

VALUING STATES FROM MULTIPLE MEASURES ON THE SAME VISUAL ANALOGUE SCALE: A FEASIBILITY STUDY

DONNA ROWEN^{a,*}, JOHN BRAZIER^a, AKI TSUCHIYA^{a,b} and MÓNICA HERNÁNDEZ ALAVA^a

^a*Health Economics and Decision Science, School of Health and Related Research (ScHARR), University of Sheffield, Sheffield, UK*

^b*Department of Economics, University of Sheffield, Sheffield, UK*

SUMMARY

Previous methods of empirical mapping involve using regressions on patient or general population self-reported data from datasets involving two or more measures. This approach relies on overlap in the descriptive systems of the measures and assumes it is appropriate to use different measures on the same population, which may not always be the case. This paper presents a feasibility study for a new approach to mapping between preference-based measures (PBM) using general population visual analogue scale (VAS) values as a common yardstick. We use data from a valuation study of 502 members of the UK general population, where, using ranking and VAS tasks, interviewees simultaneously valued health states defined by three of six PBM: EQ-5D (generic), SF-6D (generic), HUI2 (generic for children and adults), AQL-5D (asthma specific), OPUS (social care specific) and ICECAP (capabilities). Regression techniques are used to estimate the relationship between these VAS values and the original value set (i.e. 'tariff'). These results are subsequently used to estimate the relationship between all six PBM to enable 'value-based mapping' between measures. This new method of mapping potentially has a useful role in evidence synthesis and cross programme comparisons in studies using different measures. Copyright © 2011 John Wiley & Sons, Ltd.

Received 24 November 2009; Revised 19 October 2010; Accepted 18 March 2011

KEY WORDS: preference-based measures of health; quality of life; value-based mapping; visual analogue scale

1. INTRODUCTION

Health policy is increasingly being informed by economic evaluation that measures outcomes using the quality-adjusted life year (QALY). QALYs can be compared across different healthcare interventions or programmes to inform resource allocation. 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 (PBM) of health-related quality of life that reflect peoples' strength of preference for different outcomes and produce a valuation rather than simply a measurement of health. PBM of health-related quality of life generate a single index score with an upper anchor at full health (equal to 1) and a lower anchor at 0 (assuming it is equivalent to dead).

One major advantage of the QALY is that it produces a single measure that can be used to enable comparison across all types of healthcare interventions and programmes. Yet crucially, the quality adjustment weight of the QALY may be affected by which PBM is used. Ideally, all PBM should be comparable with each other, where the utility of a health state for a patient is identical regardless of the measure used, yet this is not true in practise. Many studies have illustrated that different generic PBM may assign different index scores to the same patients (see for example, Brazier *et al.*, 2004; Barton *et al.*, 2004; Espallargues *et al.*, 2005;

*Correspondence to: Health Economics and Decision Science, School of Health and Related Research (ScHARR), University of Sheffield, Regent Court, 30 Regent St, Sheffield, South Yorkshire S1 4DA, UK. Tel: +44114 222 0728. Fax: +44114 272 4095. E-mail: d.rowen@sheffield.ac.uk

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.

Recent years has seen the increasing proliferation of PBM of health. 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. PBM fall into three categories: generic measures for adults (Kaplan and Anderson, 1988; Dolan, 1997; Feeny *et al.*, 2002; Brazier *et al.*, 2002); measures for use in specific groups that are either adapted for use (for example, using simple language adjustment) in specific groups (Torrance *et al.*, 1996) or designed for those specific groups, such as older people (Grewal *et al.*, 2006) or children (Stevens, 2010); and condition-specific measures designed for specific medical conditions, such as those for asthma (Revicki *et al.*, 1998; Yang *et al.*, 2011).

Comparisons across studies using different PBM would be inaccurate if they assume that a QALY calculation is unaffected by the PBM used to generate the quality adjustment weight. One solution is to use one generic PBM in all economic evaluations (Dowie, 2002). For example, in the UK, the EQ-5D is currently recommended by NICE (2008) for use in economic evaluation. However, a generic measure is not always appropriate; for example, children require special consideration regarding language comprehension and development (Stevens, 2010), and there may be special considerations for very elderly people (Grewal *et al.*, 2006) and those needing social care (Ryan *et al.*, 2006). Also, 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). Furthermore, a generic measure is not always available even though it is appropriate. Mapping by statistical association is one option recommended by NICE (2008) to estimate EQ-5D utility data when EQ-5D data is unavailable in (but appropriate for) the study dataset, as mapping uses the health-related quality-of-life measures included in the study dataset to estimate EQ-5D utility data.

Typically, mapping by statistical association has involved a separate dataset that contains respondents' self-reported scores for their own health using 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 PBM has been included in the study rather than to compare across different PBM (for example, Brazier *et al.*, 2010; 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 PBM was included in the original dataset. This method uses statistical association between self-reported scores for two or more 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. This is reflected in the large errors found in many mapping studies (Brazier *et al.*, 2010). The regression relationship may be affected by the estimation dataset used and hence may differ for different patient groups or conditions. This method assumes it is appropriate to use different measures 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 PBM across all patient groups and conditions.

Here, we propose a new method of mapping that relates the responses on one PBM to another using visual analogue scale (VAS) values as a common yardstick. This method of 'value-based mapping' uses general population values for states defined by the descriptive systems of different measures to provide the common yardstick for conversion between measures, thus preserving the advantages of the descriptive system of each measure. The relationship between different measures is determined directly by people's VAS values for different states and not by associations in self-reported scores.

This paper briefly outlines the valuation study used to obtain the general population values and reports on the next stage that involves using these VAS values as a common metric to map between measures. This mapping enables a score for a health state described by one measure to be converted onto the utility scale of an alternative measure. For example, value-based mapping produces an estimated EQ-5D utility value using an observed score from five PBM. This enables researchers and policymakers to use estimated

EQ-5D utility values to compare between studies and healthcare programmes even when EQ-5D was not included in the original study.

1.1. Measures of health and quality of life

The study involves six PBM of health and quality of life: EQ-5D (generic), SF-6D (generic), HUI2 (generic for children and adults), AQL-5D (asthma specific), OPUS (social care specific) and ICECAP (capabilities), all summarised in Table I. The choice of measures reflects a range of different types of measures that are currently in use or nearing use in the UK.

