Article:

Rowen, D., Brazier, J., Tsuchiya, A. et al. (2 more authors) (2009) The simultaneous valuation of states from multiple instruments using ranking and VAS data: methods and preliminary results. HEDS Discussion Paper, 09/06.

HEDS Discussion Paper 09/06

HEDS Discussion Paper 09/06

Published paper
Rowen, D, Brazier, J, Tsuchiya, A, Hernandez, M. Valuing states from multiple measure on the same VAS: A feasibility study. Health Economics

ScHARR

School of Health and Related Research

The University Of Sheffield.

ScHARR

Health Economics and Decision Science
Discussion Paper Series

No. 09/06

The simultaneous valuation of states from multiple instruments using ranking and VAS data: methods and preliminary results

Donna Rowen[a]*, John Brazier[a], Aki Tsuchiya[a,b], Mónica Hernandez[b] and Rachel Ibbotson[c]

[a] Health Economics and Decision Science, University of Sheffield, Regent Court, 30 Regent Street, Sheffield, S1 4DA, UK
[b] Department of Economics, University of Sheffield, 9 Mappin Street, Sheffield, S1 4DT, UK
[c] Centre for Health and Social Care Research, Sheffield Hallam University, Montgomery House, 32 Collegiate Crescent, Sheffield, S10 2BP, UK
* 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

Abstract

Background: Previous methods of empirical mapping involve using regressions on patient or general population self-report data from datasets involving 2 or more instruments. This approach relies on overlap in the descriptive systems of the measures, but key dimensions may not be present in both measures. Furthermore this assumes it is appropriate to use different instruments on the same population, which may not be the case for all patient groups. The aim of the study described here is to develop a new method of mapping using general population preferences for hypothetical health states defined by the descriptive systems of different measures. This paper presents a description of the methods used in the study and reports on the results of the valuation study including details about the respondents, feasibility and quality (e.g. response rate, completion and consistency) and descriptive results on VAS and ranking data. The use these results to estimate mapping functions between instruments will be presented in a companion paper (Rowen et al, 2009).

Methods: The study used interviewer administered versions of ranking and VAS techniques to value 13 health states defined by each of 6 instruments: EQ-5D (generic), SF-6D (generic), HUI2 (generic for children), AQL-5D (asthma specific), OPUS (social care specific), ICECAP (capabilities). Each interview involved 3 ranking and visual analogue scale (VAS) tasks with states from 3 different instruments where each task involves the simultaneous valuation of multiple instruments. The study includes 13 health and well-being states for each instrument (16 for EQ-5D) that reflect a range of health state values according to the published health state values for each instrument and each health state is valued approximately 75-100 times.

Results: The sample consists of 499 members of the UK general population with a reasonable spread of background characteristics (response rate=55%). The study achieved a completion rate of 99% for all states included in the rank and rating tasks and 94.8% of respondents have complete VAS responses and 97.2% have complete rank responses. Interviewers reported that it is doubtful for 4.1% of respondents that they understood the tasks, and 29.3% of respondents stated that they found the tasks difficult. The results suggest important differences in the range of mean VAS and mean rank values per state across instruments, for example mean VAS values for the worst state vary across instruments from 0.075 to 0.324. Respondents are able to change the ordering of states between the rank and VAS tasks and 12.0% of respondents have one or more differences in their rank and VAS orderings for every task.

Conclusions: This study has demonstrated the feasibility of simultaneously valuing health states from different preference-based instruments. The preliminary analyse of the results presented here provides the basis for a new method of mapping between measures based on general population preferences.

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

1. Introduction

Economic evaluation using preference-based measures of health to generate Quality Adjusted Life Years (QALYs) is being increasingly used to inform health policy. The QALY measure combines both quantity and quality of life into a single measure and quality of life is measured using a preference-based measure of health. Recent years has seen the increasing proliferation of preference-based measures of health. These fall into three categories:

1) Generic measures for adults such as the EQ-5D (Brooks, 1996), the HUI3 (Feeny et al, 2002), the QWB scale (Kaplan and Anderson, 1988) and the SF-6D (Brazier et al, 2002)
2) Measures designed for specific groups, such as those for older people (Grewal et al, 2006), children (Torrance et al, 1996; Stevens, 2009) and social care (Ryan et al, 2006).
3) Condition-specific measures designed for specific medical conditions, such as those for asthma (Revicki et al, 1998; Young et al, 2007) and urinary incontinence (Yang et al, 2008).

If all of these measures are preference-based, measured on an interval scale, with the upper anchor at full health (=1) and the lower anchor at 0 (assuming it is equivalent to dead) then 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. However this is not true in practice. The increasing number of generic and condition specific preference-based measures of health have been shown to generate different scores in the same population (Brazier et al, 2004; Longworth and Bryan, 2003; O'Brien et al, 2003). This can be attributed to differences in their descriptive systems and valuation methods (Brazier et al, 2004; Tsuchiya et al, 2006). This creates a major problem for researchers and policy makers undertaking evidence synthesis and cross programme comparisons across studies using different instruments.

One solution to cross programme comparison is to use one generic preference-based measure in all economic evaluation (Dowie, 2002). However, this is not possible for all groups of patients, for example children require special consideration regarding language comprehension and development (Stevens, 2009), and there may be special considerations for very elderly people (Coast et al, 2008) and those needing social care (Ryan et al, 2006). Furthermore generic measures of health are inappropriate or insensitive for many medical conditions (for example Barton et al, 2004; Espallargues et al, 2005; Harper et al, 1997; Kobelt et al, 1999). Even if these arguments are not accepted, the fact remains that different measures are used across studies within and between conditions and this will continue as long as there is no common agreement internationally on which measure to use.

Previous attempts at mapping between instruments have used existing datasets that have two or more measures used alongside each other, and regression analysis is applied to estimate a statistical relationship between the indices generated by the measures or their descriptive systems (for example, Franks et al, 2004; Gray et al, 2006; Nichol et al, 2001; Tsuchiya et al, 2002). This type of mapping is commonly used to estimate a mapping function that will allow the estimation of a utility score when no preference-based measure has been included in the trial. This method assumes it is appropriate to use different instruments on the same population, which as described above may not be the case for all patient groups and all conditions. This approach also relies on a degree of overlap in the descriptive systems of the measures, but key dimensions may not be present in both measures. This approach also relies on a distribution of patients across the states, to avoid extrapolation when the mapping function is applied to the trial dataset. Instead what is needed is a means of relating the responses on one measure to another using a common metric and preserving the advantages that the descriptive system may bring. This study is testing a new method for mapping between preference-based measures that could be used for evidence synthesis and cross programme comparisons in studies using different preference-based measures.

