Volume 12 • Number 8 • 2009
VALUE IN HEALTH

International Comparisons in Valuing EQ-5D Health States: A Review and Analysis

Richard Norman, MSc,[1] Paula Cronin, MPH,[1] Rosalie Viney, PhD,[1] Madeleine King, PhD,[2] Deborah Street, PhD,[1,3,4] Julie Ratcliffe, PhD[5]

[1]Centre for Health Economics Research and Evaluation (CHERE), Faculty of Business, University of Technology, Sydney, NSW, Australia; [2]Psycho-oncology Co-operative Research Group (PoCoG), University of Sydney, Sydney, NSW, Australia; [3]Centre for the Study of Choice (CenSoc), Faculty of Business, University of Technology, Sydney, NSW, Australia; [4]Department of Mathematical Sciences, Faculty of Science, University of Technology, Sydney, NSW, Australia; [5]Flinders Clinical Effectiveness, Flinders University, Adelaide, Australia

ABSTRACT

Objective: To identify the key methodological issues in the construction of population-level EuroQol 5-dimensions (EQ-5D)/time trade-off (TTO) preference elicitation studies.

Method: This study involved three components. The first was to identify existing population-level EQ-5D TTO studies. The second was to illustrate and discuss the key areas of divergence between studies, including the international comparison of tariffs. The third was to portray the relative merits of each of the approaches and to compare the results of studies across countries.

Results: While most articles report use of the protocol developed in the original UK study, we identified three key areas of divergence in the construction and analysis of surveys. These are the number of health states valued to determine the algorithm for estimating all health states, the approach to valuing states worse than immediate death, and the choice of algorithm. The evidence on international comparisons suggests differences between countries although it is difficult to disentangle differences in cultural attitudes with random error and differences as a result of methodological divergence.

Conclusions: Differences in methods may obscure true differences in values between countries. Nevertheless, population-specific valuation sets for countries engaging in economic evaluation would better reflect cultural differences and are therefore more likely to accurately represent societal attitudes.

Keywords: cost-utility analysis, EQ-5D, health economics methods, health-related quality of life.

Introduction

Cost-utility analysis (CUA), where outcomes are measured in terms of quality-adjusted life-years (QALYs), is the main approach used to measure and value the impacts of treatments. The US Panel on Cost-Effectiveness in Health and Medicine recommends the use of QALYs [1]; the UK National Institute of Health and Clinical Excellence has most commonly used CUA [2,3] and has recently recommended that it should be the preferred outcome measure; and CUA is increasingly used in Australia in the evaluation of pharmaceuticals and medical services. In the recently released PBAC guidelines, a preference is expressed for the use of CUA [4].

While CUA is simple in concept, it presents challenges in practice. QALYs are designed to allow comparisons across interventions with disparate outcomes across different health-care conditions and population groups. Eliciting valuations for all health states that may be relevant to a disease or intervention is time consuming and costly, and comparison of valuations across interventions and diseases requires comparability of methods. Multiattribute utility instruments (MAUIs), which comprise a generic descriptive quality of life instrument and a scoring algorithm that covers all health states described by the instrument (e.g., the EQ-5D, the Short Form-6 dimensions (SF-6D), Health Utilities Index, and Assessment of Quality of Life), have facilitated comparability [5,6]. The scoring algorithm for these instruments is usually generated from a stated preference experiment, typically time trade-off (TTO), standard gamble conducted in a population sample The key advantage of the MAU approach is that it provides community-based valuation of health states for patients who are experiencing the state.