The EQ-5D is the most widely used generic PBM of health-related quality of life (Brooks, 1996). The SF-6D is a generic preference-based single-index measure of health that can be estimated using SF-36 and SF-12 data (Brazier *et al.*, 2002; Brazier and Roberts, 2004). The HUI2 is a generic PBM of health for children aged from approximately 10 years and adults (Torrance *et al.* 1996). The AQL-5D (Young *et al.*, 2011) is a condition-specific PBM of health for asthma derived from the Asthma Quality of Life Questionnaire (Juniper *et al.*, 1993). All of these measures produce utility scores anchored at 0 for dead and 1 for perfect health.

ICECAP is a PBM 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 (Coast *et al.*, 2008). OPUS is a preference-based social care outcome measure for older people (Ryan *et al.*, 2006). We use the utility index that excludes the safety dimension and is 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).

Table I. Measures of health and quality of life

Measure	Summary	Dimensions	Levels	Unique states	Reference	Valuation technique and reference for value set used here	Value set range
EQ-5D	Generic	Five dimensions: Mobility, self-care, usual activity, pain/discomfort and anxiety/depression	Three levels: no problems, some problems, extreme problems	243	Brooks (1996)	Time trade-off, Dolan (1997)	−0.594 to 1
SF-6D	Generic	Six dimensions: Physical functioning, role limitations, social functioning, pain, mental health, vitality	Between four and six levels, depends on the dimensions	18 000	Brazier <i>et al.</i> (2002)	Standard gamble, Brazier and Roberts (2004)	0.271 to 1
HUI2	Generic for children and adults	Seven dimensions: Sensation, mobility, emotion, cognition, self-care, pain, fertility	Four or five levels, depends on the dimensions	8000	Torrance <i>et al.</i> (1996)	VAS mapped to standard gamble, McCabe <i>et al.</i> (2005)	−0.0552 to 1
AQL-5D	Condition specific for asthma	Five dimensions: Concern about asthma, shortness of breath, weather and pollution stimuli, sleep impact and activity limitations	Five levels: no problems to extreme problems	3125	Young <i>et al.</i> (2011)	Time trade-off, Yang <i>et al.</i> (2011)	0.431 to 1
ICECAP	Capability measure for older people in UK	Five dimensions: Attachment, security, role, enjoyment, control	Four levels: all, many, a little, none	1024	Grewal <i>et al.</i> (2006)	Best-worst scaling, Coast <i>et al.</i> (2008)	0 to 1
OPUS	Social care outcome measure for older people	Five dimensions: Food and nutrition, personal care, safety, social participation, control over daily living. Here the safety dimension is excluded.	Three levels: no unmet needs, low unmet needs, high unmet needs	243	Ryan <i>et al.</i> (2006)	Discrete choice experiment, Ryan <i>et al.</i> (2006)	0 to 1

VAS, visual analogue scale.

2. METHODS

This study has two stages. The first stage is a valuation study to determine how people value health and well-being states described by the six different descriptive systems. The second stage is to estimate relationships (i) between the VAS values for the sample of states and the published value set for each measure; and (ii) across all measures to enable value-based mapping between measures. Value-based mapping, for example, allows a value on the EQ-5D utility scale to be estimated using an observed SF-6D utility value. The relationship between the two measures is via common VAS values for states described by the two descriptive systems. Therefore, a unique valuation study was conducted to determine how people value health and well-being states described by different descriptive systems. The aim of this study was different to the usual valuation study; here, we aim to determine the relationship between measures. We do not need to produce or reproduce the entire value set for any of the measures; these common VAS values were instead used to estimate the mapping relationship between the EQ-5D and other measures. The valuation study is described here briefly, for further details see Rowen *et al.*, 2009b.

2.1. Valuation study

A valuation study was conducted, where each interview involves the valuation of health states defined by multiple PBM using three ranking and VAS tasks. The VAS task asks respondents 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'. There are concerns with the use of VAS including end point bias (Torrance *et al.*, 2001) and a concern that it does not generate 'preferences' (Brazier *et al.*, 2003). However, arguably these problems are less relevant for this study as VAS is used here to convert between measures rather than to produce a value set.

Each interview involves three rank tasks and three VAS tasks. In each task, the respondents were asked to rank eight different states: 'best state', 'dead', three states from one measure (one mild state, one moderate state, plus the worst possible state), and three states from another measure (one mild state, one moderate state, plus the worst possible state). The respondents were then asked to rate these same states using VAS and were allowed to change the implied ordering from the rank task. Table II shows an example set of states for one ranking and rating task. Each rank and subsequent VAS task contains states defined using two different measures, and each interview involves three different measures. Each interview involves 15 different states in total, five states each from three different measures, plus 'best state' and 'dead'.

Thirteen health and well-being states were included in the study for each measure except EQ-5D. States were selected to reflect a range of health state utility values according to the published value set (or most recent version for AQL-5D and ICECAP) while guaranteeing a variety of levels of severity for each dimension. Given that EQ-5D has a small descriptive system, 16 states for the EQ-5D were selected using an orthogonal design in SPSS (SPSS, Inc., Chicago, IL, USA). This will allow the estimation of a VAS value set covering all 243 states defined by EQ-5D (analysed elsewhere, see Rowen *et al.*, 2009b).

The study design contains 20 variations of combinations of states used in the ranking and VAS tasks. Each variation has a unique combination of the three measures. The number of times a measure appears in the first, second and third tasks is equal for each measure. For each variation, the ordering of the tasks (1, 2 and 3) remains the same across all respondents. There is a concern that this may lead to ordering effects, but this was chosen to reduce the complexity of the interviews for the interviewers. The study is designed so that each measure appeared an equal number of times with each other measure in rank and VAS tasks, and each health state is valued approximately 75–100 times (with the exception of 50 times for four EQ-5D states) and 500 times for each worst state. Analysis of the rank data is reported elsewhere (Hernandez *et al.*, 2009).

