Mapping between preference-based measures of health via a common yardstick

Donna Rowen[1]*, John Brazier[1], Aki Tsuchiya[1,2] and Monica Hernandez[2]

Background: Different preference-based measures may assign different index scores to the same patients and hence mapping is one solution to estimate the relationship between different preference-based measures. This paper builds on a new approach to mapping between preference-based measures (PBM) using general population preferences. A previous HESG paper reported a survey to value health states defined by the descriptive systems of different measures. This paper reports on the next stage that involves using a common yardstick to map between measures.

Methods: A valuation study, where each interview involves simultaneous valuation of health states defined by multiple PBM using 3 ranking and visual analogue (VAS) tasks. Each interview involves states from 3 of 6 PBM: EQ-5D (generic), SF-6D (generic), HUI2 (generic for children), AQL-5D (asthma specific), OPUS (social care specific), ICECAP (capabilities). Regression techniques are then used to estimate the relationship between mean VAS values from the valuation study and the original value set (i.e. tariff) using a range of models. These results are subsequently used to estimate the relationship across all 6 PBM to enable 'preference-based mapping' between PBM.

Data: Sample of 502 members of the UK general population (response rate=55%, completion rate=99% for all states in the rank and rating tasks). Original value sets for each instrument are also used.

Preliminary results: Performance of the regression models varies considerably across instruments using R^2 (0.83 to 0.97), mean absolute error of state predictions (0.023 to 0.068) and Akaike Information Criteria, with poorest performance for OPUS and EQ-5D models. Linear models are appropriate for most instruments using mean level data. Preference-based mapping here produces values outside the range of existing value sets.

Conclusions: This feasibility study on a new method of mapping between PBM using general population preferences rather than statistical association is better able to take advantage of diversity in descriptive systems across a wide range of PBM. This will enable the integration of evidence from a larger range of studies for economic evaluation and hence enable better cost effectiveness models to be produced.

Key words: Preference-based measures of health; quality of life; mapping; Visual Analogue Scale

[1] Health Economics and Decision Science, University of Sheffield, United Kingdom
[2] Department of Economics, University of Sheffield, United Kingdom
* Correspondence to: Donna Rowen, Health Economics and Decision Science, University of Sheffield, Regent Court, 30 Regent Street, Sheffield, S1 4DA, UK. Telephone: +44114 222 0728. Fax: +44114 272 4095. Email: d.rowen@sheffield.ac.uk

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. The quality adjustment weight for the QALY is estimated using preference-based measures of health-related quality of life. Preference-based measures of health-related quality of life generate a single index score which can be compared across different health care interventions or programmes to inform resource allocation. Preference-based measures 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, with an upper anchor at full health (=1) and a lower anchor at 0 (assuming it is equivalent to dead). Theoretically all instruments should be comparable to each other, where the value of a health state for a patient is identical regardless of the instrument used, yet this is not true in practice. Many studies have illustrated that different generic preference-based measures may assign different index scores to the same patients (see for example Brazier et al, 2004; Barton et al, 2004; Espallargues et al, 2005; Longworth and Bryan, 2003; O'Brien et al, 2003). This creates problems for evidence synthesis within a patient group and creates a barrier for comparisons of cost effectiveness within and between conditions.

One major advantage of the QALY is that it produces a single measure that can be used to produce a common measurement to enable comparison across all types of health care interventions and programmes. Yet crucially the quality adjustment weight of the QALY may be affected by which preference-based measure is used. Many different measures including multiple generic measures are commonly used and this will continue where there is no common agreement on the use of a single measure for all patient groups and conditions. Indeed, in recent years the number of preference-based measures has prolifically increased. Preference-based measures fall into three categories; generic measures for adults (Dolan, 1997; Feeny et al, 2002; Kaplan and Anderson, 1988; Brazier et al, 2002); measures designed for specific groups, such as older people (Coast et al, 2006) or children (Torrance et al, 1996; Stevens, 2008); and condition-specific measures designed for specific medical conditions, such as those for asthma (Revicki et al, 1998; Yang et al, 2007).

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. Mapping is one solution to estimate the relationship between different preference-based measures. Typically mapping by statistical association has involved a separate dataset that contains two or more measures, regression methods are used to estimate a statistical relationship between the indices generated by the measures or their descriptive systems and the regression results are then applied to the study dataset. This type of mapping is commonly undertaken to enable a utility score to be predicted when no preference-based measure has been included in the study rather than to compare across different preference-based measures (for example, Brazier et al, 2008a; Franks et al, 2004; Gray et al, 2006; Nichol et al, 2001; Tsuchiya et al, 2002). It therefore enables a cost utility analysis to be undertaken even when no preference-based measure was included in the original dataset. This method uses statistical association between the measures and relies on a degree of overlap in the descriptive systems of the measures, but key dimensions may not be present in both measures. Indeed, some

mapping studies have large errors and demonstrate a relationship between severity and the size of the error (see Brazier et al, 2008b; Rowen et al, 2008a) and usually there is no measurement of uncertainty. The regression relationship may be affected by the dataset used and hence may differ for different patient groups or conditions. This method also assumes it is appropriate to use different instruments on the same population, which may not be the case for all patient groups and all conditions. Therefore this method of statistical mapping seems unsuitable for mapping between preference-based measures across all patient groups and conditions.