The role of MAUIs in economic evaluation is increasing. For example, the National Institute of Clinical Excellence has recommended the use of the EQ-5D, and the Pharmaceutical Benefits Advisory Committee in Australia has stated a preference for utility weights generated from the use of a MAUI in a clinical trial setting (without specifying a preference for a particular MAUI). Nevertheless, recent reviews have noted that there are significant differences in the performance of different MAUIs [7], which can be attributed to differences in the dimensions covered by the instruments, differences in preference elicitation techniques, and differences in the methods used to derive the scoring algorithm. These differences can have significant impact on valuations of health states and the resulting cost-effectiveness of interventions [8]. There has been relatively little critical appraisal of the methods of development of MAUIs scoring algorithms. In this article, we examine these issues by considering the EQ-5D [9]. We chose the EQ-5D because it is widely used, and there have been a number of different studies undertaken to develop country-specific scoring algorithms. Because the focus of this review is on one MAUI, we do not consider the psychometric aspects of the instrument but, rather, focus on the methods for development of the scoring algorithm. Many of the issues we raise are relevant to other MAUIs.

Overview of the EQ-5D

The EQ-5D is a tool developed by the EuroQol group (Rotterdam, The Netherlands) (www.euroqol.org) and has five dimensions

Address correspondence to: Richard Norman, Centre for Health Economics Research and Evaluation (CHERE), Faculty of Business, University of Technology, Sydney, PO BOX 123, Broadway, Sydney, NSW 2007, Australia. E-mail: richard.norman@chere.uts.edu.au

10.1111/j.1524-4733.2009.00581.x

  1098-3015/09/1194 1194–1200

Table 1 The EQ-5D

Dimension	Description
Mobility	
1	I have no problem in walking about.
2	I have some problems in walking about.
3	I am confined to bed.
Self-care	
1	I have no problems with self-care.
2	I have some problems washing and dressing myself.
3	I am unable to wash and dress myself.
Usual Activities	
1	I have no problems with performing my usual activities.
2	I have some problems with performing my usual activities.
3	I am unable to perform my usual activities.
Pain/Discomfort	
1	I have no pain or discomfort.
2	I have moderate pain or discomfort.
3	I have extreme pain or discomfort.
Anxiety/Depression	
1	I am not anxious or depressed.
2	I am moderately anxious or depressed.
3	I am extremely anxious or depressed.

intended to represent the major areas in which health changes can manifest. These areas are mobility, self-care, usual activities, pain/discomfort, and anxiety/depression. Each dimension contains three levels, classified as "No Problems," "Some Problems," and "Extreme Problems." Details are shown in Table 1. Thus, there are $3^5 = 243$ potential states in the descriptive system. The TTO approach is used to value a selection of these states and then to impute values for the remainder using simple regression. The use of TTO for valuing EQ-5D states is well described in other works [10,11]. For states considered to be preferable to immediate death, a respondent is faced with a choice between 10 years of a particular chronic health state defined in EQ-5D space with a period of x years in full health. The aim of the TTO is to identify a value of x for which the individual is indifferent to the choice. The value for the better-than-death health state is defined as $x/10$.

Regarding Table 1, it should also be noted that we will treat health states with the same levels as identical throughout this article (e.g., health state 12321 is the same irrespective of language). As of March 2009, the EQ-5D has been translated into 100 different languages (with a further 24 awaiting ratification). The comparability of versions is a reasonable assumption because all translations are reviewed by EuroQol Group members and ratified by the EuroQol Group Executive Committee. The Euroqol Web site states that translation consists of two forward translations of the EQ-5D English source version, two back translations, lay assessment, and the production of a full report describing each stage of the process (http://www.euroqol.org).

Our analysis of this EQ-5D/TTO approach involves two strands: first, we look at how to elicit societal valuations for EQ-5D states under the York Research Group on the Measurement and Valuation of Health TTO protocol [9]. We begin by identifying some key themes and issues that run across the population valuation studies. Then, we look at international comparisons and discuss whether it is necessary to provide nationality-specific tariffs for the EQ-5D valuation system.

Methods

The initial target of this study was to identify all large general population valuations studies employing the EQ-5D as the tool for describing health. EMBASE and MEDLINE were searched for such articles. To be considered for inclusion, the analysis had to present primary research in English and be published since 1995. Because it was expected that a proportion of good quality reports may be unavailable in peer-reviewed publications, the reference lists of articles identified in the main search were used to identify further studies. Because all of these identified nonpeer-reviewed publications were available on the EuroQol Web site (http://www.euroqol.org), the list of EuroQol Plenary Meeting Proceedings was scanned for further studies relevant to this work. To be included, a study had to attempt to value all 243 states described by the EQ-5D. Beyond this constraint, we chose to be conservative in our approach to exclusion because we were seeking to identify divergence in approach.