Rescaled VAS ratings are obtained using the following equation (MVH group, 1994):

$$A_{ijk} = \frac{R_{ijk} - R(\text{dead})_{ik}}{R(\text{best})_{ik} - R(\text{dead})_{ik}}$$

Table II. Example set of states for one ranking and rating task

I am confined to bed
I am <u>unable</u> to wash or dress myself
I am <u>unable</u> to perform my usual activities
I have <u>extreme</u> pain or discomfort
I am <u>extremely</u> anxious or depressed
I have <u>no</u> problems in walking about
I have <u>no</u> problems with self-care
I have <u>no</u> problems with performing my usual activities
I have <u>no</u> pain or discomfort
I am <u>not</u> anxious or depressed
I am confined to bed
I am <u>unable</u> to wash or dress myself
I am <u>unable</u> to perform my usual activities
I have <u>no</u> pain or discomfort
I am <u>extremely</u> anxious or depressed
I cannot have <u>any</u> of the love and friendship that I want
I can only think about the future with a <u>lot</u> of concern
I am unable to do <u>any</u> of the things that make me feel valued
I cannot have <u>any</u> of the enjoyment and pleasure that I want
I am <u>unable</u> to be at all independent
I can have <u>all</u> of the love and friendship that I want
I can think about the future with <u>only</u> a little concern
I am able to do a <u>few</u> of the things that make me feel valued
I can have a <u>lot</u> of the enjoyment and pleasure that I want
I am able to be <u>completely</u> independent
I cannot have <u>any</u> of the love and friendship that I want
I can think about the future with <u>only</u> a little concern
I am unable to do <u>any</u> of the things that make me feel valued
I cannot have <u>any</u> of the enjoyment and pleasure that I want
I am <u>unable</u> to be at all independent
Best state
Dead

where A_{ijk} represents the adjusted VAS rating for each health state j , for each individual i and for each task k , $R(\text{dead})_{ik}$ represents the raw rating given to ‘dead’, R_{ijk} represents the raw rating given to health state j , and $R(\text{best})_{ik}$ 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.

Descriptive statistics of adjusted VAS values are reported elsewhere (Rowen *et al.*, 2009b). For the EQ-5D, predicted adjusted VAS score (obtained using regression analysis) was compared with the published VAS value set. For all other analyses, mean VAS values for health states were used rather than predicted values to minimise error. We modelled the VAS values 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.

2.2. Mapping visual analogue scale values onto published value sets

Mapping functions were estimated using regression techniques that estimate for each measure the relationship between rescaled VAS values from our valuation study and the original value set or ‘utility scale’ (i.e. tariff). The independent variables for the mapping regressions are taken from the original value set and use the anchors specified by that measure. Each regression includes states from one measure only.

Three models were estimated using mean level data for each measure: (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 utility from the original value set, $j = 1, 2, \dots, J$ represents the health states per measure (16 for EQ-5D and 13 for all other measures), and ε_j represents the error term. Squared and cubic terms are designed to pick up nonlinearities in the relationship between mean VAS values and the original value set for each measure as there is no reason why the relationship is necessarily linear. Ordinary least squares was used for all models as the study design means that the VAS data was rescaled to a maximum value of one but was not censored at one.

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

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

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

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

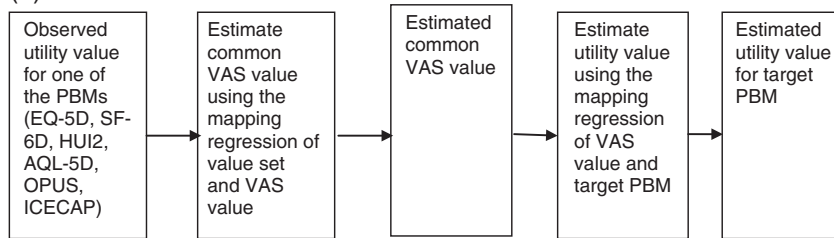
where y_{ij} represents rescaled VAS value for individual, $i = 1, 2, \dots, n$ for health state, $j = 1, 2, \dots, J$ for each measure, x represents the utility derived using the value set, and ε_{ij} represents the error term.

R^2 is often used in health economics to assess the goodness of fit of a model, but if extra variables are included in the regression, R^2 will never decrease even if the additional variables are not significant. Several alternative criteria incorporate a penalty for model complexity, for example, adjusted R^2 , Akaike's Information Criterion (AIC) (Akaike, 1973) and the Schwarz Bayesian Information Criterion (BIC) (Schwarz, 1978). AIC and BIC are reported for all models for consistency, although are not directly comparable across the models that use mean level data and individual level data. Models with lower AIC and BIC are preferred, but these criteria will not necessarily choose the same model. The penalty for the inclusion of additional variables in BIC is larger than in AIC, and therefore BIC tends to choose models with less variables than AIC. Mean squared error (MSE) of state predictions is a good indicator for mapping regressions of how large the prediction errors are. This is a useful indicator given that the purpose of the mapping function is to predict VAS values using observed utility values of one of the measures in external trial datasets. Performance of all regression models is reported using R^2 , MSE of state predictions, AIC and BIC.

2.3. Value-based mapping across all measures

Value-based mapping using common VAS values to convert between measures involves two stages: (1) the tariff values of the start measure is mapped onto the common VAS metric, and (2) the common VAS metric is mapped onto the tariff values of the target measure. Figure 1 illustrates the process, where, as an example, the SF-6D state 444544 (with utility value 0.5 on the SF-6D utility scale) is mapped onto the EQ-5D utility scale. First, the regression equation for SF-6D is used to convert the utility value of SF-6D for this state (0.5) into the common VAS metric for the state (0.464). Then second, based on the EQ-5D regression equation, this VAS value is used to estimate the corresponding utility value for SF-6D state 444544 on the EQ-5D utility scale (0.152). If the two regression equations are solved as simultaneous equations, then a single mapping function can be derived that goes straight from a value on the utility scale of the start measure to the corresponding value on the utility scale of the target measure, without stopping for the common VAS value. These mapping functions provide a method of translating one set of existing value sets into another, so that

(a) Procedure



(b) Example mapping of SF-6D onto EQ-5D

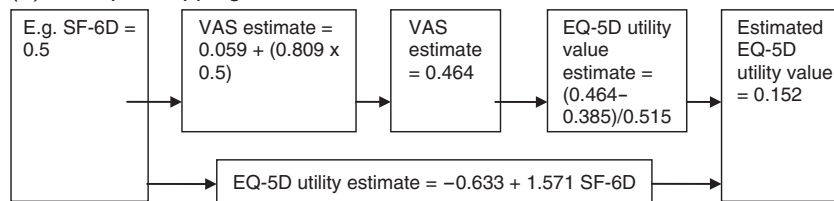


Figure 1. Value-based mapping across measures. PBM, preference-based measures. VAS, visual analogue scale

the values on the common VAS themselves are not the ultimate objective of the exercise. Also note that the variables involved in the regressions are the tariff values and the common VAS values; they do not involve the dimensions and levels of either descriptive system.