Here we propose a new method of mapping that relates the responses on one preference-based measure to another by using a common metric whilst preserving the advantages of the descriptive system of each instrument. This method of 'preference-based mapping' uses general population values for hypothetical health states defined by the descriptive systems of different measures to provide the common metric for conversion between instruments. This means that the relationship between different instruments is determined directly by people's preferences for different hypothetical states and not by associations in self-reported values. A previous HESG paper reported the valuation survey involving health states defined by descriptive systems of different measures (Rowen et al, 2008b). This paper reports on the next stage that involves using the new common yardstick to map between measures such that a score generated by one preference-based measure can be converted into the equivalent score generated using an alternative preference-based measure.

1.1 Measures of health and quality of life

The study involves 6 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 measures reflects a range of different types of measures that are currently in use or nearing use in the UK. These are summarised in table 1 and detailed in the previous HESG paper (Rowen et al, 2008b – available on request).

- Insert table 1 here –

2. Methods

The study to enable preference-based mapping has three stages. The first stage is a valuation study to determine people's preferences for health and well-being states described by different descriptive systems. The second stage is an analysis of the valuation study to estimate the relationship between the states included in the valuation study. The final stage is to estimate the relationship (i) between the states included in the valuation study and the published value set for each instrument; and (ii) across all instruments in order to enable preference-based mapping between instruments. This paper reports on the final stage of the study for the VAS data collected in the valuation survey. The first two stages are described here briefly, for further details see Rowen et al, 2008b.

2.1 Valuation study

The aim of this study is different to the usual valuation study; here we aim to determine the relationship between instruments, and do not need to produce or reproduce the entire value set for any of the instruments. A valuation study was conducted where each interview involved the simultaneous valuation of multiple

preference-based measures using 3 ranking and visual analogue (VAS) tasks. Respondents were from the geographical areas in the North of England including urban and rural areas with a mix of socio-economic characteristics. Each interview involved hypothetical states from 3 of 6 preference-based measures outlined above: EQ-5D, SF-6D, HUI2, AQL-5D, OPUS, ICECAP. Each task contains 8 states across 2 instruments: one mild state, one moderate state and the worst state for one instrument, one mild state, one moderate state and the worst state for a second instrument, and a generic 'best state' and 'dead'. Each interview involves 15 health and well-being states (5 for each instrument: two mild states, two moderate states and the worst state), the same health and well-being states are used for the 3 rank and VAS tasks.

Sixteen states for the EQ-5D are included in the valuation study, selected using an orthogonal design in SPSS. Thirteen health and well-being states for all other instruments are included, selected to reflect a range of health state values according to the published value set (or most recent version for AQL-5D and ICECAP) whilst guaranteeing a variety of levels for each dimension. The valuation study was designed so that each instrument appeared with each other instrument an equal number of times and each health state is valued approximately 75-100 times (with the exception of 50 times for 4 EQ-5D states) and 500 times for each worst state.

2.2 Relationship between states included in the study – VAS results

Raw VAS ratings measured on the 0 to 100 scale were rescaled for each task completed using the following equation (MVH group, 1994):

$$A_j = \frac{R_j - R(dead)}{R(best) - R(dead)} \quad (1)$$

where A_j represents the adjusted VAS rating for each health state $j = 1, \quad 2, \quad ..., \quad J$, *R(dead)* represents the raw rating given to 'dead', R_j represents the raw rating given to health state *j* and *R(best)* represents the raw rating given to the best health state.[3] This rescales the values such that the highest valued state equals 1 and dead equals 0, hence states can have a value worse than dead.

Mean, SD, median and interquartile range (IQR) were analysed for raw and adjusted VAS data for all states, and adjusted VAS values were compared to published value sets. For the EQ-5D predicted adjusted VAS score was compared to the published VAS value set. Mean VAS values for health states were used rather than predicted values to minimise error. We modelled the VAS utility value on the vector of all health states (not dimension and level dummies as used in usual health state valuation modelling as this is not feasible for this study design) and background characteristics using a maximum likelihood random effects model but no background characteristics variables were significant.

[3] One alternative rescaling method of VAS data is the Parducci Range-Frequency model (see Parducci and Wedell (1986)). However this requires the arbitrary setting of crucial weighting parameters and hence has not been used here. The MVH EQ-5D study adjusted VAS values using equation (1) where *R(best)* represents the raw rating given to state 11111, generic full health. This adjustment cannot be used here as we do not have instrument specific best state for all instruments.

The rank data will be analysed elsewhere using a mixed logit model that will take into account the panel structure of the data and at the same time relax the IIA property. The results will be reported elsewhere.

2.3 Relationship between VAS values and published value sets

Regression techniques are used to estimate for each instrument the relationship between mean VAS values from the valuation study and the original value set. The dependent variable, the rescaled VAS score, is anchored such that highest valued state equals 1 and dead equals 0. The independent variable is taken from the original value set and uses the anchors specified by that instrument. Each regression includes states from one instrument only.