For each included study, details most relevant to the analysis of the methods used were identified. Key areas for discussion were selected. These areas were the precise formulation of the algorithm, the number of states directly valued in the survey to generate weights, the method to value states worse than death, the influence of time preferences of results, and international comparisons in predicted values across EQ-5D space.

The algorithms were compared by expanding the approach used by Busschbach et al. [12], who compare the directly valued states in the UK, Germany, and Spain. For this, Busschbach et al. used the UK results as the benchmark. The predicted preference scores for the states under the UK algorithm were then ranked in descending order. The preference scores under each of the other algorithms are generated by using the same ordering as in the UK study. We extended this approach by including all identified algorithms. Thus, we can identify any tendency for countries to trade off quantity of life for quality of life, and identify whether countries differ in their relative valuations of the five dimensions.

Results

10 articles [11,13–21] that met the inclusion criteria were identified, of which eight were published in peer-reviewed journals. These are listed in Table 2. It should be noted that there are, at present, no such results for Canada or Australia, two countries strongly supportive of the use of CUA in health-care decision-making. Two studies utilized the visual analog scale (VAS) as the primary method of valuation [14,15]. Although this technique is widely used in preference elicitation more generally, the age of the two VAS studies in this area suggests that it has been superseded by the TTO although work by Parkin and Devlin suggests that the VAS remains a valuable tool [22].

Three significant methodological differences emerged regarding the survey structure and the development of the algorithm. The first regarded the number of states that need to be directly valued to estimate valuations for the complete EQ-5D space. The second is the approach to valuing states considered to be worse than death. The third is the choice of the algorithm to model those states not directly valued. There were a number of additional issues that might also be considered such as the validity of the TTO method and the assumption of constant proportional trade-off that it is founded on. Nevertheless, it was felt that this had been adequately covered elsewhere [23,24,25].

The Number of Directly Valued States

Given that the EQ-5D has 243 individual possible states, it is unsurprising that no study has attempted to ask respondents to directly value each of these states. Therefore, the pertinent question becomes how best to form a representative fraction of the entire space that allows a good estimation of the remainder of the EQ-5D states in whichever way that is defined. Two approaches have been adopted to form this representative fraction. The

Table 2 Identified studies

Study	Country	Sample size	Age limit/range	States directly valued	Valuation method	Preferred algorithm
Dolan [10]	UK	3395	>18	13 from 43	Time trade-off	Main effects plus NI
Tsuchiya et al. [11]	Japan	543	>20	17 plus 11111 and dead	Time trade-off	Main effects plus NI
Badia et al. [1]	Spain	975	Unspecified	13 from 43	Time trade-off	Main effects plus NI
Bjork and Norinder [14]	Sweden	1000	Range between 18 and 78	13	Visual Analog Scale	Not stated
Lamers et al. [19]	The Netherlands	300	Range between 18 and 75	17 plus 11111 and dead	Time trade-off	Main effects plus NI
Greiner et al. [17]	Germany	339	>15	13 from 43	Time trade-off	Main effects plus NI
Wittrup-Jensen et al. [21]	Denmark	1331	Range between 18 and 91	14 states used per respondent (22222, 33333, 2 mild states, 8 other states, 2 further states related to patients with diabetes or heart disease), plus death and 11111	Time trade-off	Main effects plus NI
Jelsma et al. [17]	Zimbabwe	2488	>15	7 from 38, including at least one of each severity level (see Table 3)	Time trade-off	Main effects
Shaw et al. [23]	USA	4048	Range between 18 and 99.3	Unconscious, dead, 11111, 33333 plus 2/5 mild states and 9/36 remaining	Time trade-off	DI
Devlin et al. [9]	New Zealand	2741	>18	Three versions were sent out, each containing death and 12 other health states (defined in terms of EQ-5D).	Visual Analog Scale	Main effects with and without NI

original Dolan et al. approach valued 43 states, and each respondent directly valued a subset of these 43 [9,26]. An alternative approach (described here as the Tsuchiya approach), which uses 17 states, all rated by each respondent, was developed [11]. Both sets of states are given in Table 3.

