Article:
Hernández, M., Brazier, J., Rowen, D. et al. (2012) Common scale valuations across different preference-based measures: estimation using rank data. HEDS Discussion Paper 12/02.

HEDS Discussion Paper 12/02

Common scale valuations across different preference-based measures: estimation using rank data.

Mónica Hernández Alava, John Brazier, Donna Rowen and Aki Tsuchiya

Common scale valuations across different preference-based measures: estimation using rank data.*

Mónica Hernández Alava,[1] John Brazier[1], Donna Rowen[1] and Aki Tsuchiya[1,2]

[1]School of Health and Related Research, University of Sheffield
[2]Department of Economics, University of Sheffield

February 2012

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

ADDRESS FOR CORRESPONDENCE: Mónica Hernández Alava. Health Economics and Decision Science. School of Health and Related Research. University of Sheffield. Regent Court. 30 Regent Street Sheffield S1 4DA. E-mail: monica.hernandez@sheffield.ac.uk. Tel: +44 (0)114 2220736. Fax: +44 (0)114 2724095

* We are grateful to the Medical Research Council for financial support of this research under project grant R/112915. We would like to thank Professor Coast and Professor Netten for the use of ICECAP and OPUS respectively and all the interviewees who took part in the survey. We would also like to thank participants at the HESG in Sheffield for their helpful comments, in particular Arne Risa Hole and Allan Wailoo. All responsibility for the analysis and interpretation of the data presented here lies with the authors.

ABSTRACT:

Background: Different preference-based measures (PBMs) used to estimate Quality Adjusted Life Years (QALYs) provide different utility values for the same patient. Differences are expected since values have been obtained using different samples, valuation techniques and descriptive systems. Previous studies have estimated the relationship between pairs of PBMs using patient self-reported data. However, there is a need for an approach capable of generating values directly on a common scale for a range of PBMs using the same sample of general population respondents and valuation technique but keeping the advantages of the different descriptive systems.

Methods: General public survey data (n=501) where respondents ranked health states described using subsets of six PBMs were analysed. We develop a new model based on the mixed logit to overcome two key limitations of the standard rank ordered logit model, namely, the unrealistic choice pattern (Independence of Irrelevant Alternatives) and the independence of repeated observations.

Results: There are substantial differences in the estimated parameters between the two models (mean difference 0.07) leading to different orderings across the measures. Estimated values for the best states described by different PBMs are substantially and significantly different using the standard model, unlike our approach which yields more consistent results.

Limitations: Data come from a exploratory study that is relatively small both in sample size and coverage of health states.

Conclusions: This study develops a new, flexible econometric model specifically designed to reflect appropriately the features of rank data. Results support the view that the standard model is not appropriate in this setting and will yield very different and apparently inconsistent results. PBMs can be compared using a common scale by implementation of this new approach.

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 preference weights are estimated using various preference-based measures (PBMs) of health-related quality of life, generating a single index score which can be compared across different health care interventions or programmes. All PBMs are valued on an interval scale where full health is the upper anchor with an assigned value of 1 and 0 is usually assumed equivalent to dead with negative values indicating states worse than dead.

A number of different PBMs are available for use in economic evaluation but there is no common agreement on the use of a single measure for all patient groups, disease areas and interventions. The quality adjustment weights generated by different generic PBMs can differ substantially for the same patients [1],[2],[3],[4],[5]. Such differences are to be expected. The weights are obtained by different valuation techniques, different descriptive systems and using different samples of respondents (often from different countries). One solution to this problem has been to estimate a regression between two PBMs using patient self-reported data. This regression is usually called in the literature a mapping function. While this might be a useful short term pragmatic solution, it relies on a large degree of overlap between the descriptive systems of the PBMs [6]. For this reason, this paper explores a different approach that does not require overlap at the descriptive level.

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

In this paper we analyse primary data collected by interview in which members of the general public rather than patients are asked to rank hypothetical health states described using a subset of six different PBMs. The relationship between different measures is,

therefore, determined directly by people's preferences for different hypothetical states. Ranked data are often analysed using a rank ordered logit model [7]. However, there are two limitations of this standard model that make it unattractive in our case: the unrealistic choice pattern (Independence of Irrelevant Alternatives) and the inability to accommodate repeated choices unless they are assumed to be independent. To overcome these limitations of the standard model, we develop a new model based on the mixed logit [8], [9] which we call a rank ordered mixed logit model.

The main aim of this paper is to estimate values for all the health states included in the study on a common scale so that direct comparisons across different descriptive systems are possible. This common metric could also be used to convert weights on their original scales across PBMs.

2 Methods

2.1 Data

We analyse primary data from a pilot study collected by interview in the North of England during June to October 2007. Full details of the data collection process, the study design and characteristics of the respondents can be found in [10] and additional analysis in [11]. Here we present a brief summary of the data.

The study involves six PBMs of health and quality of life: EQ-5D (generic, [12]), SF-6D (generic, [13]), HUI2 (generic for children and adults, [14]), AQL-5D (asthma specific, [15], [16]), OPUS (social care specific, [17]), ICECAP (capabilities, [18]). These PBMs reflect a range of different types of measures and are summarised in Table 1. Each respondent was asked to perform three ranking tasks. In each task, respondents were shown eight cards with descriptions of eight states and were asked to rank them. The survey design contains twenty variations of the ranking tasks. Ties in the ranking are possible if states are considered equal by a respondent. Each interview involves hypothetical states from three of the six PBMs indicated above. The eight states in each task always include two generic states 'best state' and 'dead'. The remaining six states comprise three from each of two different PBMs. For each of the two PBMs, the three states always consisted of

the worst state, one mild and one moderate state. Table 2 shows an example of a set of eight cards seen by a respondent in one ranking task. This particular example set of cards includes OPUS and EQ-5D.

Although the worst states for all the PBMs are included in the study, there are only two best states included: EQ-5D (health) and OPUS (social care). Their value is 1 since they are used as the top anchors in their respective preference-based metrics**.** One of the issues that we aim to determine in this study is how these two states compare according to people's preferences.

The dataset contains data for 501 individuals. Two respondents were dropped from the sample since they provided no ranking 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 state') in a task and another respondent only ranked two cards ('best state' and 'dead') in a task and, therefore, these two tasks were dropped from the sample. There were also seven instances of a state not being ranked in a task. In these cases we excluded the state description from the choice set of the respondent and kept the remaining observations in the task.