Three models were estimated using mean level data for each instrument: (1) linear; (2) quadratic; and (3) cubic:

$$\bar{y}_j = \alpha + \beta x_j + \varepsilon_j \quad (1)$$

$$\bar{y}_j = \alpha + \beta_1 x_j + \beta_2 x_j^2 + \varepsilon_j \quad (2)$$

$$\bar{y}_j = \alpha + \beta_1 x_j + \beta_2 x_j^2 + \beta_3 x_j^3 + \varepsilon_j \quad (3)$$

where $\bar{y}_j$ represents mean rescaled VAS value, x represents the original value from the original value set, $j = 1, 2, ..., J$ which represents the health states per instrument (16 for EQ-5D and 13 for all other instruments) and ε_j represents the error term. Squared and cubic terms are designed to pick up non-linearities in the relationship between mean VAS values and the original value set for each instrument as there is no reason why the relationship is necessarily linear. OLS is used for all models as the study design means that the VAS data is rescaled to a maximum value of 1 but is not censored at 1. Mean observed values are used here rather than predicted values to minimise error as no background characteristics variables were significant when we modelled the VAS utility value on the vector of all health states and background characteristics using a random effects model.

Three alternative models are also estimated using individual level data for each instrument using a maximum likelihood random effects model: (4) linear; (5) quadratic; and (6) cubic:

$$y_{ij} = \alpha + \beta x_j + \varepsilon_{ij} \quad (4)$$

$$y_{ij} = \alpha + \beta_1 x_j + \beta_2 x_j^2 + \varepsilon_{ij} \quad (5)$$

$$y_{ij} = \alpha + \beta_1 x_j + \beta_2 x_j^2 + \beta_3 x_j^3 + \varepsilon_{ij} \quad (6)$$

where y_{ij} represents rescaled VAS value for individual $i = 1, 2, ..., n$ for health state $j = 1, 2, ..., J$ for each instrument, x represents the value derived using the value set and ε_{ij} represents the error term. The R^2 is sometimes used to judge the goodness of fit of a model. However, it is clear that if extra variables are included in the regression the R^2 will never decrease even if the additional variables are not significant. It is important not to include many variables that may be collinear given the low number of observations in the mean level regressions. There are several alternative criteria

that incorporate a penalty for model complexity, for example the adjusted R^2, the Akaike's Information Criterion (*AIC*) proposed by Akaike (1973) and the Schwarz Bayesian Information Criterion (*BIC*) proposed by Schwarz (1978). Models with higher adjusted R^2 or lower *AIC* and *BIC* are preferred, but these criteria will not necessarily choose the same model. Adjusted R^2 has been shown to increase whenever an additional variable has a t-ratio bigger than one in absolute value. This variable would be considered insignificant at standard significance levels and would not be included if one was to base the selection on standard t-tests. The penalty for the inclusion of additional variables in *BIC* is larger that in *AIC* and therefore *BIC* tends to choose models with less variables than *AIC*. In this paper, performance of all regression models is reported using R^2, mean absolute error of state predictions and AIC.

2.4 Preference-based mapping across all instruments

The regression results are subsequently used to estimate the relationship across all 6 preference-based measures to enable preference-based mapping between preference-based measures. Figure 1 illustrates the process. The preferred regression equations estimated above for each instrument are used to enable the utility value for one instrument, instrument *A* to be converted into an equivalent predicted utility value for another instrument, instrument *B*. This is done by taking the utility value for instrument *A* and using the prediction equations firstly to predict the VAS valuation value and secondly to predict the utility value for instrument *B* using the predicted VAS value. These are solved using the regression equations as simultaneous equations. The valuation study described above involving the simultaneous valuation of multiple instruments provides the common metric to enable conversion between instruments. The method uses the statistical relationship between value sets and the values derived in this valuation study as a means of prediction; this is the relationship between valuations of the same descriptive system using different valuation methodology.

- Insert figure 1 here -

3. The data

There were 502 successfully conducted interviews, a response rate of 55% for suitable respondents answering their door at time of interview. The study achieved a completion rate of 99% for all states included in the rank and rating tasks (140 rank values and 178 VAS values missing out of 12,048 values) and 94.8% (476/502) of respondents had complete VAS responses and 97.2% (488/502) of respondents had complete rank responses. Two respondents (0.4%) had no rank or rating responses, one respondent and one further task for one respondent are excluded for unusable responses ('dead' is valued higher than all states other than 'best state'). All other responses are used in the analysis reported here.

4. Preliminary results

4.1 Mapping VAS values and published value sets

Tables 2 and 3 present the regression results using mean and individual level data respectively for the linear, quadratic and cubic models for all instruments and reports R^2, absolute error of state predictions and Akaike Information Criteria (AIC). Performance of the regression models varies considerably across instruments but not much within instruments across models using R^2 (0.83 to 0.97 for mean models),

mean absolute error of state predictions (0.023 to 0.068) and AIC, with poorest performance for OPUS and EQ-5D models. Generally the linear models are preferred for the mean level data according to AIC and mean absolute error of state predictions. Table 2 shows that the quadratic and cubic regression models using mean level data suffer from multicollinearity as evidenced by high R^2 values in regressions where none or almost none of the coefficients are significant. The estimated OLS coefficients still remain unbiased but standard errors are large. However, we are interested in predictions and multicollinearity has no significant effect on them. Regressions estimated on mean and individual level data have similar mean absolute error of state predictions, yet the linear model is not generally preferred for the individual level data according to AIC and mean absolute error of state predictions.