One possibility would be to use time trade-off (TTO) or standard gamble (SG) preference elicitation techniques on multiple instruments in one sitting, for example see Tsuchiya et al. (2006). However, given limited resources and increasing interest in using ordinal methods a ranking task was used alongside a visual analogue scale (VAS) in our current study. In addition, strictly speaking, TTO and SG are valuation methods where respondents deal with one state at a time, and therefore there is no direct head-to-head comparison of states from different instruments. Ranking and VAS offer alternative techniques that allow for the simultaneous valuation of states. In this paper we present the study design of this new approach to mapping. The paper presents a brief description of the instruments involved in this study, the methodology of the valuation study and the results of the valuation survey in terms of the background of the respondents, response rates, completion rates, consistency of responses and descriptive statistics on the rank and VAS data by states defined by the instruments. The main aim of this paper is to demonstrate the feasibility of simultaneously valuing multiple instruments and to provide the basis for the companion paper (Rowen et al, 2009) that uses these results to map between instruments.

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.

The EQ-5D is the most widely used generic preference-based measure of health-related quality of life which produces utility scores anchored at 0 for dead and 1 for perfect health (Dolan, 1997). The SF-6D is a generic preference-based single index measure for health that can be estimated using SF-36 and SF-12 data (Brazier and Roberts, 2004). The SF-6D is derived from a selection of SF-36 items (Ware et al, 1993) and produces utility scores anchored at 0 for dead and 1 for perfect health. The HUI2 is a generic preference-based measure of health for children (Torrance et al, 1996) and produces utility scores anchored at 0 for dead and 1 for perfect health. The AQL-5D (Young et al, 2007) is a condition-specific preference-based measure of health for asthma derived from the Asthma Quality of Life Questionnaire (AQLQ) (Juniper, 1993) and produces utility scores anchored at 0 for dead and 1 for perfect health (Yang et al, 2007).

ICECAP is a preference-based measure of capability for older people in the UK (Grewal et al, 2006). Utility scores are anchored at 0 for zero capabilities and 1 for perfect health. OPUS (older persons' utility scale) is a preference-based social care outcome measure for older people (Ryan et al, 2006). We use the utility index excluding the safety dimension. Utility scores are anchored at 0 for the worst outcome (high unmet needs on all dimensions) and 1 for the best outcome (no unmet needs on all dimensions). The choice of measures reflects a range of different types of measures that are currently in use or nearing use in the UK.

2. Methods

The aim of the study is to develop a preference-based method of mapping between preference-based instruments. Respondents' preferences are elicited for hypothetical health and well-being states defined by different descriptive systems. This means that the relationship between different instruments is determined directly by people's preferences for different states and not by associations in self-reported values. Unlike mapping by statistical association this approach does not rely on conceptual overlap of instruments and does not rely on the distribution of patients self-reported scores to estimate a mapping function suitable for the full value set range. In the valuation study a sample of health states defined by each of 6 instruments presented above (EQ-5D, SF-6D, HUI2, AQL-5D, OPUS and ICECAP) are valued by a representative sample of the general population.

2.1 Study design

The overall aim of the study is to be able to map between instruments via a common yardstick. It has two stages. The first stage is a valuation study to determine people's preferences for health and well-being states described by the descriptive systems of different instruments using VAS and ranking. The second stage is to estimate mapping functions between the preference-based instruments via the common yardstick of VAS and rank values and the published value sets of the instruments and to use these to map between instruments. The second part of this study is presented in a companion paper (Rowen et al, 2009).

2.2 Valuation task

For use in economic evaluation the health state valuation technique should be choice based (Drummond et al, 2005). Conventionally this has included cardinal methods such as standard gamble (SG) and time-trade off (TTO), yet over recent years there has been an increasing interest in using ordinal methods, such as ranking or pair wise comparisons (Salomon, 2003, 2007; McCabe et al, 2006). In a conventional ranking task respondents are asked to order a set of states from best to worst.

An advantage of using ranking is it is arguably cognitively less complex when eliciting preferences over health states from different instruments than TTO or SG. More importantly, it provides a direct means of comparing health states from different instruments and enables preference elicitation for a greater number of health states per respondent than SG or TTO. This use of ranking to value health states is quite new, so we decided to also use VAS to provide another means of simultaneously valuing states from different instruments and another common yardstick for testing this new method of mapping.

A conventional VAS task was used where respondents are asked to rate health states on a scale from 0 to 100 using a vertical line on a page, where 0 is the 'worst imaginable state' and 100 is the 'best imaginable state'. The main difference with previous applications of VAS is that the end points do not specify health because some of the instruments do not limit themselves to health (e.g. ICECAP). There are concerns with the use of VAS including end point bias (Torrance et al, 2001) and that it does not generate 'preferences' (Brazier et al, 2003). However these problems are less relevant for this study since the aim is not to produce a new a value set per se, but to allow the conversion of scores from one instrument into those of another. VAS and ranking provide cardinal scales to value sets of states and these 'common yardsticks' can be used as a means to relate utility scores generated using the value set for one instrument to utility scores generated using the value set for another instrument.

2.3 Valuation study design

In order to determine the relationship between instruments and to estimate this relationship each respondent values health states from multiple descriptive systems alongside each other. A respondent may value, for example, EQ-5D and SF-6D states during the same valuation task. This may make the respondent more aware of different dimensions or ways of expressing health when seeing health states from two descriptive systems rather than one. This may alter the way they value an instrument when individuals realise that some dimensions of health are explicit in one descriptive system but not the other. One aim of the study design was to reduce any systematic bias brought about through the combination of instruments and states that were valued by each individual both within a valuation task and across all valuation tasks for that individual. The study was designed so that each instrument appeared with each other instrument an equal number of times.

Each interview involves 3 rank tasks and 3 VAS tasks. Respondents are asked to rank a set of 8 cards; 3 states from one instrument (mild, moderate, worst state), 3 states from another instrument (mild, moderate, worst state), plus ‘best state’ and ‘dead’. Respondents are then asked to rate these cards using VAS and are able to change the ordering from the previous ranking. Each interview involves 15 states in total, 5 states each (2 mild states, 2 moderate states, worst state) from 3 different instruments. Each time an instrument is included in a valuation task the worst state defined by that instrument is valued.

There are 20 variations of the interview as there are 20 different combinations of r=3 instruments from the available n=6 instruments using the binomial coefficient equation n!/(r!(n-r)!). Each variation of the interview is different and has a different ‘card bloc’, making 20 card blocs in total. Each instrument appears in 10 out of the 20 card blocs and hence 10 variations of the interview. The interviews are designed so that 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 (since they appear in different combinations).