There are several features of the data which are as expected but will nevertheless become important in the model specification. The generic 'best state' is ranked first 99.73% of the time: it is ranked first on its own 98.38% of the time and the remaining 1.35% of the time it is ranked first but tied with another state. On only four occasions 0.27% of the time) is another state ranked higher than 'best state', three of which are observations for the same individual in different ranking tasks. Of the 20 times the generic 'best state' is tied with other states, 12 times it is tied with EQ-5D 11111 and 4 times it is tied with OPUS 1111.

The bottom of the ranking presents more variation than the top. Table 3 shows the frequencies with which states are ranked last. This table only includes the cases with no ties at the end of the ranking which amounts to 93.00% of the bottom rankings. The highest percentage corresponds to 'dead' being ranked last. Including ties, 'dead' is ranked last with a frequency of 80.55% in the full sample. There are 11 states under the heading 'Other' 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 PBMs.

2.2 Model specification

2.2.1 Background and limitations of the standard model

Rank ordered data are usually analysed using a logit model [19], [20]. A brief explanation of the main features of this model is given below to describe its limitations and compare it with our new model developed for this task (see appendix for a more technical description).

Individual i faces J different alternative states in each of the T choice situations. The sets of states each individual faces are different by study design and the number of choice situations also differs across individuals if a full set of rankings is not completed. Therefore, we should use J_{it} and T_i but for simplicity we just 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}, typically assumed to be a linear function of some fixed parameters β and an unknown stochastic part, ε_{ijt} assumed independent and identically distributed (IID) type I extreme value.

$$U_{ijt} = \nu_{ijt} + \varepsilon_{ijt} \qquad i = 1, 2, \ldots, n; j = 1, 2, \ldots, J; t = 1, 2, \ldots, T \tag{1}$$

In each choice situation, the individual chooses the alternative with the highest utility. When using rank data, a ranking of J alternatives is expressed as $J - 1$ successive and independent choices by the individual. The alternative ranked first is chosen from the full set of alternatives. Then, the alternative ranked second is chosen from the remaining alternatives. This continues until all alternatives are exhausted. For this reason this model is often called the exploded logit [21].

One complication is that sometimes two or more alternatives are given the same rank. Allison and Christakis [22] proposed a generalisation of the likelihood of the logit model for tied alternatives by assuming that each individual has a preferred order of the alternatives but this is not observed. Thus, it is assumed in the estimation that any possible permutation of the tied alternatives would be possible.

This method is computationally demanding and approximations have been suggested

in the literature [23], [24]. However, these approximations have been shown to be inaccurate [25] especially in cases with a large number of ties. 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 approximations might not perform well and we explore the sensitivity of the results to approximations in the results section.

Two limitations of this standard model are particularly relevant here. The first is the property of Independence of Irrelevant Alternatives (IIA). This 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 some situations, this property might represent behaviour correctly but in cases where some alternatives are either very similar or have a natural order, this property will be too restrictive. In our case, respondents are ranking across different states from different PBMs and some of the states across these PBMs could be viewed as very similar, for example EQ-5D 11111 and OPUS 1111 or some of the worst states across the PBMs. In addition, within each descriptive system, the alternatives have a logically determined order, for example EQ-5D 11112 is better than EQ-5D 12233.

A second limitation of the standard model relates to the way repeated choices are handled. In the present case, each individual performs three different ranking tasks and in addition each ranking task is exploded into a number of successive choices. 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 clustering by respondents but the first limitation would remain. Misspecification of the model will lead to inconsistent estimates making inferences unreliable.

2.2.2 A general rank ordered mixed logit model

We relax these two limitations by developing a new model for rank ordered data based on the mixed logit model [8], [9]. To our knowledge, this is the first application of a mixed logit model to rank data. The mixed logit model is very flexible and has been shown to be able to approximate any random utility model [26]. Different derivations of the

mixed logit model exist [27] but its main difference with the logit model is the inclusion of additional stochastic terms, ξ_{ijt}^s, in the utility in equation (1) to give

$$\begin{aligned} U_{ijt} &= \nu_{ijt} + \xi_{ijt}^1 + \ldots + \xi_{ijt}^S + \varepsilon_{ijt} \\ &= \nu_{ijt} + \boldsymbol{\xi}_{ijt} + \varepsilon_{ijt} \end{aligned}$$

The new error component, $\boldsymbol{\xi}_{ijt}$, can be correlated between alternatives and choice situations and can be heteroskedastic. This is the key to the flexibility of the model but also responsible for its higher computational burden. This additional random term is assumed to have zero mean and a distribution $f(\boldsymbol{\xi}|\Psi)$ where Ψ is a vector of fixed parameters defining the distribution which are estimated jointly with the other parameters in the model.

The mixed logit model, like the logit model, cannot handle ties in the alternatives in this form but it can be generalised using Allison and Christakis [22] method (see appendix). This new model requires bespoke programming which we undertook in GAUSS 9.0 [28].

2.2.3 Application of the model.

In this particular application we do not have any explanatory variables which vary across alternatives; we only use alternative specific constants. Although each respondent only sees a maximum of 17 different alternatives, the number of total alternatives across all respondents is 83. It is not possible to include a different error component for each alternative and estimate a full covariance matrix which would be required to completely eliminate the IIA property and allow full flexibility. Instead, we try to allow for a flexible enough structure in the covariance matrix of the utilities that can approximate adequately respondents' behaviour so that we obtain consistent estimates while keeping a parsimonious specification.

Given the discussion in the data section, the rank ordered mixed logit specification here adds six independent standard normal error components to the utility in equation (1); five define a nested structure and the last one represents an individual latent factor

as described below.