3. THE DATA

The dataset contains 502 successfully conducted interviews, a response rate of 55% for suitable respondents answering their door at the time of interview. The respondents were from the geographical areas in the north of England including urban and rural areas with a mix of socioeconomic characteristics. The study has a completion rate of 99% for all states included in the VAS tasks (178 VAS values missing out of 12 048 values). Two respondents (0.4%) had no VAS response, one respondent and one further task for one respondent were excluded for unusable responses ('dead' is valued higher than all states other than 'best state'). All other responses were used in the analysis reported here. Characteristics of all respondents with usable VAS responses are presented in Table III and compared with the general population in South Yorkshire and England. The sample had a higher proportion of retired persons and home owners and a lower proportion of employed persons.

Mean adjusted VAS value for the best states for those measures where their best state was valued, EQ-5D (state 11111) and OPUS (state 1111) are lower than one (0.909 and 0.899, respectively), and these measure-specific best states are regarded as statistically significantly worse than the generic best state. The worst state (known as 'PITS') was valued for all measures, and the VAS values range between 0.075 and 0.324. This range reflects the different 'floors' of the measures, with generic EQ-5D having the lowest floor and the disease-specific AQL-5D having the highest. There are no logical inconsistencies in the mean VAS scores or value set for these states for each instrument.

4. RESULTS

4.1. Mapping visual analogue scale values onto published value sets

Tables III and IV present the regression results mapping common VAS metric values onto the utility scale for each measure using mean and individual level data, respectively. Performance of the regression models varies

Table III. Characteristics of respondents

	Included respondents (<i>n</i> = 499)	South Yorkshire ¹	England
Mean age (SD)	48.5 (17.9)	–	–
Female	49.9%	51.2%	51.3%
Married/ with 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 (SD)	0.85 (0.23)	–	0.86 (0.23) ²
Found valuation task difficult (judged by respondent)	29.3%	–	–
Doubtful whether the respondent understood the tasks (judged by interviewer)	4.1%	–	–

¹Statistics for South Yorkshire Health Authority and for England in the Census 2001. Questions were used in this study, and the census was not identical. The census includes persons aged 16 and older, whereas this study only surveys persons aged 18 and older.

²Interviews conducted in the Measurement and Valuation of Health study in 1993 (Kind *et al.*, 1999).

SD, standard deviation.

across measures with the poorest performance for OPUS and EQ-5D models. There is less variation within measures across models using R^2 , MSE of state predictions, AIC and BIC. R^2 is generally high, varying from 0.83 (EQ-5D) to 0.97 (SF-6D) for the mean models. Although their R^2 is generally lower, the linear models are preferred according to AIC and BIC, with the exception of OPUS.

The models selected by BIC also tend to be selected by AIC and in cases where this does not happen, AIC is likely affected by its smaller penalty for the inclusion of additional variables. The models selected by AIC and BIC do not have the smallest MSE, but the differences in MSE between the model selected by AIC and BIC and the model with the smallest MSE are small. The cubic regression is selected as a better-fitting model using AIC, BIC and the smallest MSE for OPUS models for both mean and individual level data. However, plots of the data, available from the corresponding author on request, clearly show that the cubic model for both mean and individual data is oversensitive to outliers and that the relationship is not cubic. Table IV shows that many of the quadratic and cubic regression models suffer from multicollinearity as evidenced by high R^2 values in regressions, where none or almost none of the coefficients are significant. For these reasons, the linear models are used for subsequent analysis.

Table V shows that regressions estimated using individual level data have similar MSE of state predictions (note, there is no R^2 for maximum likelihood estimation). The mean level models are used for subsequent analysis, but results using the individual level models are available from the corresponding author on request.

4.2. Mapping across all measures

Table VI presents the mapping functions that enable a start PBM to be mapped onto the target measure to produce an estimated utility value for that measure. For example, for an observed SF-6D utility value equal to 0.7, we obtain an estimated score on the EQ-5D utility scale using the equation:

$$\begin{aligned} \text{EQ-5D} &= -0.633 + 1.571 \text{ SF-6D} \\ \text{EQ-5D} &= -0.633 + 1.571 * 0.7 = 0.467 \end{aligned} \quad (7)$$

These mapping functions have been generated by solving the linear mapping regressions in Table IV as simultaneous equations (detailed workings available from the authors on request). The mapping functions vary

Table IV. Mapping visual analogue scale values to value set using mean level data