Each card bloc involves 3 instruments with one ‘card set’ for each instrument, and each card set consists of 4 unique states of various severity plus the worst state as the fifth state. Each card set is used in 3 or 4 blocs[1] and across all blocs any given card set appears at least once in the first ranking task, at least once in the second ranking task and at least once in the third ranking task. This is to minimise any data variation due to respondents changing their understanding, consideration and concentration across the duration of the interview.

Figure 1 shows the task design for 6 out of the 20 card blocs. Instruments are labelled A, B, C, D, E and F and health states for each instrument are numbered 1 to 13, where 1 represents the worst state of each instrument.

2.4 Selection of health states

Sixteen states were selected for the EQ-5D using an orthogonal design in SPSS to generate an orthogonal array of states that enable the estimation of an additive function to value all states. Selection of states using an orthogonal design was not possible for every instrument since the orthogonal array was too large, for example 49

[1] With the exception of EQ-5D where each card set is used in 2 or 3 card blocs due to the larger number of states and hence card sets.

states for the SF-6D alone. For all other instruments twelve health and well-being states were selected plus the worst state. States were selected to reflect a range of health state values according to the published value set (or most recent version for AQL-5D and ICECAP) for each instrument, whilst guaranteeing that a variety of levels for each dimension are included in the chosen states. All health and well-being states were checked to guarantee that the combination of levels and dimensions were feasible and realistic and hence appropriate for use in ranking and VAS tasks undertaken by members of the general population.

Instrument specific full health was not used for all instruments as instead the generic state 'best state' was used. This approach was chosen as the aim is not to focus upon differences or deficiencies in descriptive systems per se but to focus upon differences in values given to health states with different health problems. The instrument specific best state was only included for two instruments: EQ-5D since it is specified by orthogonal array, and OPUS, since it has the smallest descriptive system of all instruments included.

2.5 The interviews

All interviews were conducted by trained and experienced interviewers in the respondent's own home. The interviews were undertaken by the Centre for Research and Evaluation (CRE) at Sheffield Hallam University who have done numerous valuation surveys including the UK valuation of HUI2 (McCabe et al, 2005), AQL-5D (Yang et al, 2007), OAB-5D (Yang et al, 2008), King's Health Questionnaire (Brazier et al, 2008) and TTO and SG valuation of EQ-5D alongside SF-6D (Tsuchiya et al, 2006). Respondents were not offered any financial reward for their participation.

The interview began with the respondent being asked to report their own health using the descriptive system for the EQ-5D and subsequently all instruments that the individual would value during the interview. This familiarised the respondent with the idea of describing states and the items and levels in the descriptive system for each of the instruments involved in their interview.

In the next stage of the interview the respondent was asked to rank a shuffled set of 8 cards (task 1) as described above. The respondent was asked to place the cards in order of how good or bad they think they are. Respondents were asked to imagine that they themselves were actually in each state and that it is going to last for the rest of their life without changing. The respondent was then asked to rate these same states (without reshuffling the cards) using the VAS, allowing respondents to change the ordering from the previous ranking. The respondent was then asked to repeat the ranking and VAS tasks twice with different sets of cards (tasks 2 and 3).

The final stage of the interview involved self-completion questions on background characteristics and how difficult the respondent found the interview and interviewer feedback on whether the respondent understood the tasks.

2.6 Selection of respondents

Respondents were from the geographical areas in the North of England including urban and rural areas with a mix of socio-economic characteristics. Streets were sampled from the selected areas and all willing participants within quota were interviewed. Letters were sent to households informing them the interviewers will be in their area. Interviewers then visited houses and interviewed all people who were willing to participate and within quota from those addresses.

2.7 Analysis

Raw VAS ratings measured on the 0 to 100 scale are rescaled using the following equation (MVH, 1994):

$$A_{ijk} = \frac{R_{ijk} - R\,dead_{ik}}{R\,best_{ik} - R\,dead_{ik}} \qquad (1)$$

where A_{ijk} represents the adjusted VAS rating for each health state j for each individual i for each task k, $R(dead)_{ik}$ represents the raw rating given to 'dead', R_{ijk} represents the raw rating given to health state j and $R(best)_{ik}$ represents the raw rating given to the best health state. This rescales the values per task per individual such that the highest valued state (including the generic 'best state') equals 1 and dead equals 0, hence states can have a value worse than dead.

Rank responses are scored according to their rank from 1 to 8 where 1 is the most preferred state and 8 is the least preferred state. Ties in rank data are scored at the highest rank. For example, the data is coded 1,1,1,4,5,6,7,8 for a respondent who ties 3 states as the most preferred state.

Duration, completion, difficulty self-reported by respondent, understanding and concentration reported by interview are reported for the easiest and most difficult card blocs to demonstrate the range across all card blocs. Mean and standard deviation (SD) of adjusted VAS score and rank score for highest and worst states by instrument are presented. Mean, SD, median and interquartile range (IQR) are reported for raw and adjusted VAS data for all states, and adjusted VAS values and rank scores are compared to published value sets and logical consistency is reported. For the EQ-5D, the predicted adjusted VAS score is compared to the published VAS value set.

3. The data

There were 502 successfully conducted interviews, a response rate of 55% for suitable respondents answering their door at time of interview. Amongst responders, 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. Higher response rates for ranking over VAS were also achieved in the MVH EQ-5D valuation survey (Gudex, 1997). Two respondents (0.4%) have 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.

Characteristics of all respondents with usable rank or rating responses are presented in table 2 and compared to the general population in South Yorkshire and England. The sample population sample is broadly comparable to the regional and national populations, but it has a higher proportion of retired persons and home owners and fewer employed persons. This is not a problem for this study since it is mainly concerned with the feasibility of this approach rather than generating a definitive set of values.

4. Results

4.1 Feasibility

Table 3 presents duration of interview, self-reported difficulty of questions and understanding, and effort and concentration as reported by the interviewer for a selection of card blocs. These card blocs were selected as they were reported as being the easiest and most difficult card blocs using the self-report data from respondents. Yet responses for difficulty, understanding and effort and concentration are similar for all card blocs, suggesting little variation in feasibility across blocs. Mean duration varies from 32 to 41 minutes across all blocs, missing VAS responses range from 0% (0 out of 576) to 5.1% (33 out of 648) and missing rank responses range from 0% (0 out of 600) to 4.9% (32 out of 648) for each card bloc, with missing values seeming unrelated to difficulty. There is no clear pattern of difficult combinations of instruments, as one of the easiest and both of the most difficult blocs contain AQL-5D and ICECAP.

4.2 Results

Table 4 shows the highest and worst states by instrument. The mean adjusted VAS value for the best states for those instruments where their best state was valued, namely EQ-5D (state 11111) and OPUS (state 1111), are lower than one (i.e. 0.909 and 0.899 respectively), indicating that these instrument specific best states are regarded as worse than the generic best state. The worst state is valued for all instruments and the VAS values range between 0.075-0.324 and rank score range from 5.39-6.87. This range reflects the different ‘floors’ of the instruments, with generic EQ-5D having the lowest floor and the disease specific AQL-5D having this highest.

