. Transportation Research A 14A, 367-378.

Brady, H. E. 1989. Factor and ideal point analysis for interpersonally incomparable data. Psychometrika 54, 181-202.

Brady, H. E. 1990. Dimensional analysis of ranking data. American Journal of Political Science. Vol. 34, 1017-1048.

Brazier, J., Roberts, J., Tsuchiya, A. and Busschbach, J. 2004. A comparison of the EQ-5D and SF-6D across seven patient groups. Health Economics 13, 873-884.

Breslow, N. E. 1974. Covariance analysis of censored survival data. Biometrics 30, 89-100.

Cardell, N. S. and Dunbar, F. C. 1980 Measuring the societal impacts of automobile downsizing. Transportation Research A 14A, 423-434.

Chapman, R. G. and Staelin, R. 1982. Exploiting rank ordered choice set data within the stochastic utility model. Journal of Marketing Research 14, 288-301.

Chiou, L. and Walker, J. L. 2007. Masking identification of discrete choice models under simulation methods. Journal of Econometrics 141, 683-703.

Craig, B. M., Busschbach, J. J. V. and Salomon, J. A. 2008. Keep it simple: Ranking health states yields values similar to cardinal measurement approaches. Journal of clinical Epidemiology Vol ??,

Durbin, J. 1951. Incomplete blocks in ranking experiments. British Journal of Psychology 4, 85-90.

Efron, B. 1977. The efficiency of Cox's likelihood function for censored data. Journal of the American Statistical Association 72, 557-565.

Espallargues, M., Czoski-Murray, C., Bansback, N., Carlton, J., Lewis, G., Hughes, L., Brand, C. and Brazier, J. 2005. The impact of Age Related Macular Degeneration on Health Status Utility Values. Investigative Ophthalmology and Visual Science 46, 4016-4023.

Farewell, V. T. and Prentice, R. L. 1980. The approximation of partial likelihood with emphasis on case-control studies. Biometrika 67, 273-278.

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

Hausman, J. A. and Wise, D. A. 1978. A conditional probit model for qualitative choice: discrete decisions recognizing interdependence and heterogeneous preferences. Econometrica 46, 403 - 426.

Kalbfleisch, J. D. and Prentice, R. L. 2002. The statistical analysis of failure time data. (second edition) New York: Wiley.

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

Longworth, L. and Bryan, S. 2003. An empirical comparison of EQ-5D and SF-6D in liver transplant patients. Health economics 12,1061-1067.

Luce, R. D. 1959. Individual Choice Behaviour. New York: Wiley.

McFadden, D. 1974/3?. Conditional logit analysis of qualitative choice behavior In Frontiers in Econometrics, edited by P. Zarembka, 105-142. New York. Academic Press.

McFadden, D. and Train, K. 2000. Mixed MNL models for discrete response. Journal of Applied Econometrics 15, 447-470.

O'Brien, B. J., Spath, M., Blackhouse, G., Severens, J.L. and Brazier, J. E. 2003. A view from the Bridge: agreement between the SF-6D utility algorithm and the Health Utilities Index. Health Economics 12, 975-982.

Peto, R. 1972 Contribution to the discussion of a paper by D. R. Cox. Journal of the Royal Statistical Society, Series B 34, 205-207.

Plackett, R. L. 1975. The analysis of permutations. Journal of the Royal Statistical Society, 24, 193-202.

Rowen, D., Brazier, J., Tsuchiya, A. and Hernández, M. (2009). Simultaneous valuation of multiple instruments: a new approach to mapping. Health Economics and Decision Sciences Discussion Paper, University of Sheffield.

Skrondal, A. and Rabe-Hesketh, S. 2003. Multilevel logistic regression for polytomous data and rankings. Psychometrica 68, 267-287.

Train, K. E. Discrete choice methods with simulation. Cambridge University Press: Cambridge, UK.

Walker, J. L., Ben-Akiva, M. and Bolduc, D. 2007. Identification of parameters in normal error component logit-mixture (NECLM) models. Journal of Applied Econometrics 22, 1095-1125.