The structure for the top states is very general to allow us to directly test hypotheses about their equality. A common structure is also allowed for the alternatives 'dead' and the worst states. When individuals are given the eight alternatives to rank, the cards 'dead' and 'best state' are very prominent (see Table 2). In addition to this, similar wording clearly separates the alternatives into two groups of three alternatives since they come from two different PBMs. Within each group of three states, one clearly stands out as being the worst and in fact always corresponds to the worst state for that PBM. Given that it is easy to identify 'dead' and the two worst states in each set of cards as the most undesirable ones, it might be appropriate to allow for the correlation structure of utilities to be similar within this group[1]. Therefore, the initial model includes five separate error components ($\xi_i^s,\ s = 1, \ldots, 5$) in this nested structure: three different error components for the three alternatives at the top ('best state', EQ-5D 11111 and OPUS 1111), a fourth error component for a nest encompassing the alternative 'dead' and the worst states of all six PBMs and a fifth error component for a final nest covering all the remaining alternatives. The final nest might not be significant given the range of states included in it but it will be damaging for the consistency of the estimates to ignore it if it exists [29]. The five nests are allowed to have different variances, ϖ_s^2. The nested structure induces two different types of correlations. First, the utilities of alternatives in the same nest are now allowed to be correlated and differ by nest. Second, the utilities of the same alternative across different choice situations for the same individual are perfectly correlated. Therefore, this nested structure goes some way in relaxing the restrictive properties of the logit model. However, it is still rather limited. The utilities of alternatives in a specific nest are assumed to have exactly the same correlation and the utilities of alternatives in different nests are still uncorrelated. Furthermore, it only partially addresses the issue of repeated choices. To be able to relax these assumptions, we add to this nested structure another standard normal error component, an individual latent factor, ξ_i^6. 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,

[1] A variation excluding 'dead' from this group was also attempted (see Results).

τ_j, so that its impact differs by alternative. The utility in equation (1) augmented with the error components structure becomes

$$U_{ijt} = \begin{cases} \beta_j + \varpi_{1i}\xi_{1i} + \tau_j\xi_i^6 + \varepsilon_{ijt} & \text{if } j = \text{ 'Best State'} \\ \beta_j + \varpi_{2i}\xi_{2i} + \tau_j\xi_i^6 + \varepsilon_{ijt} & \text{if } j = \text{ 'EQ-5D 11111'} \\ \beta_j + \varpi_{3i}\xi_{3i} + \tau_j\xi_i^6 + \varepsilon_{ijt} & \text{if } j = \text{ 'OPUS 1111'} \\ \beta_j + \varpi_{4i}\xi_{4i} + \tau_j\xi_i^6 + \varepsilon_{ijt} & \text{if } j = \text{ 'Dead' or any of the worst states} \\ \beta_j + \varpi_{5i}\xi_{5i} + \tau_j\xi_i^6 + \varepsilon_{ijt} & \text{otherwise} \end{cases} \quad (2)$$

Since the individual latent factor enters the utilities of all alternatives but with different factor loadings, it allows for different correlations across the utilities of all the alternatives. These correlations will depend on both the factor loadings and the variance of the nests the utilities belong to, that is, both respondent specific and alternative specific characteristics. For instance, the utilities of the alternatives in the nest containing the alternative 'dead' and the worst states of all six PBMs will no longer share the same correlation structure unless all the τ_j of all the states in the nest are equal. This model allows for a very rich pattern of respondent behaviour relaxing the restrictive properties of the logit model.

A number of alternative specifications to the initial model in equation (2) were estimated and those are discussed in the results section.

2.3 Relationships between current weights and common scale weights.

After estimation of the parameters of the model using a rank ordered mixed logit model, we estimate the relationship between the published sets of values for each of the PBMs used (EQ-5D [12], SF-6D [30], HUI2 [31], AQL-5D [15], OPUS [17], ICECAP [18]) and the health state values of the PBMs in our common metric. Due to the relatively small number of states for each PBM included in this study, the estimated relationships can only support very simple linear functional forms and are probably not of sufficient accuracy. The following relationship is postulated:

$$\beta = W\alpha + \eta \quad (3)$$

where β is the $((J-1)\times 1)$ vector of true values measured in a common scale for the states included in the study, α is a $k \times 1$ vector $(k < J)$ of parameters of interest to be estimated and W is a block diagonal data matrix. Each block W_1 to W_6 relates to one of the six PBMs and includes as a variable the currently used values of the states included in this study. The η's are IID normally distributed, mean zero and heteroskedastic with block diagonal covariance matrix

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

The true parameter vector β is not known; instead we have an estimate $\hat{\beta}$ so that $\hat{\beta} = \beta + \zeta$. Substituting this expression in (3), we get the following regression

$$\begin{aligned} \hat{\beta} &= W\alpha + \eta + \zeta \\ &= W\alpha + \omega \end{aligned}$$

where ω is a composite error term. It is assumed that η and ζ are independent and normally distributed. We use maximum likelihood to estimate α together with the six θ's using $\Omega + Cov\left(\hat{\beta}\right)$ as an estimate of the covariance matrix of ω, where $Cov\left(\hat{\beta}\right)$ is the inverse of the Hessian matrix of the rank ordered mixed logit model.

3 Results

First we checked the accuracy of an approximation (Efron's [24]) to handle ties in the standard rank order logit model. The use of Efron's approximation generates differences in the estimated values of the states of up to 0.08 and a difference of 0.12 in the estimated value of the 'best state'. These differences are large relative to the range of estimated values and although the computational cost is great both in terms of the additional programming complexity and increased estimation time we do not use any approximations.

A number of alternative specifications to the initial model in equation (2) were estimated[2] and a restricted version was selected according to the Bayesian Information Criterion (BIC). The restricted model had a common error component and equal alternative specific constants for 'EQ-5D 11111' and 'OPUS 1111'. Table 4 presents the estimated parameters of the rank order logit model and the preferred specification (selected by BIC) of the rank ordered mixed logit model rescaled so that the difference in expected utility between the most preferred state from a PBM (OPUS 1111) and 'dead' equals one. The estimated parameters seem to broadly reflect the logical ordering of each state within each PBM. Based on the significance of the estimated factor loadings, τ_j, it is clear that IIA will be rejected in support of the rank ordered mixed logit model and a likelihood ratio test between the two models emphatically rejects the rank order logit with a test statistic of 1129.37 and zero p-value. Unfortunately, some of the parameters 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 [32] and therefore would not change the present conclusion.