Table 5 presents mean and standard deviation of adjusted VAS scores (where dead=0) and rank scores for all states included in the valuation study. The number of valuations per state other than the worst state is between 70 to 100 for all instruments, except the EQ-5D. There were fewer valuations per EQ-5D state (between 48 to 98), since there were more states valued from that instrument. The worst state of each instrument had between 490 to 501 valuations. Perhaps unsurprisingly all instruments have a wide range of VAS values, where the AQL-5D (condition specific measure) has a smaller range and this is consistent with the published value set utility range using TTO (0.431 to 1). The results suggest important differences in the value range of each of the instruments (with mean worst state values ranging from 0.075 to 0.324). The inter-quartile range suggests that adjusted values vary across individuals and potentially across card blocs. There are no logical inconsistencies in the mean VAS scores, mean rank values or value set for these states for each instrument.

We modelled the VAS utility value on the vector of all health states and background characteristics using a maximum likelihood random effects model. No background characteristics variables were significant and hence to minimise error the mean values for health states are reported here rather than predicted values using the regression model.

Figure 2 plots adjusted VAS values and the published value sets for each instrument, indicating the different relationships for each instrument. Note that the published value sets are typically not based on VAS. It shows that EQ-5D VAS values from our study are mostly higher than the published TTO-based values. Overall our VAS values have a smaller range and spread than the published value set utility scores using other valuation methods. The results indicate that for milder health and well-being states the VAS value is often lower than the published value set utility score, whereas for more severe states the VAS value is often higher than the published value set utility score. This relationship is present for each instrument with the exception of the AQL-5D where all VAS values are lower than the published value set utility scores, and EQ-5D where most VAS values are higher than the published TTO and VAS value set utility scores (as shown in table 5). Figure 3 plots rank scores and the published value sets for each instrument. The relationship is largely similar yet the pattern cannot be precisely determined as our rank values and the value sets are measured using different scales.

4.3 VAS and rank differences

Respondents were able to change the ordering of the health states for the rank and VAS tasks. The Spearman rank correlation coefficient between all rank and implied ranking observations using VAS values is 0.9778, demonstrating high correlation. However, rankings and implied rankings using the VAS values are different for 32.7% of respondents, yet only for 7.5% of all possible responses. Although one third of all respondents have different rankings and implied rankings using the VAS values, this is often only for a small number of states and not for entire rank tasks. In total 12.0% of all respondents have one or more differences for every task between their rankings and implied rankings. Differences in rankings and implied rankings vary across card bloc from 2.2% to 12.0% of total responses per bloc, but there is no clear pattern between differences and the combinations of instruments valued.

4.4 EQ-5D results

The orthogonal array used to select the EQ-5D states in the valuation study enables the estimation of an additive function to value all 243 EQ-5D states. Figure 4 shows the relationship between predicted adjusted VAS score and the MVH EQ-5D VAS published value set for all EQ-5D states. Predicted adjusted VAS score is estimated using a maximum likelihood random effects model using the model specified in Dolan (1997) using the adjusted VAS data. All main effects coefficients are of the expected sign and all are significant at the 1% level with the exception of the dummy variable representing severe problems in usual activities. The ‘N3’ term, a dummy variable for states with at least one dimension at the most severe level, is significant but smaller (0.107) than that in the MVH EQ-5D VAS published value set (0.269). Figure 4 shows that predicted adjusted VAS scores are similar to the MVH VAS value set for mild states, yet indicates a different relationship for more severe states where typically our model overpredicts. The majority of states with at least one dimension at the most severe level, that is those with an ‘N3’ term were over predicted (ICC = 0.899).

5. Discussion

One advantage of VAS or ranking for this type of study over TTO or SG is that respondents see multiple states at the same time. Respondents rank and rate states from different instruments simultaneously and hence the ordering of states is explicit rather than implicit. Respondents were able to value states from different instruments alongside each other using VAS and ranking methods and extremely high completion rates were achieved at the task level.

Each card bloc appears feasible and comparable in terms of completion, understanding, difficulty and effort. Yet the combination of instruments in a given task may affect VAS values. Three of the instruments included in the study are designed for specific patient groups; the HUI2 is designed for children and ICECAP and OPUS are designed for older people. However, the descriptive systems for OPUS and ICECAP do not include any words or dimensions that are applicable only for older people or that indicate the measure is designed for older people. The HUI2 descriptive system used in the study does not include words that indicate that the measure is designed for children, and this is in accordance with the HUI questionnaire that is used to determine the HUI2 and HUI3 health states and utility scores for patients. Therefore respondents will not realise that the measures are designed for different patient groups. Table 3 indicates that simultaneous valuation of HUI2 and ICECAP or OPUS is feasible as the two easiest blocs include HUI2 and either OPUS or ICECAP. One of the instruments included in the study is condition-specific, the AQL-5D for asthma. Valuing AQL-5D alongside other instruments may make respondents more aware of other aspects of health that are not explicitly included in the descriptive system than when the AQL-5D is valued alone. Table 3 suggests that valuing AQL-5D alongside generic measures is feasible as although AQL-5D appears in the most difficult card blocs it also appears in one of the easiest..

The rankings and implied rankings from the VAS values were found to differ for one third of respondents. This is expected as they are different tasks and respondents were able to revise their orderings at any time. However, the relationship between the published value set and rank and VAS values is largely similar as demonstrated in figures 2 and 3 and only 12.0% of respondents have differences in every task between their rankings and implied rankings using VAS. A separate piece of work is being undertaken to convert the rank data into cardinal values using a mixed logit model. The next part of the study will involve analysis on the rank and VAS data to estimate the relationship between the instruments. This stage will indicate the degree to which the use of VAS or rank values are important.

This paper presents the detailed methods of the valuation study and basic descriptive statistics of the rank and VAS values. Further analyses of the VAS and rank data will be reported elsewhere. The next stage of the study will be reported in Rowen et al, 2009. The final stage of the study involves the use of these results to map between the six measures of health and quality of life using the rank and VAS data. This will involve estimating the relationship between the 13 health states valued for each instrument and the original value set. This will be done using the rank and VAS data collected in the valuation study. This mapping uses preferences rather than statistical association and is better able to take advantage of diversity in descriptive systems for different measures. This provides a way of mapping between different preference-based measures that can be used for evidence synthesis and cross programme comparisons in studies using different preference-based measures. 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.

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 the MRC, project number R/112915. The usual disclaimer applies.

References

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., 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, 2008.

Brazier, J., Greene, C., McCabe, C., and Stevens, K., Use of visual analog scales in economic evaluation. Expert Review of Pharmacoeconomics Outcomes Research 3:293-302, 2003.