- Insert tables 2 and 3 here –

4.2 Mapping across all instruments

Tables 4 and 5 are look-up tables presenting the predicted values for all instruments in comparison to the common VAS metric from the valuation study. Table 4 is a look-up table presenting the predicted value for all instruments using the linear mean level regressions reported in table 2. Table 5 is a look-up table using the linear individual level regressions reported in table 3. For example, table 4 illustrates that a health state valued at 0.5 using our common VAS metric is equivalent to an EQ-5D state with tariff value 0.223 and an SF-6D state with tariff value 0.545. Tables 6 and 7 are look-up tables presenting the predicted values for EQ-5D in comparison to values from all other instruments for the mean level and individual level regressions respectively. For example, table 6 illustrates that a health state with SF-6D value of 0.5 has an equivalent EQ-5D value of 0.152. The predicted values go above 1 and below the bottom of the range for some instruments. This is a feature of the way the prediction equations have been generated as they have not been constrained to the value set range of each instrument.

Figure 2 illustrates observed EQ-5D and SF-6D values for a patient dataset with a line demonstrating the preference-based mapping conversion between EQ-5D and SF-6D. The dataset consists of 2515 cases of self-reported SF-6D and EQ-5D created using seven patient group datasets (chronic obstructive airways disease, osteoarthritis, irritable bowel syndrome, lower back pain, leg ulcers, post menopausal women and elderly) collected in various studies undertaken at the University of Sheffield, UK (see Brazier et al., 2004). This comparison demonstrates the difference between preference-based mapping to convert between EQ-5D and SF-6D values and self-reported EQ-5D and SF-6D values.

- Insert tables 4, 5, 6 and 7 and figure 2 here -

5. Discussion

The relationships between VAS values derived here and the existing value sets are not straightforward. The value sets use a variety of preference elicitation techniques; time trade-off, standard gamble, best-worst scaling and discrete choice experiment. The differences between the values elicited using different techniques is well established in the literature for time trade-off, standard gamble and VAS but is less established for best-worst scaling and discrete choice experiments. However the relationship between our VAS values and the value set is affected not only by the different

valuation technique. We expect differences in the values as the populations are different (locations, year, sampling method), the set of health states for each instrument is not identical to the valuation study and all states are rated on VAS alongside other states. We therefore do not wish to focus upon differences in VAS values and the value sets but instead observe logical inconsistencies. Indeed, we expect differences; the question is whether we can map between VAS values and the value sets. There are two logical inconsistencies for AQL-5D and one for ICECAP in the mean VAS values and none in the value set which presents a challenge to the mapping functions for these instruments.

The purpose of a mapping function is not simply to estimate the relationship between the dependent and independent variables. Rather, the purpose of a mapping function is to predict values of the dependent variable using a separate dataset that contains only the independent variables. Explanatory power focuses upon how well the model explains the dataset it was estimated on, which is of general interest here but is not a useful basis for assessing the model performance of the mapping function. Mean absolute error is a better indicator of how large the prediction errors are and whether this of a minimal important difference. All regressions in tables 2 and 3 have errors below 0.07, which is lower than many errors reported in papers using mapping by statistical association (for example see Rowen et al., 2008a, and Brazier et al., 2008b). Mapping functions are often tested using a separate dataset that contains both the dependent and independent variables to analyse the performance of the mapping function (for example see Franks et al, 2004; Gray et al, 2006). However this is not possible here as there is no other dataset that contains similar data and our dataset is not sufficiently large to split into separate estimation and validation datasets.

Background characteristics are not included in the models since in preliminary data analysis they were found not to be significant. These mapping functions can therefore be used even on published data where these variables may not be available. AIC chooses the linear regression models in five out of the six instruments when using the mean level data. The only exception to this is the regression for OPUS where AIC clearly selects the cubic regression as a better fitting model. The models selected by AIC also tend to have the smallest mean absolute error. In cases where this does not happen the differences in mean absolute error between the model selected by AIC and the model with the smallest mean absolute error are small. Some of the regressions including higher order terms have obvious signs of multicollinearity, although the models, had they been selected, would still give unbiased predictions. When the individual level data is used the picture is not so clear cut. AIC chooses the linear regressions only for three out of the six instruments (EQ-5D, AQL-5D and ICECAP). For the remaining three instruments (SF-6D, HUI2 and OPUS) AIC chooses the cubic regression. However for simplicity and comparability the linear models are used here for mapping between instruments reported in tables 4, 5, 6 and 7.

It is unclear whether the mean or individual level regressions are appropriate for estimating for each instrument the relationship between mean VAS values from the valuation study and the original value set. The mean level models are deterministic analyses that ignore issues of uncertainty and the distribution around the mean, but pragmatically are in accordance with techniques used in health economics. For example, Torrance et al. (1996) and Stevens et al. (2006) use mean level regression models to estimate the relationship between standard gamble and VAS responses.

Individual level models suffer from the problem that the value set does not differ across individuals. This is one important difference to the existing mapping literature between different methods of eliciting health state values (for example Dolan and Sutton, 1997; Stevens et al., 2006). In principle, individual level models would be preferred to be able to exploit a richer dataset. However, independent variables are not variables but estimated parameters. Future research in this area should involve exploration of how to address this problem using different estimation methods. The use of the individual or mean level regression results affects the mapping results reported in tables 4, 5, 6 and 7, yet the results are largely similar.

The look-up tables presented here should be interpreted as preliminary results from a feasibility study that enable mapping between instruments using preference-based mapping rather than mapping by statistical association. The meaning of mapping between measures that are used on different populations or different patient groups or between health-related quality of life measures and quality of life measures is not straightforward. The preference-based mapping undertaken here enables conversion from a utility estimate from one instrument to an equivalent predicted utility value for an alternative instrument. This places no restriction upon whether conversion between the two instruments makes sense theoretically. Indeed, conversion from a condition-specific measure to a generic measure is often undertaken for economic evaluation yet the reverse is rarely pursued.