Although the standard errors seem to indicate 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, again, testing is 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 an individual latent factor (ξ_i^6) and compared it with the full model. The likelihood ratio test rejects the model with only the latent factor with a χ^2 test statistic of 38.12 and p-value of zero. Thus, the 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 values of health states between the standard model and our model. Excluding the greatest difference in parameter estimates which corresponds to the generic 'best state', the largest difference is 0.17 and the mean

[2]These included amongst others a model with only a latent factor ξ_i^6 and no nested structure and a model with a nest for 'EQ-5D 11111' and 'OPUS 1111'. We also attempted to estimate alternative specifications where the alternative 'dead' had a separate error component but it seems that the data cannot empirically sustain these models.

and the median difference are 0.07 and 0.08 respectively. Surprisingly, the point estimates for the top states described by EQ-5D and OPUS are substantially different when estimated using the standard model (0.83 and 1 respectively). Thus, the absence of social care problems described by OPUS1111 is preferred to the absence of health problems described by EQ-5D1111. The restriction of equality is rejected with a χ^2 statistic of 8.85 and a p-value of zero. In contrast, when the more flexible rank ordered mixed logit model is estimated, the specification that fits best according to BIC is one where both parameters are equal. That is, the unexpected large difference disappears once the more flexible rank order mixed logit model is used. Thus, the restrictive assumptions of the rank order logit appear to be affecting the estimates in such a way as to cause the expected utilities of the top two states described by EQ-5D and OPUS to appear significantly different.

Unfortunately, due to the study design, there are few instances of logically determined orderings between the states included in each PBMs beyond the best states being ranked at the top and the worst states at the bottom, therefore we cannot compare the models performance in this matter.

There are further features of our model results that are worth noting. First, the factor loadings of the latent factor are all 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 the 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 a 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 5 shows the implied correlations of the utility differences between the worst states described by each of the PBMs and 'dead'. All these correlations are very high and clearly different from 0.5; a respondent who prefers the worst state on one measure to 'dead' is more likely to prefer all the rest of the worst states to 'dead'. This pattern of correlations is not unique

to the worst states; the mean and median correlations between all utility differences in our model are 0.86 and 0.89 respectively, a clear departure from 0.5 for a large number of utility differences.

In addition, in both models the estimated parameter value corresponding to the generic 'best state' is significantly above that of the top two states defined by EQ-5D and OPUS. This difference is larger for the rank order mixed logit model. The alternative 'best state' was clearly viewed as different and preferred to the top two states in the raw data and these coefficients reflect this fact. Given that the 'best state' dominates so clearly, the assumption of random utility for this alternative may not be adequate in this case but nevertheless since 'best state' is in a nest of its own and not used to set the scale of the model it is unlikely to have a noticeable influence on the rest of the estimated parameters.

Figure 1 plots the current published state values against the estimated values on a common metric for all states included in the study. The published state value of the worst state of EQ-5D is -0.594 [12] and that of the worst state of HUI2 is -0.0552 [31]. This apparently large difference is not observed in the estimated values in the common metric. In fact, on average, the worst HUI2 state seems to be regarded as worse than the worst EQ-5D state by the respondents of this study with coefficients of 0.29 (0.03) and 0.23 (0.03) respectively (standard errors in brackets).

Another important point to note is the large difference in the published state values between the top two EQ-5D states included in this study (see Figure 1), this will become an important issue in the following section.

3.1 Relationship between current weights and common scale weights.

Estimates of the health state values and their covariance matrix are used in this section to provide an illustrative example of the estimated relationship between the current published state values and those values estimated in a common scale using the rank ordered mixed logit model.

Visual inspection of scatter plots suggests that linear relationships are probably adequate, given the caveats about the small number of observations raised earlier, for all

PBMs apart from OPUS. For this PBM the scatter plot indicates that a cubic relationship might fit the data better. This is confirmed by the smaller BIC of the cubic model (158.5) compared to the model with only linear terms (175.1). Table 6 shows the estimated coefficients and standard errors for these regressions on the original scale. Table 7 provides the implied scaled relationship between health state values on their original and the common scales.

These regressions allow us to calculate how each PBM in our study might map onto any other. Table 8 shows how the ranges of the PBMs relate to EQ-5D. These estimates need to be treated with caution given the large gap evident in Figure 1 between the highest two or three EQ-5D states included in the study. A few issues are of particular note. The lowest published value of HUI2 (-0.0552) maps onto a value of EQ-5D which is substantially lower (-0.689). Indeed, this is even lower than the lowest published value for EQ-5D (-0.594). This suggests that the difference between the worst states described by EQ-5D and HUI2 is not as large as the current published values suggest. Given that these two states were included in the study, this issue was also evident from the estimated parameters of the rank ordered mixed logit model (see Figure 1). The point estimates of these two states are followed by the worst states of ICECAP and OPUS (with similar values), then followed by the worst state described by SF-6D and finally by the worst AQL-5D estate. This ascending ordering of the worst states does indeed appear reasonable. However, obtaining values for SF-6D and HUI2 upper states that are significantly lower than one is surprising, although this might be a consequence of the study design and the large gap between the EQ-5D states included in the study which makes the specification of the function unreliable. This is a useful point arising from this feasibility study which will need addressing in future work in this area.

4 Discussion and conclusions

Many PBMs coexist since there is no agreement on the use of a single measure for all patient groups, disease areas and interventions. Different PBMs generate different quality adjustment weights even for the same patients. Therefore, comparisons across studies

which assume that the QALY calculation is unaffected by the PBM used would be inaccurate.

This paper contributes to the literature in this area with two important developments. First, the paper opens a new avenue for research by proposing an alternative way of comparing across PBMs using a common scale. Second, the paper develops a new, flexible econometric model, the rank ordered mixed logit, to be able to analyse the type of data (rankings) required to compare across PBMs in this alternative approach. Furthermore, this new econometric model can be effectively used to analyse rank data in many other situations where the assumptions required by the rank ordered logit model are rejected by the data.