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.

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

Coast J, Flynn T, Natarajan L, Sprotson K, Lewis J, Louviere JL, Peters TJ. Valuing the ICECAP capability index for older people. Social Science and Medicine 67:874-882, 2008.

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

Dowie, J., Decision validity should determine whether a generic or condition-specific HRQOL measure is used in health care decisions. Health Economics 11:1-8, 2002.

Drummond, M.F., Schulpher, M.J., Torrance, G.W., O'Brien, B.J., Stoddart, G.L., Methods for the Economic Evaluation of Health Care Programmes Oxford: Oxford University Press, 2005.

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.

Gudex, C., Dolan, P., Kind, P., Thomas, R., Williams, A., Valuing health states: Interviews with the general public. European Journal of Public Health 7:441-448, 1997.

Harper, R., Brazier, J.E., Waterhouse, J.C., Walters, S.J., Jones, N.M.B., Howard, P. (1997) A comparison of outcome measures for patients with chronic obstructive pulmonary disease (COPD) in an outpatient setting. Thorax 52:879-887, 1997.

Juniper, E.F., Guyatt, G.H., Ferrie, P.J., Griffith, L.E., Measuring quality of life in asthma. American Review of Respiratory Disease 147: 832-838, 1993.

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

Kobelt, G., Kirchberger, I., Malone-Lee, J., Quality of life aspects of the overactive bladder and the effect of treatment with tolterodine. British Journal of Urology 83:583–90, 1999.

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.

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

The MVH Group. The Measurement and Valuation of Health; First report on the main survey, May 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., Tsuchiya, A., Hernandez, M., Mapping between preference-based measures of health via a common yardstick. Health Economics and Decision Science Discussion Paper, forthcoming 2009.

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.

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

Salomon, J.A., Using ordinal data to estimate cardinal valuations. In: Measuring and Valuing Health Benefits for Economic Evaluation, Brazier, J., Ratcliffe, J., Salomon, J.A., and Tsuchiya, A. (Eds). Oxford University Press: Oxford, 2007.

Stevens, K., Working with children to develop dimensions for a preference based generic, paediatric, health related quality of life measure. Qualitative Health Research, forthcoming 2009.

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

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, Health Economics and Decision Science Discussion Paper, 2002.

Tsuchiya, A., Brazier, J., Roberts, J., Comparison of valuation methods used to generate the EQ-5D and SF-6D value sets. Journal of Health Economics 25:334-346, 2006.

Yang, Y., Brazier, J., Tsuchiya, A., Coyne, K., Estimating a Preference-Based Single Index from the Overactive Bladder Questionnaire. Value in Health 12(1): 159-166, 2008.

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 Science Discussion Paper, 2007.

Young T, Yang Y, Brazier J, Tsuchiya A The use of Rasch analysis as a tool in the construction of a preference based measure: the case of AQLQ. Health Economics and Decision Science Discussion Paper, 2007.

Ware, J.E., Snow, K.K., Kolinski, M., Gandeck, B., SF-36 Health survey manual and interpretation guide. The Health Institute, New England Medical Centre, Boston, MA, 1993.

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	Young 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 et al. (2008)	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 Characteristics of respondents

	Included respondents (n=499)	South Yorkshire[2]	England
Mean age (s.d.)	48.5(17.9)	-	-
Female	49.9%	51.2%	51.3%
Married/Partner	69.1%	-	-
Employed or self-employed	49.1%	56.1%	60.9%
Unemployed	1.4%	4.1%	3.4%
Long-term sick	3.6%	7.7%	5.3%
Full-time student	4.2%	7.5%	7.3%
Retired	30.5%	14.4%	13.5%
Own home outright or with a mortgage	86.1%	64.0%	68.7%
Renting property	13.8%	36.0%	31.3%
Secondary school is highest level of education	43.9%	-	-
EQ-5D score (s.d.)	0.86(0.23)	-	-
Found valuation task difficult (judged by respondent)	29.3%	-	-
Doubtful whether the respondent understood the tasks (judged by interviewer)	4.1%	-	-

[2] Statistics for South Yorkshire Health Authority and for England in the Census 2001. Questions used in this study and the census are not identical. The census includes persons aged 16 and above whereas this study only surveys persons aged 18 and above.

Table 3 Duration, completion, difficulty, understanding and effort by card bloc

	Card bloc			
	Most difficult blocs		Easiest blocs	
Instruments	AQL-5D OPUS ICECAP	SF-6D ICECAP AQL-5D	HUI2 OPUS EQ-5D	AQL-5D HUI2 ICECAP
N	24	25	27	25
Duration (minutes)	34	41	32	32
Number of missing VAS responses	0(0%)	1(0.2%)	33(5.1%)	5(0.8 %)
Number of missing rank responses	0(0%)	0(0%)	32(4.9%)	1(0.2%)
Spearman rank correlation coefficient for rank and implied rank from VAS values	0.959	0.967	0.964	0.989
Difficulty of questions, self-reported	n	n	n	n
Very difficult	3(12.5%)	3(12.0%)	0(0%)	0(0%)
Quite difficult	8(33.3%)	4(16.0%)	4(14.8%)	4(16.0%)
Neither	8(33.3%)	4(16.0%)	4(14.8%)	10(40.0%)
Fairly easy	3(12.5%)	13(52.0%)	13(48.1%)	9(36.0%)
Very easy	2(8.3%)	1(4.0%)	4(14.8%)	1(4.0%)
Understanding, reported by interviewer				
Understood and performed exercises easily	17(70.8%)	15(60.0%)	23(85.2%)	16(64.0%)
Some problems but seemed to understand in the end	7(29.2%)	8(32.0%)	1(3.7%)	8(32.0%)
Doubtful whether the respondent understood the exercises	0(%)	2(8.0%)	2(7.4%)	0(0%)
Effort and concentration, reported by interviewer				
Concentrated very hard	15(62.5%)	16(64.0%)	19(70.4%)	12(48.0%)
Concentrated fairly hard	9(37.5%)	6(24.0%)	5(18.5%)	11(44.0%)
Didn't concentrate very hard	0 (0%)	1(4.0%)	1(3.7%)	0 (0%)
Concentrated at start but lost interest towards end	0(0%)	2(8.0%)	1(3.7%)	1(4.0%)

Table 4 Adjusted VAS score and mean rank of worst and highest states by instrument

Instrument	Highest state valued			Worst state		
	Health state	Mean VAS (s.d.)	Mean rank (s.d.)	Health state	Mean VAS (s.d.)	Mean rank (s.d.)
EQ-5D	11111	0.909 (0.159)	2.12 (1.02)	33333	0.075 (0.228)	6.75 (1.03)
SF-6D	211111	0.860 (0.133)	2.37 (0.92)	645655	0.266 (0.250)	5.63 (1.22)
HUI2	112222	0.706 (0.210)	2.71 (0.97)	455445	0.077 (0.229)	6.87 (0.94)
AQL-5D	13321	0.717 (0.205)	3.14 (1.21)	55555	0.324 (0.240)	5.39 (1.23)
ICECAP	12321	0.872 (0.114)	2.26 (0.46)	44444	0.227 (0.275)	5.76 (1.29)
OPUS	1111	0.899 (0.111)	2.05 (0.57)	3333	0.223(0.238)	5.75 (1.24)

Note: Rank scores are coded in order of preference from 1 to 8, allowing ties, where 1 is the most preferred state.