Lamers et al. [19] investigate these alternative approaches. Using data from Dolan et al. [9], they assumed that all respondents would value 11111 (full health) and in addition value 12,17,22,27,32,37 or 42 of the remaining 42 states. Samples of size 50, 100, 200, 300, 400, 600, and 800 were assumed. The outcome for each of these combinations is the mean absolute error (MAE) between the predicted values from the subsequent algorithm and the values observed in the data set. MAE is a useful tool for estimating appropriateness because it shows the fit of the model to the data. Nevertheless, other diagnostics might also be of value, for example, out-of-sample or split-sample prediction (of directly valued states or otherwise).

As expected, the MAE is negatively associated with both the sample size and the number of health states directly valued. Additionally, they contrast these data with the results of Dolan et al. [26], which suggest that not only does the 17-state approach used by Tsuchiya et al. [11] lead to a lower MAE than that of Dolan et al. but also it may lead to a lower MAE than if each respondent valued 17 (or even 22) randomly assigned states from the 42 (although the difference does not appear to be statistically significant). The mean correlation for the predicted and actual values if 22 states from 42 are randomly selected is 0.986 (SD = 0.006), whereas the figures for the 17 states used by Tsuchiya et al. was 0.989 (SD = 0.002).

A related question concerns whether the 17- and 43-state approaches are optimal in terms of study design. For equal precision in each of the effect estimates to be allowed, it is necessary to have equal frequency of appearance for each of the levels for each of the attributes. Because there is a disproportionate number of the better health states, that is, states with attributes at level 1, in the 43 Dolan states [9] or the 17 Tsuchiya states [11], there is greater precision at that healthy end of the scale. The other related issue involves the estimation of interactions. Although only 10 degrees of freedom are required for the estimation of main effects, a further 40 are required to estimate two-factor interactions. Of course, if certain level combinations do not appear together (and perhaps do not make sense together), then estimation of all two-factor interactions becomes impossible.

Transformation of Values for Worse than Dead States

Although it is plausible that the poorer states in the EQ-5D might be considered worse than immediate death, certain methodological issues arise from generating an algorithm with a subset of states that includes states worse than death. While anchoring death at 0 and full health at 1 gives meaning to states that lie in that range, it is difficult to interpret different values below 0. The lack of a tool that is well suited to this task means that existing articles have taken a variety of approaches to valuing these states, some of which raise further questions.

All articles begin from the same starting point, by asking respondents to choose between immediate death and a period of 10 years of life, some of which is spent in the state worse than death and some in full health. In the majority of articles [11,13,16,17,19,21], if the individual is indifferent between immediate death and x years of the bad state followed by $(10 - x)$ years of full health, the score for the state worse than death is then calculated in the following way:

$$\text{Preference score (State worse than death)} = (x/10) - 1 \quad (1)$$

Table 3 The states selected by Dolan and Tsuchiya et al. (common states in bold)

Category	Tsuchiya	Dolan (the number valued by each respondent is in parentheses)
Full health	**11111**	**11111**
Very mild	**11112, 11121, 11211, 12111, 21111**	(2 from 5) **11112, 11121, 11211, 12111, 21111**
Mild	**11113, 11131, 11133, 11312**	(3 from 12) 11122, **11131**, **11113**, 21133, 21222, 21312, 12211, **11133**, 22121, 12121, 22112, **11312**
Moderate	**13311, 32211, 32313, 22222**	(3 from 12) 13212, 32331, **13311**, 22122, 12222, 21323, **32211**, 12223, 22331, 21232, **32313**, **22222**
Severe	**23232, 32223, 33323**	(3 from 12) 33232, **23232**, 23321, 13332, 22233, 22323, **32223**, 32232, 33321, **33323**, 23313, 33212
Pits state	**33333**	**33333**