Tables 4 and 5 indicate that for high and low VAS values in our study the equivalent predicted utility values for the other instrument are higher and lower than the value set range for all instruments. However, tables 6 and 7 indicate that for all instruments the equivalent predicted EQ-5D value is not lower than the EQ-5D value set range when the other instrument is within its value set range and predicted EQ-5D value is only higher than the value set range when converting from ICECAP. Figure 2 shows observed self-reported EQ-5D and SF-6D values for a patient dataset and the conversion between EQ-5D and SF-6D using preference-based mapping. The plot suggests that preference-based mapping predicts a lower EQ-5D value for mild states than may be expected using mapping by statistical association (which uses the relationship between observed EQ-5D and SF-6D values).

Overall these results raise a number of interesting points for further discussion including:

- the relationships between the instruments
- the implications of mapping from a child to an adult measure, from a generic to a condition specific measure, or from a health-related quality of life to a wellbeing or quality of life measure
- what it means when we have a predicted value above 1 for an instrument or a predicted value below the bottom of the range for a given instrument
- whether the predictions for each instrument should be bounded by the range of their value set

6. Conclusions

This paper presents a new approach to mapping between preference-based measures. The paper demonstrates the feasibility of this approach and provides a number of interesting results about the upper and lower anchors of the instruments. For the first time it permits comparisons between instruments on level terms through the use of a

standardised metric. There are important concerns with the statistical methods used to estimate the mapping functions and particularly the estimation of uncertainty needs to be addressed.

Acknowledgements
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. This study is funded by a project grant from the MRC. The usual disclaimer applies.

References

Akaike, H., Information Theory and an Extension of the Maximum Likelihood Principle. In: B. N. Petrov and F. Csaki, eds. *Second International Symposium on Information Theory*. Budapest: Akademiai Kiado, pp. 267–281, 1973.

Barton, G.R., Bankart, J., Davis, A.C., Summerfield, Q.A., 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, 2004.

Brazier. J., Roberts, J., The estimation of a Preference-Based Measure of Health From the SF-12. *Medical Care* 42:851-859, 2004.

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

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

Brazier, J., Czoski-Murray, C., Roberts, J., Brown, M., Symonds, T., Kelleher, C., Estimation of a preference-based index from a condition-specific measure: the King's Health Questionnaire. Medical Decision Making 28:113-126, 2008a.

Brazier J, Yang Y, Tsuchiya A., A review of studies mapping (or cross walking) from non-preference based measures of health to generic preference-based measures. HEDS Discussion paper, University of Sheffield, 2008b.

Brooks, R., Euroqol: the current state of play. *Health Policy* 37:54-72, 1996.

Coast, J., Flynn, T., Grewal, I., Natarajan, L., Lewis, J., Sprotson, K., Developing an index of capability for older people: A new form of measure for public health interventions?. Paper for the Future of Health Conference, Cambridge, July 2006.

Dolan, P., Modeling Valuations for EuroQol Health States. *Medical Care* 35:1095-1108, 1997.

Dolan, P., Sutton, M., Mapping visual analogue scale health state valuations onto standard gamble and time trade-off values. *Social Science and Medicine* 44:1519-1530, 1997.

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

Feeny, D., Furlong, W., Torrance, G.W., Goldsmith, C.H., Zhu, Z., DePauw, S., Denton, M., Boyle, M., Multi-attribute and single-attribute utility functions for the Health Utilities Index Mark 3 system. *Medical Care* 40,113-128, 2002.

Franks, P., Lubetkin, E. I., Gold Marthe R, Tancredi, D. J., Haomiao, J., Mapping the SF-12 to the EuroQol EQ-5D Index in a National US Sample. *Medical Decision Making* 24:247-254, 2004.

Gray, A. M., Rivero-Arias, O., Clarke, P. M., Estimating the Association between SF-12 Responses and EQ-5D Utility Values by Response Mapping. *Medical Decision Making* 26:18-29, 2006.

Grewal, I., Lewis, J., Flynn, T.N., Brown, J., Bond, J., Coast, J., Developing attributes for a generic quality of life measure for older people: preferences or capabilities? *Social Science and Medicine* 62:1891-1901, 2006.

Kaplan, R.M., Anderson, J.P., A general health policy model: Update and applications. *Health Services Research* 23:203-235, 1998.

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

McCabe, C., Stevens, K., Roberts, J., Brazier, J., Health state values for the HUI 2 descriptive system: Results from a UK survey. *Health Economics* 14:231-244, 2005.

MVH group, The measurement and valuation of health: first report on the main survey. Centre for Health Economics, University of York, York, UK, 1994.

Nichol, M.B., Sengupta, N., Globe, D.R., Evaluating Quality Adjusted Life Years: Estimation of the Health Utility Index (HUI2) from the SF-36. *Medical Decision Making* 21:105-112, 2001.

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

Revicki, D.A., Leidy, N.K., Brennan-Deimer, F., Sorenson, S. and Togias, A., Integrating patients preferences into health outcomes assessment: The multi-attribute asthma symptom utility index. *Chest* 114:998-1007, 1998.

Rowen, D., Brazier, J., Roberts, J., Mapping SF-36 onto the EQ-5D index: how reliable is the relationship? HEDS Discussion paper, University of Sheffield, 2008a.