Table 5 Health states with observed raw and adjusted VAS scores and published value set

Health state description	Number of valuations	Mean adjusted VAS (s.d.)	Median (Interquartile range) for adjusted VAS	VAS published value set utility value (where available[3]	Mean rank (s.d.)	Median (Interquartile range) rank
EQ-5D						
11111 (best state)	74	0.909(0.159)	0.95(0.900-1.000)	1	2.12(1.02)	2(2-2)
11322	50	0.709(0.200)	0.70(0.647-0.861)	0.403	2.70(0.97)	2(2-3)
12311	76	0.676(0.213)	0.70(0.540-0.848)	0.457	2.95(1.24)	3(2-3)
21113	75	0.604(0.223)	0.67(0.500-0.750)	0.435	3.48(1.17)	3(3-4)
13211	74	0.604(0.234)	0.60(0.498-0.800)	0.455	3.24(1.10)	3(2-4)
11223	75	0.577(0.238)	0.60(0.433-0.750)	0.392	3.71(1.23)	4(3-4)
22212	98	0.570(0.231)	0.60(0.400-0.750)	0.587	3.19(1.16)	3(2-4)
21331	50	0.540(0.212)	0.59(0.374-0.700)	0.308	3.98(1.12)	4(3-5)
23121	75	0.492(0.248)	0.50(0.330-0.688)	0.330	3.72(1.02)	4(3-4)
13132	76	0.438(0.299)	0.40(0.250-0.619)	0.251	4.32(1.22)	4(3-5)
12133	47	0.381(0.384)	0.37(0.211-0.612)	0.243	4.23(1.37)	4(3-5)
31112	47	0.392(0.261)	0.40(0.200-0.600)	0.385	4.47(1.46)	4(3-6)
31231	50	0.347(0.224)	0.30(0.166-0.493)	0.247	4.60(1.31)	4(4-6)
32121	48	0.288(0.264)	0.25(0.150-0.450)	0.272	5.29(1.41)	6(5-6)
33313	75	0.197(0.235)	0.15(0.053-0.300)	0.099	5.33(1.15)	6(5-6)
33333 (worst state)	496	0.075(0.227)	0.05(0.000-0.150)	-0.072	6.75(1.03)	7(6-7)
SF-6D						
211111	99	0. 860(0.133)	0.90(0.847-0.947)		2.37(0.92)	2(2-2)
211211	76	0.777(0.182)	0.85(0.697-0.900)		2.49(1.10)	2(2-2)
112221	74	0.754(0.204)	0.80(0.675-0.900)		2.55(1.04)	2(2-3)
111453	99	0.666(0.227)	0.75(0.526-0.850)		2.94(1.16)	3(2-3)
214411	76	0.652(0.222)	0.70(0.500-0.830)		3.04(1.33)	3(2-4)

[3] EQ-5D VAS values quoted in ‘The Measurement and Valuation of Health; First report on the main survey’, The MVH Group, May 1994. Values used are the VAS tariff of means: Whole population – 10 year duration.

Health state description	Number of valuations	Mean adjusted VAS (s.d.)	Median (Interquartile range) for adjusted VAS	VAS published value set utility value (where available[3]	Mean rank (s.d.)	Median (Interquartile range) rank
623133	76	0.540(0.224)	0.58(0.400-0.700)		3.97(1.26)	4(3-5)
424421	74	0.532(0.241)	0.55(0.350-0.730)		3.38(1.26)	3(2-4)
311655	99	0.490(0.237)	0.48(0.300-0.700)		4.25(0.99)	4(4-5)
545622	74	0.484(0.234)	0.50(0.300-0.650)		4.18(1.22)	4(4-5)
422655	76	0.411(0.267)	0.40(0.239-0.650)		4.76(1.24)	5(4-6)
624343	74	0.411(0.195)	0.40(0.250-0.575)		4.22(1. 22)	4(3-5)
535645	99	0.340(0.242)	0.33(0.150-0.500)		4.90(1.21)	5(4-6)
645655 (worst state)	498	0.266(0.250)	0.25(0.1()00-0.400)		5.63(1.22)	6(5-7)
AQL-5D						
13321	72	0.717(0.205)	0.79(0.590-0.876)		3.14(1.21)	3(2-3.75)
21223	98	0.701(0.205)	0.76(0.598-0.850)		3.02(1.25)	3(2-3)
53411	72	0.647(0.212)	0.70(0.500-0.830)		3.13(1.28)	3(2-4)
32441	76	0.619(0.197)	0.63(0.500-0.790)		3.35(1.27)	3(2-4)
12543	100	0.608(0.203)	0.60(0.500-0.750)		3.16(1.10)	3(2-4)
45143	76	0.551(0.200)	0.55(0.444-0.700)		3.59(1.12)	4(3-4)
23534	76	0.550(0.205)	0.60(0.400-0.700)		3.67(1.12)	3(3-4)
52314	100	0.515(0.226)	0.52(0.350-0.689)		3.90(1.32)	4(3-5)
15355	76	0.510(0.205)	0.50(0.365-0.600)		4.37(1.26)	5(3-5)
34254	72	0.424(0.324)	0.40(0.261-0.637)		4.07(1.09)	4(3-5)
55424	98	0.432(0.210)	0.43(0.300-0.586)		4.67(1.13)	5(4-6)
34554	72	0.403(0.274)	0.40(0.250-0.600)		4.69(1. 12)	5(4-5.75)
55555 (worst state)	494	0.324(0.240)	0.30(0.158-0.480)		5.39(1.23)	5(5-6)
HUI2						
112222	75	0.706(0.210)	0.75(0.600-0.856)		2.71(0.97)	2(2-3)
121132	76	0.661(0.194)	0.69(0.550-0.800)		3.20(1.06)	3(2-4)