Model		x (SE.)	x^2 (SE)	x^3 (SE)	Constant (SE)	n	R^2	AIC	BIC	MSE (state level)
EQ-5D	(1)	0.515** (0.063)			0.385** (0.026)	16	0.83	-29.93	-28.39	0.00702
	(2)	0.531** (0.080)	-0.039 (0.116)		0.388** (0.028)	16	0.83	-28.08	-25.76	0.00696
	(3)	0.523** (0.119)	-0.055 (0.205)	0.026 (0.268)	0.390** (0.037)	16	0.83	-26.09	-23.00	0.00695
SF-6D	(1)	0.809** (0.054)			0.059 (0.035)	13	0.95	-44.87	-43.74	0.00136
	(2)	0.388 (0.357)	0.344 (0.288)		0.173 (0.101)	13	0.96	-44.60	-42.91	0.00119
	(3)	2.354 (1.619)	-3.13 (2.808)	1.905 (1.532)	-0.167 (0.291)	13	0.97	-44.66	-42.40	0.00102
HUI2	(1)	0.724** (0.068)			0.112** (0.034)	13	0.91	-35.55	-34.42	0.00279
	(2)	0.731* (0.233)	-0.009 (0.289)		0.112* (0.044)	13	0.91	-33.56	-31.86	0.00279
	(3)	0.927 (0.451)	-0.879 (1.717)	0.842 (1.636)	0.114* (0.046)	13	0.91	-31.93	-29.67	0.00271
AQL-5D	(1)	0.745** (0.060)			0.016 (0.043)	13	0.93	-50.06	-48.93	0.00091
	(2)	0.914 (0.611)	-0.122 (0.436)		-0.040 (0.207)	13	0.93	-48.17	-46.47	0.00091
	(3)	3.576 (4.458)	-4.159 (6.710)	1.976 (3.276)	-0.605 (0.961)	13	0.94	-46.68	-44.42	0.00087
OPUS	(1)	0.522** (0.070)			0.218** (0.042)	13	0.84	-29.41	-28.28	0.00448
	(2)	0.150 (0.235)	0.366 (0.222)		0.282** (0.055)	13	0.87	-30.52	-28.83	0.00353
	(3)	1.245** (0.363)	-2.45* (0.846)	1.871** (0.552)	0.209** (0.044)	13	0.94	-39.21	-36.95	0.00155
ICECAP	(1)	0.669** (0.057)			0.240** (0.032)	13	0.93	-34.18	-33.05	0.00311
	(2)	0.551* (0.224)	0.125 (0.227)		0.257** (0.046)	13	0.93	-32.56	-30.87	0.00301
	(3)	0.643 (0.546)	-0.136 (1.404)	0.186 (0.989)	0.251** (0.057)	13	0.93	-30.61	-28.35	0.00300

SE, standard error; AIC, Akaike's Information Criterion; BIC, Bayesian Information Criterion; MSE, mean squared error.

*significant at 5%;

**significant at 1%.

across measures, reflecting differences in measures and differences in the relationships between measures. Background characteristics are not included in the models because in preliminary data analysis, they were found not to be significant.

To demonstrate how these results can be used, we show how the observed values of all measures can be expressed on the EQ-5D utility scale (though this can be carried out for any instrument using the mapping functions in Table VI). Table VII is a look-up table presenting estimated values on the EQ-5D utility scale for observed values of other measures. For example, Table VII illustrates that SF-6D health state 444544 with utility value of 0.5 has an estimated value on the EQ-5D utility scale of 0.152 using the common VAS metric to convert between measures. Figure 1(b) illustrates how the estimate of the score on the EQ-5D utility scale is produced using the mapping regressions in Table IV and the mapping functions in Table VI. Estimates are outside of the EQ-5D value set range (1 to -0.594) for the highest ICECAP value (EQ-5D estimate of 1.017) and the lowest HUI2 value (EQ-5D estimate of -0.608).

Table V. Mapping visual analogue scale values to value set using individual level data

Model		x (SE)	x^2 (SE)	x^3 (SE)	Constant (SE)	n	Groups	AIC	BIC	MSE (state level)
EQ-5D	(4)	0.513** (0.011)			0.382** (0.011)	1482	249	-248.46	-227.25	0.00703
	(5)	0.515** (0.011)	-0.024 (0.023)		0.388** (0.012)	1482	249	-247.56	-221.06	0.00698
	(6)	0.48** (0.028)	-0.063 (0.037)	0.081 (0.061)	0.396** (0.013)	1482	249	-247.30	-215.50	0.00703
SF-6D	(4)	0.832** (0.020)			0.044** (0.015)	1487	249	-453.44	-432.22	0.00139
	(5)	0.660** (0.133)	0.152 (0.116)		0.084* (0.034)	1487	249	-453.16	-426.64	0.00131
	(6)	2.578** (0.667)	-3.416** (1.222)	2.02** (0.689)	-0.221* (0.109)	1487	249	-459.73	-427.91	0.00104
HUI2	(4)	0.711** (0.017)			0.116** (0.013)	1486	251	-220.32	-199.11	0.00281
	(5)	0.684** (0.052)	0.043 (0.079)		0.116** (0.013)	1486	251	-218.61	-192.09	0.00281
	(6)	0.956** (0.115)	-1.392* (0.547)	1.457** (0.550)	0.132** (0.014)	1486	251	-223.60	-191.78	0.00284
AQL-5D	(4)	0.764** (0.026)			-0.001 (0.020)	1474	248	-490.62	-469.44	0.00094
	(5)	0.91** (0.257)	-0.111 (0.195)		-0.045 (0.080)	1474	248	-488.95	-462.47	0.00092
	(6)	4.018 (2.198)	-4.963 (3.414)	2.427 (1.705)	-0.680 (0.453)	1474	248	-488.97	-457.20	0.00088
OPUS	(4)	0.509 (0.014)			0.225** (0.012)	1481	249	-374.62	-353.42	0.00450
	(5)	0.321 (0.048)	0.221** (0.054)		0.239** (0.013)	1481	249	-389.31	-362.81	0.00375
	(6)	1.275 (0.111)	-2.674** (0.310)	2.065** (0.218)	0.223** (0.013)	1481	249	-473.88	-442.08	0.00163
ICECAP	(4)	0.696** (0.015)			0.231** (0.012)	1460	244	-325.31	-304.17	0.00319
	(5)	0.712** (0.054)	-0.020 (0.063)		0.230** (0.013)	1460	244	-323.41	-296.98	0.00321
	(6)	0.794** (0.135)	-0.291 (0.416)	0.208 (0.315)	0.229** (0.013)	1460	244	-321.85	-290.13	0.00317

SE, standard error; AIC, Akaike's Information Criterion; BIC, Bayesian Information Criterion; MSE, mean squared error.

*significant at 5%;

**significant at 1%.

Figure 2 plots observed EQ-5D utility values for a patient dataset and estimated EQ-5D utility values using mapping by statistical association and value-based mapping using observed SF-6D utility values. The patient dataset consists of 2515 cases of self-reported SF-6D and EQ-5D scores created using seven patient group datasets (chronic obstructive pulmonary 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 (Brazier *et al.*, 2004). The EQ-5D estimates for mapping by statistical association are produced using a linear regression model estimated for this patient dataset using ordinary least squares (EQ-5D = $-0.343 + 1.490$ SF-6D (standard error of coefficients = 0.20, 0.30)).