Rowen, D., Brazier, J., Tsuchiya, A., Hernandez, M., Simultaneous valuation of multiple instruments: a new approach to mapping. HESG discussion paper Summer 2008, 2008b.

Ryan, M., Netten, A., Skatun, D. and 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* 25:927-944, 2006.

Schwarz, G., Estimating the dimension of a model. *Annals of Statistics* 6(2):461-464, 1978.

Stevens, K.J., McCabe, C.J., Brazier, J.E., Mapping between Visual Analogue Scale and Standard Gamble data; results from the UK Health Utilities Index 2 valuation survey. *Health Economics* 15:527-533, 2006.

Stevens, K.J., Working with children to develop dimensions for a preference based generic paediatric health related quality of life measure. *Health Economics and Decision Sciences Discussion Paper*, 2008.

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

Tsuchiya, A., Brazier,J., McColl,E., Parkin, D., Deriving preference-based single indices from non-preference based condition-specific instruments: Converting AQLQ into EQ5D indices, *Sheffield Health Economics Group Discussion Paper*, 2002.

Yang, Y., Tsuchiya, A., Brazier, J., Young, T., Estimating a preference-based single index from the Asthma Quality of Life Questionnaire (AQLQ). *Health Economics and Decision Sciences Discussion Paper*, 2007.

Table 1 Measures of health and quality of life

Instrument	*Summary*	*Dimensions*	*Levels*	*Unique states*	*Reference*	*Valuation technique and reference for value set used here*	*Value set range*
EQ-5D	Generic	5 dimensions: Mobility, self-care, usual activity, pain/discomfort and anxiety/depression	3 levels: no problems, some problems, extreme problems	243	Brooks (1996)	Time trade-off, Dolan (1997)	-0.594 to 1
SF-6D	Generic	6 dimensions: Physical functioning, role limitations, social functioning, pain, mental health, vitality	Between 4 and 6 levels, depends on the dimensions	18,000	Brazier et al. (2002)	Standard gamble, Brazier and Roberts (2004)	0.271 to 1
HUI2	Generic for children	7 dimensions: Sensation, mobility, emotion, cognition, self care, pain, fertility	4 or 5 levels, depends on the dimensions	8,000	Torrance et al. (1996)	VAS mapped to standard gamble, McCabe et al. (2005)	-0.0552 to 1
AQL-5D	Condition specific for asthma	5 dimensions: Concern about asthma, shortness of breath, weather and pollution stimuli, sleep impact and activity limitations	5 levels: no problems to extreme problems	3,125	Yang et al. (2007)	Time trade-off, Yang et al. (2007)	0.431 to 1
ICECAP	Capability measure for older people in UK	5 dimensions: Attachment, security, role, enjoyment, control	4 levels: all, a lot, a little, none	1,024	Grewal et al. (2006)	Best-worst scaling, Coast (2006)	0 to 1
OPUS	Social care outcome measure for older people	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	243	Ryan et al. (2006)	Discrete choice experiment, Ryan et al. (2006)	0 to 1

Table 2 Mapping VAS values to tariff values using mean level data

	Model	*Constant*	*t statistic*	*x*	*t statistic*	x^2	*t statistic*	X^3	*t statistic*	*n*	R^2	*AIC*	*Mean absolute error (state level)*
EQ-5D	(1)	0.385	(14.99)**	0.515	(8.16)**					16	0.83	-29.933	0.067
	(2)	0.388	(13.71)**	0.531	(6.63)**	-0.039	(0.34)			16	0.83	-28.075	0.067
	(3)	0.39	(10.63)**	0.523	(4.40)**	-0.055	(0.27)	0.026	(0.1)	16	0.83	-26.088	0.067
SF-6D	(1)	0.059	(1.7)	0.809	(15.07)**					13	0.95	-44.871	0.030
	(2)	0.173	(1.7)	0.388	(1.09)	0.344	(1.19)			13	0.96	-44.600	0.029
	(3)	-0.167	(0.57)	2.354	(1.45)	-3.13	(1.11)	1.905	(1.24)	13	0.97	-44.661	0.026
HUI2	(1)	0.112	(3.34)**	0.724	(10.60)**					13	0.91	-35.554	0.039
	(2)	0.112	(2.54)*	0.731	(3.14)*	-0.009	(0.03)			13	0.91	-33.556	0.039
	(3)	0.114	(2.49)*	0.927	(2.05)	-0.879	(0.51)	0.842	(0.51)	13	0.91	-31.932	0.041
AQL-5D	(1)	0.016	(0.38)	0.745	(12.37)**					13	0.93	-50.065	0.024
	(2)	-0.04	(0.19)	0.914	(1.5)	-0.122	(0.28)			13	0.93	-48.165	0.025
	(3)	-0.605	(0.63)	3.576	(0.8)	-4.159	(0.62)	1.976	(0.6)	13	0.94	-46.680	0.023
OPUS	(1)	0.218	(5.14)**	0.522	(7.49)**					13	0.84	-29.407	0.048
	(2)	0.282	(5.10)**	0.15	(0.64)	0.366	(1.65)			13	0.87	-30.524	0.051
	(3)	0.209	(4.71)**	1.245	(3.43)**	-2.45	(2.90)*	1.871	(3.39)**	13	0.94	-39.210	0.034
ICECAP	(1)	0.248	(8.28)**	0.687	(12.46)**					13	0.93	-35.611	0.043
	(2)	0.257	(6.03)**	0.626	(2.92)*	0.066	(0.30)			13	0.93	-33.724	0.044
	(3)	0.259	(4.88)**	0.586	(1.11)	0.181	(0.13)	-0.083	(0.08)	13	0.93	-31.734	0.044