Health state description	Number of valuations	Mean adjusted VAS (s.d.)	Median (Interquartile range) for adjusted VAS	VAS published value set utility value (where available[3]	Mean rank (s.d.)	Median (Interquartile range) rank
112123	99	0.639(0.235)	0.69(0.500-0.849)		3.19(1.31)	3(2-4)
331131	98	0.499(0.375)	0.50(0.327-0.713)		3.85(1.25)	4(3-5)
323331	76	0.490(0.224)	0.50(0.376-0.650)		3.76(1.20)	4(3-5)
314431	75	0.442(0.267)	0.45(0.246-0.622)		4.52(1.36)	5(3-6)
234111	74	0.428(0.192)	0.40(0.308-0.533)		4.18(1.34)	4(3-5)
344222	76	0.407(0.244)	0.40(0.250-0.550)		4.45(1.27)	4(4-6)
133444	98	0.350(0.355)	0.30(0.150-0.600)		4.50(1.38)	4.5(4-6)
144325	75	0.369(0.296)	0.30(0.150-0.621)		4.79(1.45)	5(4-6)
125425	77	0.323(0.269)	0.30(0.119-0.495)		5.13(1.22)	5(4-6)
445234	98	0.155(0.223)	0.15(0.037-0.275)		5.98(0.79)	6(6-6)
455445 (worst state)	501	0.077(0.229)	0.05(0.000-0.150)		6.87(0.94)	7(7-7)
ICECAP						
12321	100	0.872(0.114)	0.90(0.850-0.950)		2.26(0.46)	2(2-2.75)
21131	73	0.857(0.149)	0.90(0.817-0.950)		2.12(0.47)	2(2-2)
31212	72	0.805(0.151)	0.85(0.709-0.900)		2.21(0.60)	2(2-2)
22242	72	0.735(0.217)	0.80(0.650-0.900)		2.86(1.35)	2(2-3)
23324	100	0.682(0.217)	0.75(0.513-0.879)		2.82(1.06)	3(2-3)
33333	72	0.567(0.290)	0.62(0.400-0.780)		3.35(1.24)	3(2-4)
14344	100	0.569(0.261)	0.58(0.350-0.800)		3.46(1.09)	3(3-4)
43111	72	0.538(0.227)	0.55(0.400-0.706)		3.40(1.33)	3(2-4)
44143	73	0.398(0.275)	0.44(0.211-0.553)		4.33(1.08)	4(3-5)
43443	72	0.386(0.217)	0.36(0.250-0.600)		4.65(1.06)	5(4-6)
43334	72	0.352(0.245)	0.30(0.190-0.500)		4.96(1.37)	5(4-6)
42444	99	0.343(0.258)	0.30(0.150-0.550)		4.60(1.05)	4(4-5)
44444 (worst state)	490	0.226(0.275)	0.16(0.052-0.400)		5.76(1.29)	6(5-7)
OPUS						
1111	74	0.899(0.111)	0.94(0.891-0.952)		2.05(0.57)	2(2-2)

Health state description	Number of valuations	Mean adjusted VAS (s.d.)	Median (Interquartile range) for adjusted VAS	VAS published value set utility value (where available[3]	Mean rank (s.d.)	Median (Interquartile range) rank
2121	74	0.663(0.213)	0.71(0.538-0.823)		2.78(1.02)	2.5(2-3)
3121	76	0.641(0.199)	0.65(0.500-0.800)		2.92(0.88)	3(2-3)
2212	99	0.596(0.224)	0.60(0.450-0.798)		3.04(1.21)	3(2-4)
3132	76	0.519(0.194)	0.50(0.350-0.690)		3.69(0.87)	4(3-4)
2123	74	0.451(0.300)	0.46(0.285-0.692)		3.78(1.56)	4(2-5)
1322	99	0.480(0.258)	0.50(0.248-0.700)		3.94(1. 43)	4(3-5)
1233	99	0.451(0.244)	0.45(0.250-0.650)		3.97(1.12)	4(3-5)
3221	99	0.445(0.241)	0.40(0.260-0.645)		4.10(1.35)	4(3-5)
2331	74	0.409(0.253)	0.41(0.239-0.600)		4.39(1.21)	4(3-5)
3313	74	0.386(0.224)	0.40(0.219-0.550)		4.66(1.02)	5(4-5)
1333	74	0.318(0.329)	0.27(0.150-0.550)		4.64(1.53)	5(3.75-6)
3333 (worst state)	496	0.223(0.238)	0.20(0.071-0.350)		5.75(1.24)	6(5-7)
Best state	1488	0.999(0.010)	1.00(1.000-1.000)		1.00(0.08)	1(1-1)
Dead	1488	0.000(0.001)	0.00(0.000-0.000)		7.59(0.85)	8(7-8)

Figure 1 States for the ranking and VAS tasks for blocs 1-6 out of 20

BLOC 1 – A, B, C. Respondents, n=25
e.g. A=EQ-5D, B=SF-6D, C=HUI2.
A1 = EQ-5D state 11111, B1 = SF-6D state 645655

Task 1	Task 2	Task 3
A1	A1	B1
A2	A4	B4
A3	A5	B5
B1	C1	C1
B2	C2	C4
B3	C3	C5
Best state	Best state	Best state
Dead	Dead	Dead

BLOC 2 – D, E, F. n=25

Task 1	Task 2	Task 3
D1	D1	E1
D2	D4	E4
D3	D5	E5
E1	F1	F1
E2	F2	F4
E3	F3	F5
Best state	Best state	Best state
Dead	Dead	Dead

BLOC 3 – E, F, A. n=25

Task 1	Task 2	Task 3
E1	E1	F1
E6	E8	F8
E7	E9	F9
F1	A1	A1
F6	A6	A8
F7	A7	A9
Best state	Best state	Best state
Dead	Dead	Dead

BLOC 4 – B, C, D. n=25

Task 1	Task 2	Task 3
B1	B1	C1
B6	B8	C12
B7	B9	C13
C1	D1	D1
C10	D10	D12
C11	D11	D13
Best state	Best state	Best state
Dead	Dead	Dead

BLOC 5 – D, A, E. n=25

Task 1	Task 2	Task 3
D1	D1	A1
D6	D8	A12
D7	D9	A13
A1	E1	E1
A10	E6	E8
A11	E7	E9
Best state	Best state	Best state
Dead	Dead	Dead

BLOC 6 – F, B, C. n=25

Task 1	Task 2	Task 3
F1	F1	B1
F6	F8	B12
F7	F9	B13
B1	C1	C1
B10	C10	C12
B11	C11	C13
Best state	Best state	Best state
Dead	Dead	Dead