For this dataset, covering seven different patient populations, the value-based mapping typically produces lower EQ-5D estimates than mapping by statistical association. MSE for the value-based mapping approach is higher (0.104) than for mapping by statistical association (0.048). This is expected, given that they are produced using very different approaches estimated using different datasets. Mapping by statistical association estimates the relationship between self-reported scores and is likely to produce reasonably accurate estimates

Table VI. Mapping functions to map between measures

Start measure (observed)	Target measure (predicted)				
	EQ-5D	SF-6D	HUI2	AQL-5D	OPUS
EQ-5D		SF-6D = 0.403 + 0.637 EQ-5D	HUI2 = 0.377 + 0.711 EQ-5D	AQL-5D = 0.495 + 0.691 EQ-5D	OPUS = 0.320 + 0.987 EQ-5D
SF-6D	EQ-5D = -0.633 + 1.571 SF-6D		HUI2 = -0.073 + 1.117 SF-6D	AQL-5D = 0.058 + 1.086 SF-6D	OPUS = -0.305 + 1.550 SF-6D
HUI2	EQ-5D = -0.530 + 1.406 HUI2	SF-6D = 0.066 + 0.895 HUI2		AQL-5D = 0.129 + 0.972 HUI2	OPUS = -0.203 + 1.387 HUI2
AQL-5D	EQ-5D = -0.717 + 1.447 AQL-5D	SF-6D = -0.053 + 0.921 AQL-5D	HUI2 = -0.133 + 1.029 AQL-5D		OPUS = -0.387 + 1.427 AQL-5D
OPUS	EQ-5D = -0.324 + 1.014 OPUS	SF-6D = 0.197 + 0.645 OPUS	HUI2 = 0.146 + 0.721 OPUS	AQL-5D = 0.271 + 0.701 OPUS	
ICECAP	EQ-5D = -0.282 + 1.299 ICECAP	SF-6D = 0.224 + 0.827 ICECAP	HUI2 = 0.177 + 0.924 ICECAP	AQL-5D = 0.301 + 0.898 ICECAP	OPUS = 0.042 + 1.282 ICECAP
					ICECAP = 0.217 + 0.770 EQ-5D
					ICECAP = -0.271 + 1.209 SF-6D
					ICECAP = -0.191 + 1.082 HUI2
					ICECAP = -0.335 + 1.114 AQL-5D
					ICECAP = -0.033 + 0.780 OPUS

Table VII. Value-based mapping look-up table: Mapping onto EQ-5D

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	1.017
0.9	0.781	0.9	0.735	0.9	0.59	0.9	0.588	0.9	0.888
0.8	0.624	0.8	0.595	0.8	0.44	0.8	0.487	0.8	0.758
0.7	0.467	0.7	0.454	0.7	0.30	0.7	0.385	0.7	0.628
0.6	0.310	0.6	0.313	0.6	0.15	0.6	0.284	0.6	0.498
0.5	0.152	0.5	0.173	0.5	0.01	0.5	0.183	0.5	0.368
0.4	-0.005	0.4	0.032	0.4	-0.14	0.4	0.081	0.4	0.238
0.3	-0.162	0.3	-0.108			0.3	-0.020	0.3	0.108
		0.2	-0.249			0.2	-0.122	0.2	-0.022
		0.1	-0.390			0.1	-0.223	0.1	-0.152
		0	-0.530			0	-0.324	0	-0.282

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

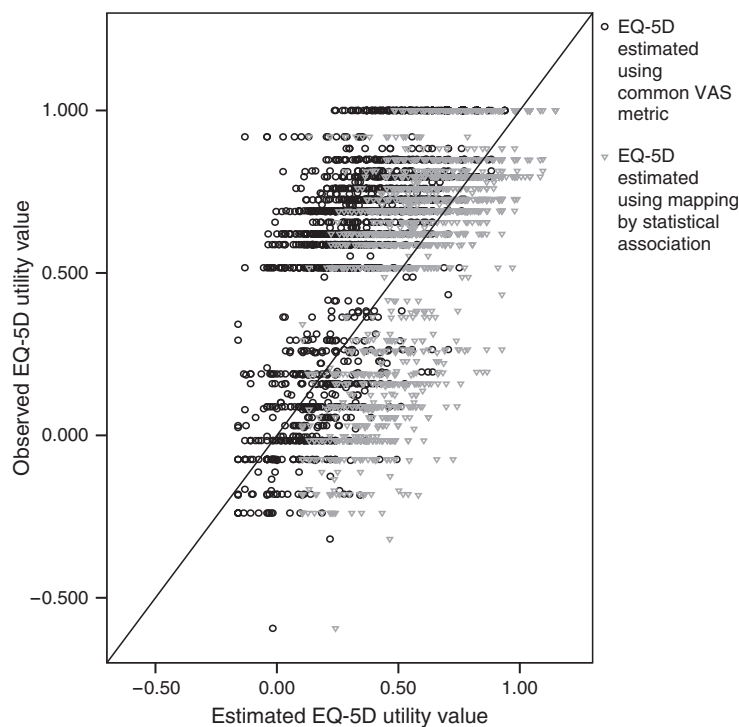


Figure 2. Comparison of observed EQ-5D utility values for a patient dataset and estimated EQ-5D utility values using observed SF-6D utility values. VAS, visual analogue scale

for a dataset of this nature, where both measures are appropriate. The direct statistical mapping approach requires reasonable conceptual overlap between all the dimensions in EQ-5D and SF-6D. Where dimensions are only represented in one measure, for example, vitality appears in SF-6D but not EQ-5D, the conventional methods fail to reflect this (for example, nonsignificant coefficients for the vitality dimension in mapping from SF-36 to EQ-5D (Rowen *et al.*, 2009a, 2009b)). Mapping each measure onto a third scale, in this case VAS, means that important dimensions like vitality in the source instrument are not excluded when represented on the target utility scale.