Note: * significant at 5%; ** significant at 1%

Table 3 Mapping VAS values to tariff values using individual level data

	Model	*Constant*	*t statistic*	*x*	*t statistic*	x^2	*t statistic*	x^3	*t statistic*	*n*	*Groups*	*AIC*	*Mean absolute error (state level)*
EQ-5D	(4)	0.382	(35.56)**	0.513	(48.86)**					1482	249	-248.46	0.068
	(5)	0.388	(32.25)**	0.515	(48.72)**	-0.024	(1.05)			1482	249	-247.56	0.067
	(6)	0.396	(29.40)**	0.48	(16.92)**	-0.063	(1.69)	0.081	(1.32)	1482	249	-247.30	0.066
SF-6D	(4)	0.044	(3.00)**	0.832	(41.20)**					1487	249	-453.44	0.029
	(5)	0.084	(2.47)*	0.66	(4.96)**	0.152	(1.31)			1487	249	-453.16	0.027
	(6)	-0.221	(2.02)*	2.578	(3.86)**	-3.416	(2.80)**	2.02	(2.93)**	1487	249	-459.73	0.025
HUI2	(4)	0.116	(9.03)**	0.711	(41.98)**					1486	251	-220.32	0.039
	(5)	0.116	(9.04)**	0.684	(13.13)**	0.043	(0.54)			1486	251	-218.61	0.040
	(6)	0.132	(9.29)**	0.956	(8.30)**	-1.392	(2.54)*	1.457	(2.65)**	1486	251	-223.60	0.043
AQL-5D	(4)	-0.001	(0.03)	0.764	(28.90)**					1474	248	-490.62	0.024
	(5)	-0.045	(0.56)	0.91	(3.54)**	-0.111	(0.57)			1474	248	-488.95	0.024
	(6)	-0.680	(1.5)	4.018	(0.83)	-4.963	(1.45)	2.427	(1.42)	1474	248	-488.97	0.023
OPUS	(4)	0.225	(18.04)**	0.509	(36.08)**					1481	249	-374.62	0.047
	(5)	0.239	(18.51)**	0.321	(6.69)**	0.221	(4.10)**			1481	249	-389.31	0.051
	(6)	0.223	(17.23)**	1.275	(11.51)**	-2.674	(8.63)**	2.065	(9.48)**	1481	249	-473.88	0.034
ICECAP	(4)	0.236	(19.19)**	0.716	(46.31)**					1460	244	-317.67	0.042
	(5)	0.232	(18.11)**	0.781	(14.11)**	-0.082	(1.24)			1460	244	-317.19	0.041
	(6)	0.231	(17.79)**	0.801	(5.98)**	-0.146	(0.36)	0.049	(0.16)	1460	244	-315.22	0.042

Note: * significant at 5%; ** significant at 1%

Table 4 Preference-based mapping look-up table: Mapping from VAS results using mean level regressions

VAS	*EQ-5D*	*SF-6D*	*HUI2*	*AQL-5D*	*OPUS*	*ICECAP*
1	*1.194*	*1.163*	*1.227*	*1.321*	*1.498*	*1.095*
0.9	1.000	*1.040*	*1.088*	*1.187*	*1.307*	0.949
0.8	0.806	0.916	0.950	*1.052*	*1.115*	0.803
0.7	0.612	0.792	0.812	0.918	0.923	0.658
0.6	0.417	0.669	0.674	0.784	0.732	0.512
0.5	0.223	0.545	0.536	0.650	0.540	0.367
0.4	0.029	0.422	0.398	0.515	0.349	0.221
0.3	-0.165	0.298	0.260	*0.381*	0.157	0.076
0.2	-0.359	*0.174*	0.122	*0.247*	*-0.034*	*-0.070*
0.1	-0.553	*0.051*	-0.017	*0.113*	*-0.226*	*-0.215*
0	*-0.748*	*-0.073*	*-0.155*	*-0.021*	*-0.418*	*-0.361*

Note: Values above and below the range of each value set are in italics.

Table 5 Preference-based mapping look-up table: Mapping from VAS results using individual level regressions

VAS	*EQ-5D*	*SF-6D*	*HUI2*	*AQL-5D*	*OPUS*	*ICECAP*
1	*1.205*	*1.149*	*1.243*	*1.310*	*1.523*	*1.067*
0.9	*1.010*	*1.029*	*1.103*	*1.179*	*1.326*	0.927
0.8	0.815	0.909	0.962	*1.048*	*1.130*	0.788
0.7	0.620	0.788	0.821	0.918	0.933	0.648
0.6	0.425	0.668	0.681	0.787	0.737	0.508
0.5	0.230	0.548	0.540	0.656	0.540	0.369
0.4	0.035	0.428	0.399	0.525	0.344	0.229
0.3	-0.160	0.308	0.259	*0.394*	0.147	0.089
0.2	-0.355	*0.188*	0.118	*0.263*	*-0.049*	*-0.050*
0.1	-0.550	*0.067*	-0.023	*0.132*	*-0.246*	*-0.190*
0	*-0.745*	*-0.053*	*-0.163*	*0.001*	*-0.442*	*-0.330*

Note: Values above and below the range of each value set are in italics.