Figure 2 Comparison of adjusted VAS value and published value set

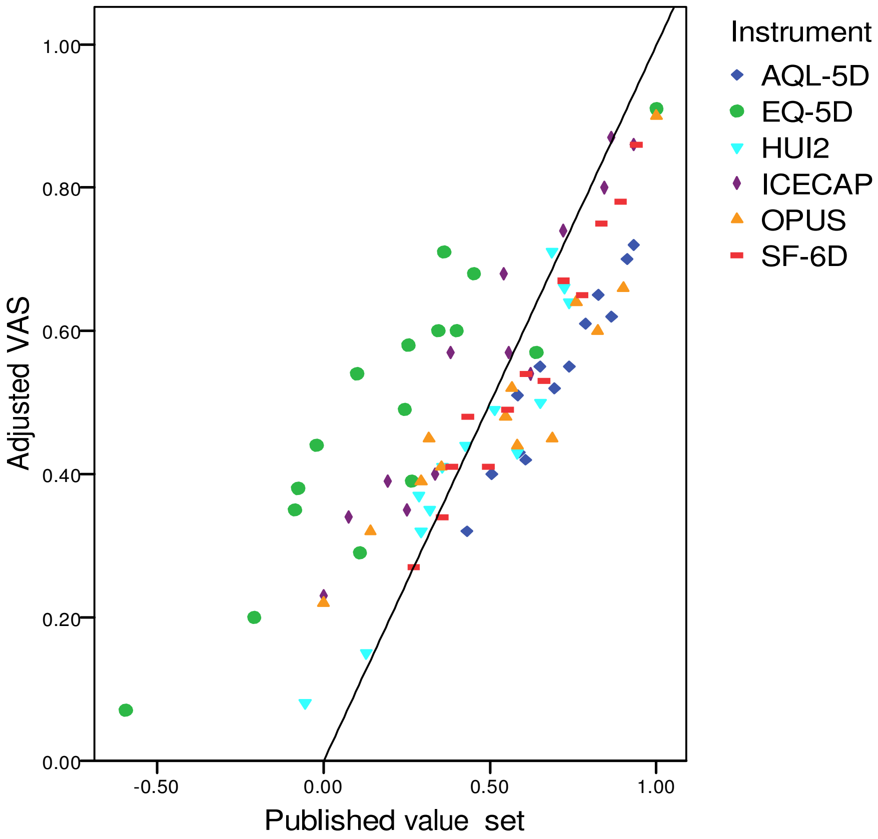


Figure 3 Comparison of rank value and published value set

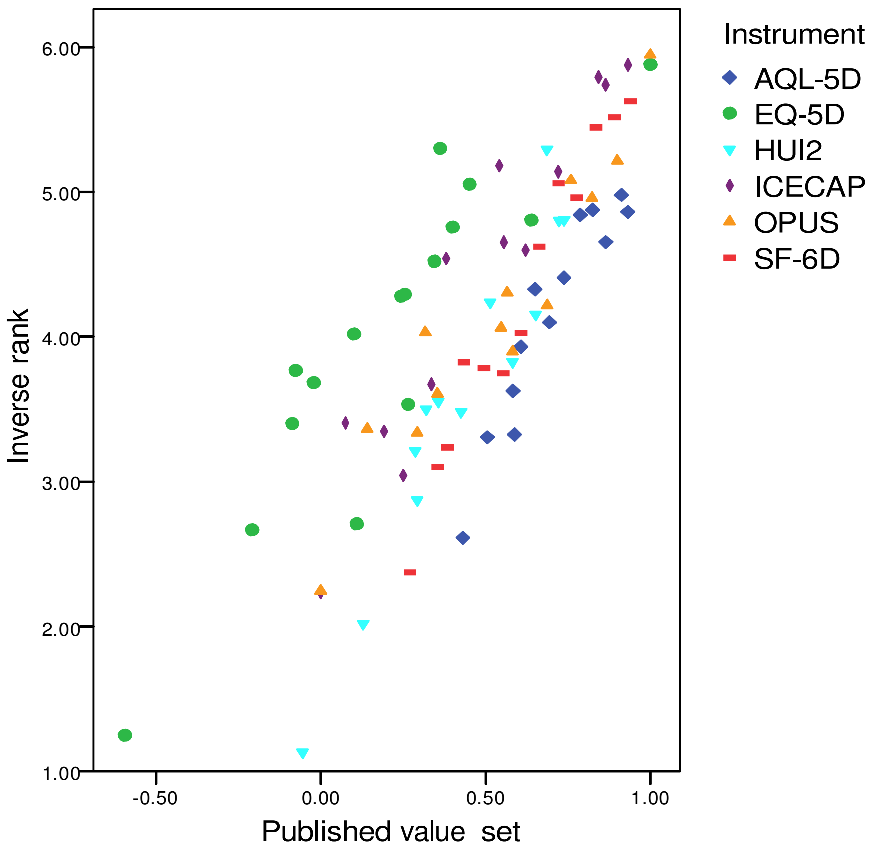
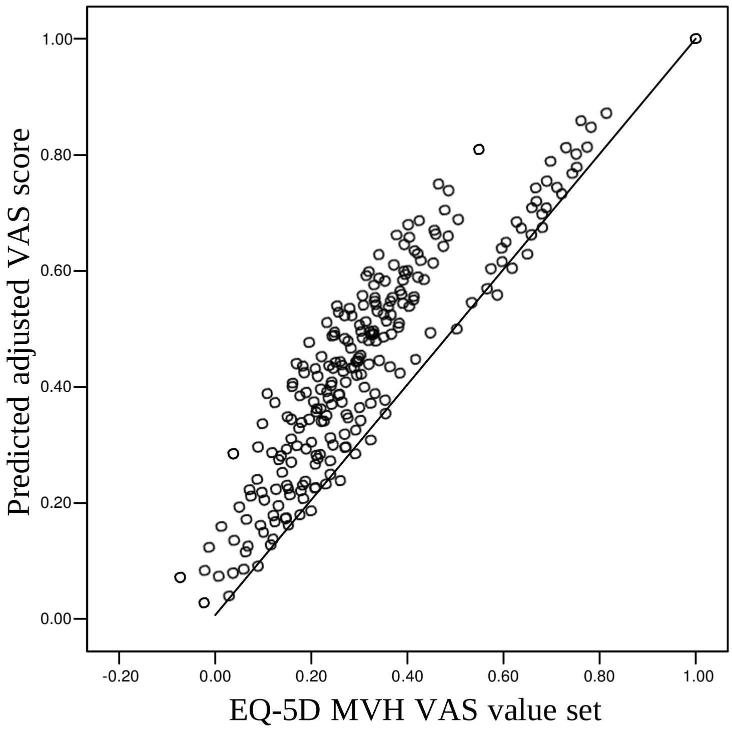


Figure 4 Comparison of EQ-5D VAS value set for all EQ-5D states and predicted adjusted VAS score using maximum likelihood random effects regression[4]



4 EQ-5D VAS values quoted in MVH, 1994. Values used are the VAS tariff of means: Whole population – 10 year duration.