This method allows the relationship between PBMs to be determined directly by people's preferences for different hypothetical states. This provides a method of comparing across PBMs using a common scale that could be used to provide a new means of mapping between them. Rank data is commonly analysed using a rank ordered logit model. This model is straightforward to estimate but assumes IIA and independent repeated choices which, very often, are too strong and rejected by the data. In these cases, the rank ordered logit model will give inconsistent estimates of the parameter values and inferences based on this model could be misleading. This paper develops a very flexible model, the rank ordered mixed logit model. The general model has been tailored to our dataset reflecting the specific characteristics of the study but it can be easily adapted for estimation of other rank datasets with a different error components structure. We have shown that there are considerable differences in the estimated parameter values corresponding to the states in the six PBMs between these two models. The estimated parameters of the rank ordered mixed logit model seem to reflect the logical ordering of each state within each PBM and provide a direct comparison across the states included in the study.

In addition to this direct comparison, we have also presented how this common metric might be used to convert different quality adjusted weights across PBMs. Due to the small number of states for each PBM included in this study, the estimated relationships can only support very simple linear functional forms and are quite possibly not of sufficient accuracy. Therefore, these relationships should only be regarded as an illustrative example

of how the issue could be tackled. The conversion takes into account the estimated nature of the common metric and allows for clustering in the error term around the PBMs. A clear future research development would be to conduct a larger, definitive study allowing the results of applying the methods described here to be considered sufficiently robust for reliably informing decision making.

References

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

[2] O'Brien BJ, Spath M, Blackhouse G, Severens JL, Brazier JE. A view from the Bridge: agreement between the SF-6D utility algorithm and the Health Utilities Index. Health Economics. 2003;12:975-982.

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

[4] Barton GR, Bankart J, Davis AC, Summerfield QA. 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. 2004;3:103-105.

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

[6] Brazier JE, Yang Y, Tsuchiya A, Rowen DL. A review of studies mapping (or cross walking) non-preference based measures of health to generic preference-based measures. The European Journal of Health Economics. 2010; 11: 215-225

[7] McCabe C, Brazier J, Gilks P, Tsuchiya A, Roberts J, O'Hagan A, et al. Using rank data to estimate health state utility models. Journal of Health Economics. 2006;15:418-431.

[8] Boyd JH, Mellman RE. The effect of fuel economy standards on the U.S. automotive market: an hedonic demand analysis. Transportation Research A. 1980;14A:367-378.

[9] Cardell NS, Dunbar FC. Measuring the societal impacts of automobile downsizing. Transportation Research A. 1980;14A:423-434.

[10] Rowen D, Brazier J, Tsuchiya A, Hernández M, and Ibbotson R. The simultaneous valuation of states from multiple instruments using ranking and VAS data: methods and preliminary results. Health Economics and Decision Sciences Discussion Paper Series 09/06, University of Sheffield. 2009.

[11] Rowen D, Brazier J, Tsuchiya A, Hernández Alava M. Valuing states from multiple measures on the same VAS: A feasibility study. Health Economics. 2012; [in press].

[12] Dolan P. Modeling Valuations for EuroQol Health States. Medical Care. 1997;35:1095-1108.

[13] Brazier J, Roberts J, Deverill M. The estimation of a preference-based measure of health from the SF-36. Journal of Health Economics. 2002;21:271-292.

[14] Torrance GW, Feeny DH, Furlong WJ, Barr RD, Zhang Y, Wang Q. A multi-attribute utility function for a comprehensive health status classification system: Health Utilities Mark 2. Medical Care. 1996;34:702-722.

[15] Yang Y, Brazier J, Tsuchiya A, Young T. Estimating a Preference-Based Index for a 5-Dimensional Health State Classification for Asthma Derived From the Asthma Quality of Life Questionnaire. Medical Decision Making. 2011;31(2):281-91.

[16] Young T, Yang Y, Brazier J, Tsuchiya A. The Use of Rasch Analysis in Reducing a Large Condition-Specific Instrument for Preference Valuation: The Case of Moving from AQLQ to AQL-5D. Medical Decision Making 2011;31(1):195-210.

[17] Ryan M, Netten A, Skatun D, Smith P. Using discrete choice experiments to estimate a preference-based measure of outcome - An application to social care for older people. Journal of Health Economics. 2006;25:927-944.

[18] Coast J, Flynn T, Natarajan L, Sprotson K, Lewis J, Louviere JL et al. Valuing the ICECAP capability index for older people. Social Science and Medicine. 2008;67:874-882.

[19] Luce RD. Individual Choice Behaviour. New York: Wiley; 1959.

[20] Plackett RL. The analysis of permutations. Journal of the Royal Statistical Society. 1975; 24:193-202.

[21] Chapman RG, Staelin R. Exploiting rank ordered choice set data within the stochastic utility model. Journal of Marketing Research. 1982;14:288-301.

[22] Allison PD, Christakis NA. Logit models for sets of ranked items. Sociological Methodology. 1994;24:199-228.

[23] Breslow NE. Covariance analysis of censored survival data. Biometrics. 1974;30:89-100.

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

[25] Farewell VT, Prentice RL. The approximation of partial likelihood with emphasis on case-control studies. Biometrika. 1980;67:273-278.

[26] McFadden D, Train K. Mixed MNL models for discrete response. Journal of Applied Econometrics. 2000;15:447-470.

[27] Train KE. Discrete choice methods with simulation. Cambridge University Press: Cambridge, UK; 2003.

[28] GAUSS Mathematical and Statistical System 9.0. Aptech Systems, Inc.

[29] Walker JL, Ben-Akiva M, Bolduc D. Identification of parameters in normal error component logit-mixture (NECLM) models. Journal of Applied Econometrics. 2007;22:1095-1125.

[30] Brazier J, Roberts J. The estimation of a preference-based measure of health from the SF-12. Medical Care 2004;42:851–859.

[31] McCabe C, Stevens K, Roberts J, Brazier J. Health state values for the HUI 2 descriptive system: results from a UK survey. Health Economics 2005;14: 231–244.

[32] Andrews DWK. Testing when a parameter is on the boundary of the maintained hypothesis. Econometrica. 2001;69:683-734.

[33] Chiou L, Walker JL. Masking identification of discrete choice models under simulation methods. Journal of Econometrics. 2007;141:683-703.

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: Example set of eight cards seen by a respondent in one ranking task.