Because x is bounded by 0 and 10, the preference scores for states worse than death are bounded by 0 and –1. The one divergence from this orthodoxy is found in Shaw et al. [20], for whom

$$\text{Preference score (State worse than death)} = x/(10 - x) \quad (2)$$

They allowed the value for x to be between 0.25 and 9.75 years, meaning that the preference score is initially as low as –39. This leads to an asymmetry between states better than immediate death and those worse. This is important because it means that the impact of a brief period in the severest health state is of the same magnitude as a much longer period in full health. Although a poor state such as this might be plausible, it could be argued that the uncertainty surrounding interpretation of states worse than death means that the value we place on these states should not have a dominant influence on the final algorithm. Shaw et al. suggested that states worse than death should be bounded by –1. Thus, they applied a linear transformation to the raw scores, constraining all scores to be in this range [20]. The major problem with this linear transformation is that the valuations in this range are dependent on the minimum length of time the respondent is allowed to endure in the bad health state. If the minimum period allowable in the poor health state increases to, for example, 1 year, all negative values would be divided by nine. The effect of dividing the different health valuations by different factors (defined by the shortest allowable period in the poor health state) is illustrated in Figure 1.

As health moves away from 0 toward –1, the effect of this procedural variable becomes increasingly large and suggests that this divergence from the orthodox position is not justified.

The Construction of the Algorithm

The choice of the algorithm is intrinsically associated with the states directly valued in the TTO. Equal precision around point estimates of main effects depends on equal appearance frequency for each of the levels, which does not occur in the states valued in any of the international articles. Equally, for interactions between levels to be estimated, most of which are plausible, these interactions have to appear in the states directly valued, which certainly are not the case for all pairs of levels. In choosing an algorithm, the benchmark UK study [16] prefers the N3 model, in which the algorithm is a main effects model using dummy variables for levels in each dimension worse than "No Problems," plus the N3 dummy variable, defined as 1 when any of the dimensions are at level 3 (the worst level). Thus, Valuation = 1 – [constant + $\Sigma(\text{dummy}_{l,d} * \text{co-efficient}_{l,d}) + (\text{dummy}_{N3} * \text{coefficient}_{N3})$].

Aside from increased predictive value of the model with this interaction term [9], the intuition behind using such a value is not clear. Indeed, the Japanese data showed no improvement in model fit after inclusion of the N3 term. One potential explanation for including the N3 term is that the first dimension moving to level 3 will have significant spillover effects, perhaps not captured by the other dimensions. The need to adapt to a life with a severe impediment has a disutility that is a one-off. Thus, the second dimension to move to level 3 will have a disutility (illustrated by the coefficient associated with the respective dummy variable) but may have a lesser impact than if the move had occurred from a state with no pre-existing level 3 problems. The reverse argument, claiming that the N3 term has no intuitive appeal, might argue that the extra predictive value is a remnant of the correction methods used to adjust states worse than death to constrain them between 0 and –1. Because these states are considerably more likely to have level 3 dimensions than the general set of states, it is arguable that applying an erroneous transformation, compressing negative values into too small a range, might be identified through lower coefficients being applied to level 3 parameters beyond the first.