5. DISCUSSION

Not all clinical trials include PBM, but even when they do, they do not all use the same measure, presenting a challenge for evidence synthesis. Here, we undertake a feasibility study examining a new method of mapping that uses common VAS values to map five PBM onto EQ-5D. For example, suppose we wish to compare mean utility values of asthma patients using two different interventions from two different trials, where one trial used AQL-5D to generate utilities and the other used EQ-5D. We wish to estimate differences between each intervention and across the two trials, while controlling for the fact that the values were generated using different measures. Our mapping approach enables us to do this by converting mean observed AQL-5D utilities into equivalent EQ-5D utilities, but the equivalence is not in terms of the health state description but in terms of its impact on quality of life.

The common VAS values are obtained by asking respondents to value states from multiple instruments alongside each other. The VAS and mapping results indicate that this approach is feasible. The functions in Table VI that map across PBM indicate that the relationships vary across measures. We expect differences; the original value sets use different preference elicitation techniques including time trade-off, standard gamble, best–worst scaling and discrete choice experiment; the study populations are different (locations, year, sampling method); and here, all states are rated alongside other measures.

It is unclear whether the mean or individual level regressions are most appropriate for estimating the relationship between mean VAS values from the valuation study and the original utility scale for each measure. The mean level models ignore individual variability but pragmatically are in accordance with techniques used in health economics. For example, Torrance *et al.* (1996) and Stevens *et al.* (2006) used mean level regression models on fewer data points than used here to estimate the relationship between standard gamble and VAS responses. In principle, individual level models would be preferred to exploit a richer dataset. However, the individual level models suffer from the problem that the original value set of each instrument does not differ across individuals, that is, although the dependent variable is an individual level variable, the independent variables are mean level variables. This means that the independent variables are not ‘variables’ in the usual sense but are instead estimated parameters taken from previous valuation studies. Future research in this area should involve exploration of ways of addressing this problem. For the purposes of prediction, the main object of the present paper, the use of the mean level dependent variable will still produce unbiased point estimates.

The descriptive system of the EQ-5D covers a larger value set range (1 to –0.594) of health states than all other measures included in the study. Only the HUI2 appears to have a descriptive system that can be mapped to nearly the full EQ-5D range. The AQL-5D value set maps to only a small section of the full EQ-5D range (0.73 to –0.14). Some EQ-5D estimates are higher or lower than the EQ-5D value set range. This clearly highlights differences in the value set ranges of different measures and hence of health states covered by the different descriptive systems. This raises two issues: (i) whether the predictions in this case should be bounded by the range of the EQ-5D value set; and (ii) what it means if we have a predicted utility value above the best state or below the worst state defined by the classification system of any measure.

This study suffers from comparatively low sample size, meaning that the mapped estimates may be unreliable. Observed VAS values for EQ-5D best state (0.909) and OPUS best state (0.899) are similar, for example, yet the mapped OPUS best state has an estimate on the EQ-5D utility scale of 0.689. This may be the case because the mapping regressions are poorest for OPUS and EQ-5D models, with the highest MSE. The mapping functions may not be entirely reasonable at the upper ends of the scale, where there were only a small number of observations in the valuation study in comparison with the worst state defined by each measure, which has the highest number of observations. Further research is needed to examine differences in instrument specific full health and states at the mildest end of the descriptive system.

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. Mapping by statistical association uses self-reported scores on both measures to map between measures. This is only appropriate if each measure is suitable for that population. But here, we use general population VAS values for states defined by

both measures to map between measures. This will be appropriate even if the target measure is inappropriate for that population to self-report their health. One advantage of using this approach in the UK is that EQ-5D is currently recommended by NICE (2008) for use in economic evaluation, and NICE states that mapping can be used when EQ-5D was not included in the trial. This approach enables an EQ-5D utility estimate to be produced when the EQ-5D is either unavailable or unsuitable, provided the trial dataset includes one of the other PBM included here.

This feasibility study has demonstrated that it is possible to value multiple measures on the same VAS scale and to use these values to map between measures. This new approach of value-based mapping potentially has a useful role in evidence synthesis and cross programme comparisons in studies using different measures. However, important concerns remain with the statistical methods used to estimate the mapping functions, and particularly the estimation of uncertainty for mapped estimates needs to be addressed in future research in this area.

ACKNOWLEDGEMENTS

We would like to thank Professor Coast and Professor Netten for the use of the ICECAP and the OPUS, respectively. We would also like to thank all the interviewees who took part in the survey and the participants at the Health Economics Study Group Meeting in January 2009 and the EuroQol Plenary Meeting in September 2009 for their comments on an earlier version of this paper. This study was funded by MRC project grant R/112915. The authors have declared that there is no conflict of interest.

REFERENCES

- Akaike H. 1973. In *Information Theory and an Extension of the Maximum Likelihood Principle*, Petrov BN and Csaki F (eds.), *Second International Symposium on Information Theory*. Akademiai Kiado: Budapest, 267–281.
- Barton GR, Bankart J, Davis AC, Summerfield QA. 2004. Comparing utility scores before and after hearing aid provision: results According to the EQ-5D, HUI3 and SF-6D. *Applied Health Economics and Health Policy* **3**: 103–105.
- Brazier J, Greene C, McCabe C, Stevens K. 2003. Use of visual analog scales in economic evaluation. *Expert Review of Pharmacoeconomics Outcomes Research* **3**: 293–302.
- Brazier J, Roberts J. 2004. The estimation of a preference-based measure of health from the SF-12. *Medical Care* **42**: 851–859.
- Brazier J, Roberts J, Deverill M. 2002. The estimation of a preference-based measure of health from the SF-36. *Journal of Health Economics* **21**: 271–292.
- Brazier J, Roberts J, Tsuchiya A, Busschbach J. 2004. A comparison of the EQ-5D and SF-6D across seven patient groups. *Health Economics* **13**: 873–884.
- Brazier J, Yang Y, Tsuchiya A, Rowen D. 2010. A review of studies mapping (or cross walking) non-preference based measures of health to generic preference-based measures. *The European Journal of Health Economics* **11**: 215–225.
- Brooks R. 1996. EuroQol: the current state of play. *Health Policy* **37**: 53–72.
- Coast J, Flynn T, Natarajan L, Sprotson K, Lewis J, Louviere JL, Peters TJ. 2008. Valuing the ICECAP capability index for older people. *Social Science & Medicine* **67**: 874–882.
- Dolan P. 1997. Modeling valuations for EuroQol health states. *Medical Care* **35**: 1095–1108.
- Dowie J. 2002. Decision validity should determine whether a generic or condition-specific HRQOL measure is used in health care decisions. *Health Economics* **11**: 1–8.
- Espallargues M, Czoski-Murray C, Bansback N, Carlton J, Lewis G, Hughes L, Brand C, Brazier J. 2005. The impact of age related macular degeneration on health status utility values. *Investigative Ophthalmology & Visual Science* **46**: 4016–4023.
- Feeny D, Furlong W, Torrance GW, Goldsmith CH, Zhu Z, DePauw S, Denton M, Boyle M. 2002. Multiattribute and single-attribute utility functions for the Health Utilities Index Mark 3 system. *Medical Care* **40**: 113–128.
- Franks P, Lubetkin EI, Gold MR, Tancredi DJ, Jia H. 2004. Mapping the SF-12 to the EuroQol EQ-5D Index in a national US sample. *Medical Decision Making* **24**: 247–254.
- Gray AM, Rivero-Arias O, Clarke PM. 2006. Estimating the Association between SF-12 Responses and EQ-5D Utility Values by Response Mapping. *Medical Decision Making* **26**: 18–29.
- Grewal I, Lewis J, Flynn TN, Brown J, Bond J, Coast J. 2006. Developing attributes for a generic quality of life measure for older people: preferences or capabilities? *Social Science & Medicine* **62**: 1891–1901.
- Harper R, Brazier JE, Waterhouse JC, Walters SJ, Jones NMB, Howard P. 1997. Comparison of outcome measures for patients with chronic obstructive pulmonary disease (COPD) in an outpatient setting. *Thorax* **52**: 879–887.