I have an inadequate diet potentially resulting in a health risk I am often dirty with poor personal hygiene I am socially isolated with little or no contact from others I have no control over daily living
I do not always get appropriate food but there is little health risk I am often dirty with poor personal hygiene I am socially isolated with little or no contact from others I have as much control over daily living as possible
I have sufficient, varied timely meals I am always clean and appropriately dressed I see people as often as I would like I have as much control over daily living as possible
I have no problems in walking about I am unable to wash or dress myself I have some problems with performing my usual activities I have no pain or discomfort I am not anxious or depressed
I am confined to bed I am unable to wash or dress myself I am unable to perform my usual activities I have extreme pain or discomfort I am extremely anxious or depressed
I have some problems in walking about I am unable to wash or dress myself I have no problems with performing my usual activities I have moderate pain or discomfort I am not anxious or depressed
Dead
Best state

Table 3: Number of times a state 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 4: Scaled parameter estimates.

		RO logit		RO mixed logit			
Health State		β_j	s.e.	β_j	s.e.	τ_j	s.e.
ϖ	Best state			0.7640	0.3158*		
ϖ	Top states			0.0854	0.0630*		
ϖ	Middle states			0.0066	0.0171*		
ϖ	Bottom states, dead			0.0384	0.0072*		
EQ-5D	11111	0.8320	0.0512	1.0000	-	-0.1240	0.0423
	11322	0.7218	0.0441	0.7415	0.0488	-0.3042	0.0379
	12311	0.6493	0.0377	0.7346	0.0483	-0.2116	0.0323
	13211	0.6747	0.0356	0.7059	0.0449	-0.2900	0.0323
	21113	0.6314	0.0359	0.6752	0.0436	-0.2538	0.0304
	23121	0.6206	0.0335	0.6695	0.0431	-0.2836	0.0320
	11223	0.5976	0.0346	0.6512	0.0425	-0.2688	0.0312
	22212	0.5600	0.0310	0.6358	0.0405	-0.2381	0.0306
	21331	0.5602	0.0378	0.6321	0.0433	-0.2720	0.0353
	13132	0.4711	0.0305	0.5567	0.0392	-0.3097	0.0320
	12133	0.4507	0.0331	0.5521	0.0402	-0.3116	0.0343
	31112	0.4236	0.0332	0.5329	0.0392	-0.2781	0.0321
	31231	0.3649	0.0301	0.4803	0.0369	-0.2841	0.0344
	32121	0.3446	0.0305	0.4774	0.0382	-0.3159	0.0341
	33313	0.2786	0.0247	0.4168	0.0348	-0.3514	0.0329
	33333	0.1412	0.0121	0.2937	0.0293	-0.3705	0.0329
SF-6D	211111	0.8060	0.0425	0.8491	0.0562	-0.1450	0.0411
	112221	0.7495	0.0417	0.8202	0.0534	-0.1728	0.0345
	211211	0.7645	0.0412	0.7892	0.0507	-0.1766	0.0394
	111453	0.7005	0.0365	0.7364	0.0458	-0.2213	0.0357
	214411	0.6429	0.0365	0.7081	0.0452	-0.1940	0.0360
	424421	0.5926	0.0335	0.6499	0.0422	-0.3075	0.0323
	623133	0.5416	0.0323	0.6299	0.0414	-0.2352	0.0351
	545622	0.5337	0.0317	0.6204	0.0407	-0.2813	0.0306
	311655	0.5458	0.0302	0.6180	0.0402	-0.2823	0.0310
	624343	0.5153	0.0307	0.5991	0.0398	-0.2957	0.0312
	422655	0.4800	0.0292	0.5696	0.0391	-0.3177	0.0343
	535645	0.4744	0.0274	0.5661	0.0382	-0.3024	0.0319
	645655	0.3759	0.0191	0.4980	0.0343	-0.3168	0.0292
AQL-5D	21223	0.6915	0.0364	0.7393	0.0459	-0.2041	0.0299
	13321	0.6551	0.0361	0.6995	0.0444	-0.2170	0.0314
	12543	0.6428	0.0328	0.6965	0.0429	-0.2110	0.0299
	53411	0.6274	0.0358	0.6839	0.0435	-0.2265	0.0301
	32441	0.6145	0.0343	0.6687	0.0427	-0.2292	0.0300
	45143	0.5864	0.0334	0.6530	0.0420	-0.2287	0.0312
	23534	0.5761	0.0334	0.6471	0.0418	-0.2265	0.0309
	52314	0.5580	0.0296	0.6325	0.0402	-0.2505	0.0299
	34254	0.5315	0.0314	0.6095	0.0398	-0.2481	0.0291
	55424	0.5211	0.0291	0.6066	0.0392	-0.2291	0.0287
	15355	0.5143	0.0306	0.5963	0.0396	-0.2552	0.0298
	34554	0.5068	0.0300	0.5880	0.0390	-0.2398	0.0292
	55555	0.4174	0.0204	0.5274	0.0346	-0.2583	0.0263