Table 6 Preference-based mapping look-up table: Mapping to EQ-5D using mean level regressions

SF-6D	*EQ-5D*	*HUI2*	*EQ-5D*	*AQL-5D*	*EQ-5D*	*OPUS*	*EQ-5D*	*ICECAP*	*EQ-5D*
1	0.938	1	0.876	1	0.73	1	0.689	1	1.07
0.9	0.781	0.9	0.735	0.9	0.59	0.9	0.588	0.9	0.93
0.8	0.624	0.8	0.595	0.8	0.44	0.8	0.487	0.8	0.80
0.7	0.467	0.7	0.454	0.7	0.30	0.7	0.385	0.7	0.67
0.6	0.310	0.6	0.313	0.6	0.15	0.6	0.284	0.6	0.53
0.5	0.152	0.5	0.173	0.5	0.01	0.5	0.183	0.5	0.40
0.4	-0.005	0.4	0.032	0.4	-0.14	0.4	0.081	0.4	0.27
0.3	-0.162	0.3	-0.108			0.3	-0.020	0.3	0.13
		0.2	-0.249			0.2	-0.122	0.2	0.00
		0.1	-0.390			0.1	-0.223	0.1	-0.13
		0	-0.530			0	-0.324	0	-0.27

Note: The look-up tables from SF-6D and AQL-5D are constrained to feasible values using the range of the value set.

Table 7 Preference-based mapping look-up table: Mapping to EQ-5D using individual level regressions

SF-6D	*EQ-5D*	*HUI2*	*EQ-5D*	*AQL-5D*	*EQ-5D*	*OPUS*	*EQ-5D*	*ICECAP*	*EQ-5D*
1	0.963	1	0.867	1	0.74	1	0.686	1	1.11
0.9	0.801	0.9	0.729	0.9	0.59	0.9	0.587	0.9	0.97
0.8	0.639	0.8	0.590	0.8	0.44	0.8	0.488	0.8	0.83
0.7	0.476	0.7	0.452	0.7	0.30	0.7	0.388	0.7	0.69
0.6	0.314	0.6	0.313	0.6	0.15	0.6	0.289	0.6	0.55
0.5	0.152	0.5	0.174	0.5	0.00	0.5	0.190	0.5	0.41
0.4	-0.010	0.4	0.036	0.4	-0.15	0.4	0.091	0.4	0.27
0.3	-0.172	0.3	-0.103			0.3	-0.008	0.3	0.13
		0.2	-0.241			0.2	-0.108	0.2	-0.01
		0.1	-0.380			0.1	-0.207	0.1	-0.15
		0	-0.519			0	-0.306	0	-0.28

Note: The look-up tables from SF-6D and AQL-5D are constrained to feasible values using the range of the value set.

Figure 1 Preference-based mapping across instruments

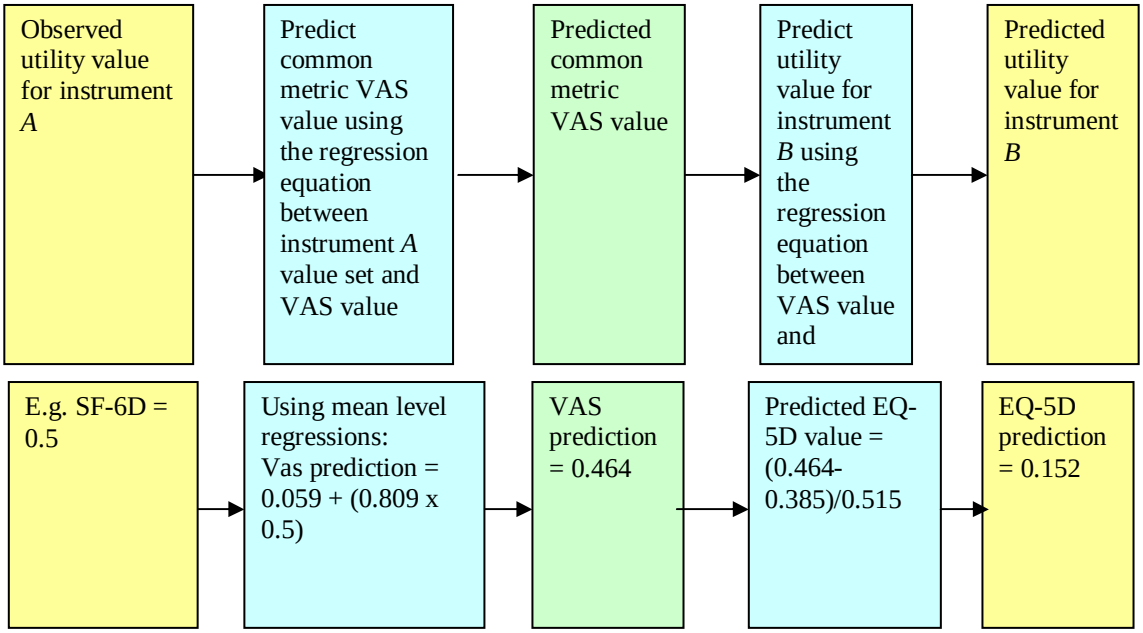

Figure 2 Observed EQ-5D and SF-6D values for a patient dataset with a line demonstrating the preference-based mapping conversion between EQ-5D and SF-6D