- Hernandez M, Brazier J, Rowen D, Tsuchiya A. 2009. Valuing health states from different instruments using ranking. *Health Economics Study Group Discussion Paper*, July 2009.
- Juniper EF, Guyatt GH, Ferrie PJ, Griffith LE. 1993. Measuring quality of life in asthma. *The American Review of Respiratory Disease* **147**: 832–838.
- Kaplan RM, Anderson JP. 1988. A general health policy model: update and applications. *Health Services Research* **23**: 203–235.
- Kind P, Hardman G, Macran S. 1999. UK population norms for EQ-5D. *Centre for Health Economics Discussion Paper Series*, University of York.
- Kobelt G, Kirchberger I, Malone-Lee J. 1999. Quality-of-life aspects of the overactive bladder and the effect of treatment with tolterodine. *British Journal of Urology* **83**: 583–590.
- Longworth L, Bryan S. 2003. An empirical comparison of EQ-5D and SF-6D in liver transplant patients. *Health Economics* **12**: 1061–1067.
- McCabe C, Stevens K, Roberts J, Brazier J. 2005. Health state values for the HUI 2 descriptive system: results from a UK survey. *Health Economics* **14**: 231–244.
- MVH group. 1994. The measurement and valuation of health: first report on the main survey. *Report to the Department of Health*. Centre for Health Economics, University of York.
- NICE. 2008. *Guide to the methods of technology appraisal*. National Institute for Health and Clinical Excellence, London. (Available from: <http://www.nice.org.uk/aboutnice/howwework/devnicetech/technologyappraisalprocessguides/guidetothemethodsoftechnologyappraisal.jsp>) [7 July 2008].
- Nichol MB, Sengupta N, Globe DR. 2001. Evaluating quality adjusted life years: estimation of the Health Utility Index (HUI2) from the SF-36. *Medical Decision Making* **21**: 105–112.
- O'Brien BJ, Spath M, Blackhouse G, Severens JL, Brazier JE. 2003. A view from the bridge: agreement between the SF-6D utility algorithm and the Health Utilities Index. *Health Economics* **12**: 975–981.
- Revicki DA, Leidy NK, Brennan-Deimer F, Sorenson S, Togias A. 1998. Integrating patient preferences into health outcomes assessment: The multiattribute asthma symptom utility index. *Chest* **114**: 998–1007.
- Rowen D, Brazier J, Roberts J. 2009a. Mapping SF-36 onto the EQ-5D index: how reliable is the relationship? *Health and Quality of Life Outcomes* **7**: 27.
- Rowen D, Brazier J, Tsuchiya A, Hernandez M, Ibbotson R. 2009b. The simultaneous valuation of states from multiple instruments using ranking and VAS data: methods and preliminary results. *Health Economics and Decision Sciences Discussion Paper 09/06*, University of Sheffield.
- Ryan M, Netten A, Skatun D, Smith P. 2006. 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.
- Schwarz G. 1978. Estimating the dimension of a model. *Annals of Statistics* **6**: 461–464.
- Stevens KJ, McCabe CJ, Brazier JE. 2006. Mapping between Visual Analogue Scale and Standard Gamble data; results from the UK Health Utilities Index 2 valuation survey. *Health Economics* **15**: 527–533.
- Stevens KJ. 2010. Working with children to develop dimensions for a preference-based, generic, pediatric health-related quality-of-life measure. *Qualitative Health Research* **20**: 340–351.
- Torrance GW, Feeny DH, Furlong WJ, Barr RD, Zhang Y, Wang Q. 1996. A multiattribute utility function for a comprehensive health status classification system. Health Utilities Mark 2. *Medical Care* **34**: 702–722.
- Torrance GW, Feeny DH, Furlong WJ. 2001. Visual Analog Scales: do they have a role in the measurement of preferences for health states? *Medical Decision Making* **21**: 329–334.
- Tsuchiya A, Brazier J, McColl E, Parkin D. 2002. Deriving preference-based single indices from non-preference based condition-specific instruments: converting AQLQ into EQ5D indices. *Health Economics and Decision Sciences Discussion Paper*, University of Sheffield.
- Yang Y, Brazier J, Tsuchiya A, Young T. 2011. Estimating a preference-based index for a 5-dimensional health state classification for asthma derived from the asthma quality of life questionnaire. *Medical Decision Making* **31**: 281–291.
- Young TA, Yang Y, Brazier JE, Tsuchiya A. (2011). The use of Rasch analysis in reducing a large condition-specific instrument for preference valuation: the case of moving from AQLQ to AQL-5D. *Medical Decision Making* **31**: 195–210.