Table 4 (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	0.7282	0.0464	-0.2801	0.0322
	121132	0.5986	0.0346	0.6715	0.0432	-0.2460	0.0334
	112123	0.5745	0.0326	0.6628	0.0416	-0.2263	0.0306
	323331	0.4887	0.0306	0.5805	0.0407	-0.3505	0.0378
	314431	0.4486	0.0310	0.5445	0.0395	-0.3235	0.0339
	234111	0.4290	0.0286	0.5414	0.0383	-0.2984	0.0312
	331131	0.4563	0.0279	0.5385	0.0374	-0.3170	0.0310
	344222	0.4208	0.0289	0.5349	0.0397	-0.3783	0.0392
	125425	0.3598	0.0275	0.4831	0.0369	-0.3163	0.0330
	133444	0.3569	0.0251	0.4693	0.0358	-0.3572	0.0335
	144325	0.3478	0.0269	0.4679	0.0369	-0.3474	0.0340
	445234	0.2266	0.0221	0.3670	0.0332	-0.3401	0.0326
	455445	0.0974	0.0109	0.2311	0.0288	-0.3912	0.0332
ICECAP	21131	0.8597	0.0551	0.9409	0.0793	-0.1595	0.0449
	31212	0.9124	0.0514	0.8939	0.0571	-0.1946	0.0345
	12321	0.8438	0.0474	0.8781	0.0509	-0.2312	0.0363
	23324	0.7034	0.0372	0.7329	0.0457	-0.2779	0.0323
	22242	0.6795	0.0380	0.7238	0.0473	-0.3334	0.0447
	14344	0.6233	0.0336	0.6735	0.0435	-0.3276	0.0329
	33333	0.6164	0.0339	0.6716	0.0433	-0.2778	0.0324
	43111	0.6046	0.0339	0.6685	0.0434	-0.2629	0.0308
	43443	0.5177	0.0312	0.5969	0.0399	-0.2688	0.0303
	43334	0.4425	0.0288	0.5499	0.0383	-0.3015	0.0327
	44143	0.4584	0.0297	0.5493	0.0383	-0.2892	0.0302
	42444	0.4401	0.0268	0.5341	0.0370	-0.3253	0.0300
	44444	0.3274	0.0175	0.4607	0.0330	-0.3239	0.0284
OPUS	1111	1.0000	-	1.0000	-	-0.2089	0.0372
	2121	0.7277	0.0404	0.7606	0.0481	-0.1975	0.0322
	3121	0.6368	0.0377	0.7172	0.0457	-0.2047	0.0314
	2212	0.5933	0.0325	0.6721	0.0432	-0.2270	0.0301
	2331	0.5613	0.0302	0.6378	0.0416	-0.2941	0.0298
	3132	0.5319	0.0333	0.6287	0.0416	-0.2387	0.0326
	1322	0.5407	0.0295	0.6176	0.0400	-0.2715	0.0312
	2123	0.5340	0.0314	0.6135	0.0403	-0.2920	0.0290
	3221	0.5340	0.0291	0.6122	0.0397	-0.2490	0.0291
	3313	0.5050	0.0311	0.5922	0.0398	-0.2608	0.0304
	1233	0.4937	0.0283	0.5895	0.0391	-0.2713	0.0301
	1333	0.4515	0.0286	0.5653	0.0394	-0.3210	0.0314
	3333	0.3419	0.0178	0.4766	0.0333	-0.3046	0.0274
Best	State	1.4901	0.0745	3.0672	0.9283	-0.0527	0.1400
Dead		0.0000	-	0.0000	-		

*Standard errors are provided for illustration purposes only.

Table 5: Correlations between utility differences between the worst states 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	0.9282	1.0000				
AQL-5D	55555	0.9080	0.9019	1.0000			
HUI2	455445	0.9412	0.9303	0.9094	1.0000		
ICECAP	44444	0.9299	0.9210	0.9020	0.9321	1.0000	
OPUS	3333	0.9249	0.9167	0.8998	0.9270	0.9170	1.0000

Table 6: Estimated coefficients for the regressions of the relationship between health state values from PBMs on their original scales to a common scale based on the rank ordered mixed logit.

Variable		**Estimate**	**s.e.**
EQ-5D	Constant	7.0651	0.4429
	Health State Value	4.6620	0.5417
SF-6D	Constant	5.1368	0.4757
	Health State Value	5.8356	0.4070
AQL-5D	Constant	5.2427	0.4791
	Health State Value	4.4695	0.3934
HUI2	Constant	4.2007	0.5164
	Health State Value	6.2987	0.6917
ICECAP	Constant	6.2803	0.5015
	Health State Value	5.6808	0.6155
OPUS	Constant	6.3458	0.4006
	Health State Value	12.3298	1.0734
	(Health State Value)^2	-28.9721	3.2768
	(Health State Value)^3	22.1348	2.5449
	θ_1	0.7186	0.1419
	θ_2	0.1865	0.0946
	θ_3	0.1562	0.0576
	θ_4	0.5430	0.1213
	θ_5	0.5442	0.1325
	θ_6	0.0000	0.1696

Table 7: Implied relationship between health state valuations on original and common scales (Dependent variable: common scale 0: 'dead', 1: 'EQ-5D 11111' and 'OPUS 1111').

	EQ-5D	**SF-6D**	**AQL-5D**	**HUI2**	**ICECAP**	**OPUS**
Constant	0.5405	0.3930	0.4011	0.3214	0.4805	0.4855
Health State Value	0.3567	0.4464	0.3419	0.4819	0.4346	0.9433
(Health State Value)^2						-2.2164
(Health State Value)^3						1.6934

Table 8: Estimated EQ-5D mapping ranges.

	Published value range		**Common metric**		**EQ-5D mapping range**	
EQ-5D	-0.594	to 1	0.329	to 0.897	-0.594	to 1
SF-6D	0.271	to 1	0.514	to 0.839	-0.074	to 0.838
AQL-5D	0.431	to 1	0.548	to 0.743	0.022	to 0.568
HUI2	-0.0552	to 1	0.295	to 0.803	-0.689	to 0.737
ICECAP	0	to 1	0.480	to 0.915	-0.168	to 1.050
OPUS	0	to 1	0.485	to 0.906	-0.154	to1.024

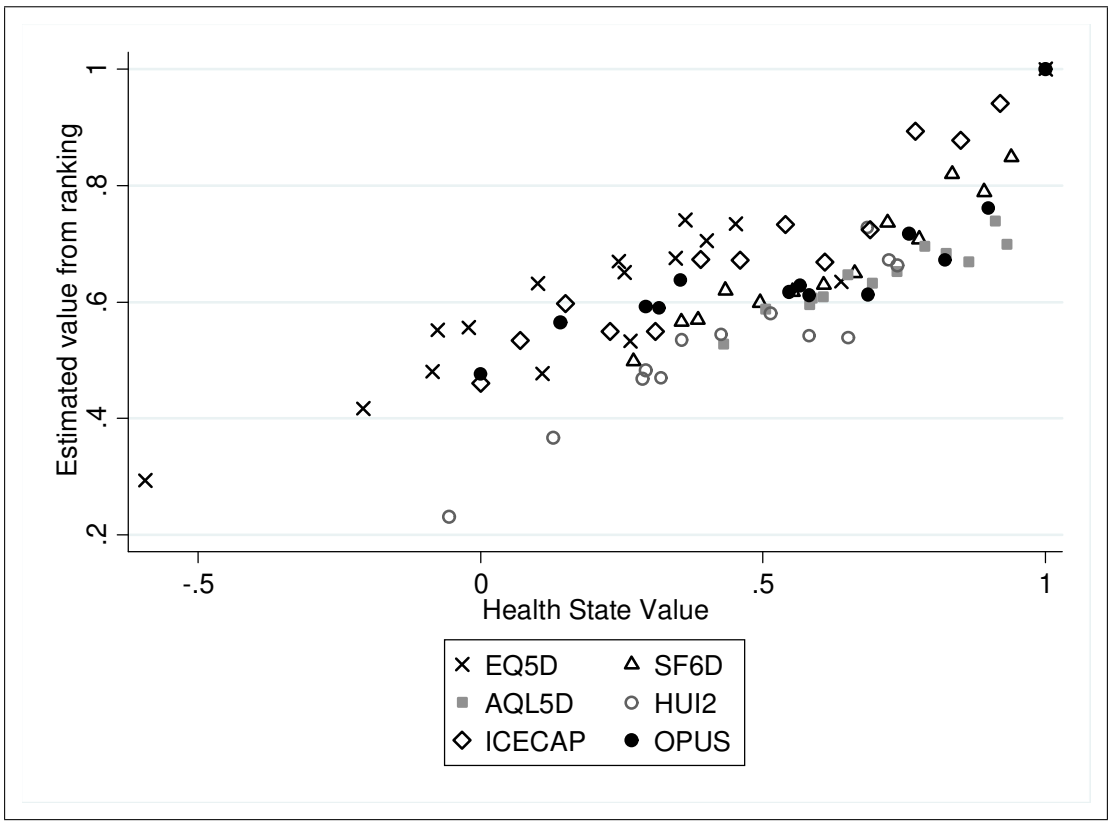

Figure 1: Scatter plot of estimated β_j versus current published state values.

Technical appendix: Model specification and identification.

Specification of the standard rank ordered logit model

It is assumed that individual i faces J different alternatives in each of the T choice situations. Both, the number of alternatives and choice situations might differ and therefore the notation J_{it} and T_i is more appropriate but for simplicity of exposition and without loss of generality we 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 assumed to be a linear function of some fixed parameters β and an unknown stochastic part, ε_{ijt} which is assumed independent and identically distributed (IID) type I extreme value.

$$U_{ijt} = \nu_{ijt} + \varepsilon_{ijt} \qquad i = 1, 2, \ldots, n; \quad j = 1, 2, \ldots, J; \quad t = 1, 2, \ldots, T \tag{A1}$$

In each choice situation, the individual chooses the alternative with the highest utility. 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

$$\Pr\left(\mathbf{R}_{it}\right) = \prod_{l=1}^{J-1} \frac{\exp\left(\nu_{ir_{it}^l t}\right)}{\sum_{s=l}^{J} \exp\left(\nu_{ir_{it}^s t}\right)} \tag{A2}$$

Effectively, each ranking is expressed as $J-1$ independent choices by the individual.

Allison and Christakis [22] 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 we do 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 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 (A2) can be generalised to

$$\Pr(\mathbf{R}_{it}) = \prod_{l=1}^{L} \sum_{p \in Q_l} \prod_{k=1}^{K_l} \frac{\exp\left(\nu_{ir_{it}^{l}[p_k]t}\right)}{\sum_{s=k}^{K_l} \exp\left(\nu_{ir_{it}^{l}[p_s]t}\right) + \sum_{s>l} \sum_{m=1}^{K_s} \exp\left(\nu_{ir_{it}^{s}[p_m]t}\right)} \tag{A3}$$

and the probability of observing a set of rankings by an individual can be written as

$$\mathbf{P}_i = \prod_{t=1}^{T} \Pr(\mathbf{R}_{it}) \tag{A4}$$

The model is estimated by maximizing the likelihood function that uses the above probability for each individual in the sample.

Specification of a general rank ordered mixed logit model

The utility that individual i gets from alternative j in choice situation t for a mixed logit model is analogous to the utility in (A1) but with an additional error component $\boldsymbol{\xi}_{ijt}$:

$$U_{ijt} = \nu_{ijt} + \boldsymbol{\xi}_{ijt} + \varepsilon_{ijt}$$

This additional error component can be correlated between alternatives and choice situations and can be heteroskedastic. It is assumed to have zero mean and a distribution $f(\boldsymbol{\xi}|\Psi)$ where Ψ is a vector of fixed parameters that determine this distribution and need to be estimated in addition to the rest of parameters in the model. Conditioning on $\boldsymbol{\xi}$ the probability of a given choice is logit and the probability of observing a certain set of rankings is analogous to that in equation (A4) and can be expressed as

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

Since $\boldsymbol{\xi}$ is not known, $\mathbf{P}_i(\boldsymbol{\xi})$ needs to be integrated over the density of ξ to obtain the

unconditional probability of the sequence of choices for person i

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

The model is usually estimated by simulated maximum likelihood since the loglikelihood using $\mathbb{P}_i$ as the probability of observing the sample data for each individual in the sample.

Identification of the specific application of the rank ordered mixed logit 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 $\pi^2/3$. Alternatively, the same normalisation can be achieved by setting two of the alternative specific constants to two different values and allowing the variance of the error term to be estimated freely as $(\lambda\pi)^2/3$. In the present case there is a natural normalisation for the model since the utility preference weights are anchored at one for full health and usually zero for dead. 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)[3] and directly estimating the scale parameter λ, 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 [29], 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

[3]In general, to set the scale, one constant needs to be set to a known value. In the present case, it seems sensible to set the constant of OPUS1111 and/or EQ-5D 11111 to one given that their published health state value is one for both.

factor analytic form as follows

$$M_i U_i = M_i \beta + M_i F V \boldsymbol{\xi}_i + M_i \varepsilon_i$$

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, τ_j, V is a (6×6) diagonal matrix containing all ϖ_s $(s = 1, \ldots, 5)$ and 1 as the diagonal elements, $\boldsymbol{\xi}_i$ is a (6×1) vector of IID standard normal random variables, ξ_i^s $(s = 1, \ldots, 6)$ and, finally, 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:

$$cov\left(\Delta U_i\right) = \Delta M_i F V V' F' M_i' \Delta' + \Delta \frac{\left(\lambda\pi\right)^2}{3} I_{J_{it}T_i} \Delta'$$

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 (2) 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 ϖ_s's are theoretically identified. The normalisation needed in the factor loadings have been shown to be arbitrary in [29] 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 [33]. 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.