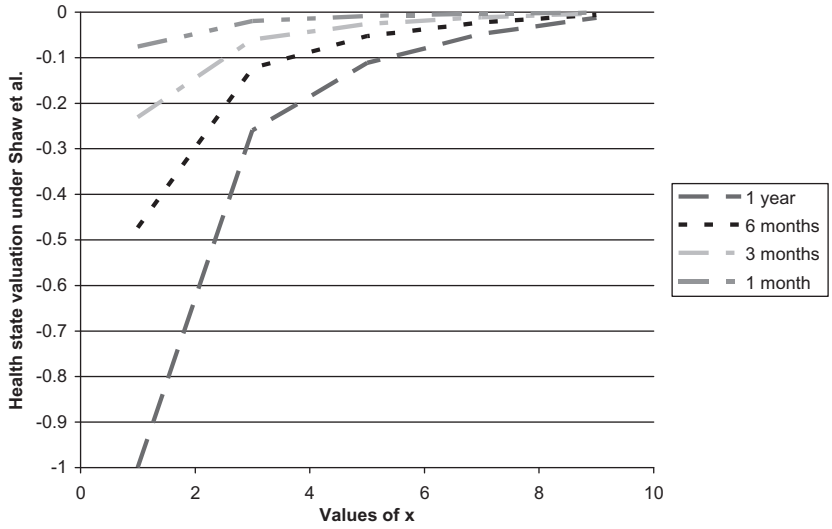

Figure 1 The effect of changing minimum time duration on valuations of states worse than death.

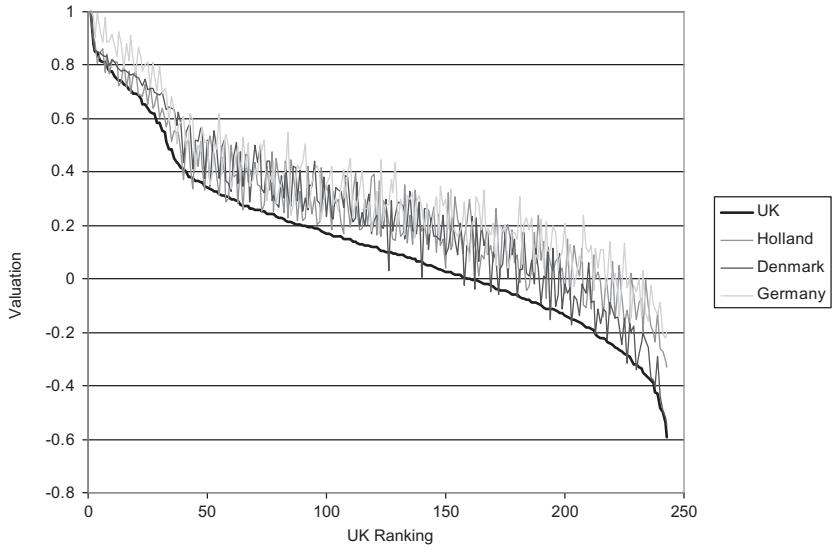

Figure 2 Comparison of Northern European algorithms.*
*This figure uses the UK preference ordering as a baseline for comparison.

Other than the N3 variable, most studies do not utilize interaction terms in their final models. Nevertheless, the intuitive argument in support of interactions can be illustrated by using a number of examples (e.g., the disutility of not being able to do usual activities may vary, depending on whether the person is mobile because this defines what usual activities consist of). A number investigate alternative model specifications containing interactions [19] but generally (and perhaps surprisingly) find that they do not improve the fit of the model [11,13,16,21].

The final issue regarding the algorithm is the use and interpretation of the constant term. Conventionally, the intercept reflects the value of the function when all explanatory variables are 0 (level 1 in the N3 model). Nevertheless, in this case, this interpretation does not hold because 11111 is axiomatically described as full health and is anchored at 1. In the identified articles, there are two approaches in the discussion of the intercept. In the majority of studies, the intercept is allowed to vary from 0 and is interpreted as the disutility associated with not being at perfect health, independent of the disutility associated with the movement within the dimension per se [10]. This could be justified in the same way as the N3 variable was justified above. An alternative approach is taken in a recent US study [20]. The full algorithm used in this study is given by

$$\text{Valuation} = 1 - (\Sigma(\text{dummy}_{l,d} * \text{coefficient}_{l,d}) + \beta_1 D1 + \beta_2 I2\text{-squared} + \beta_3 I3 + \beta_4 I3\text{-squared}), \quad (3)$$

where D1 is the number of dimensions not at level 1 beyond the first, I2 is the number of dimensions at level 2 beyond the first, and I3 is the number of dimensions at level 3 beyond the first. The differences between this approach and the more commonly utilized N3 approach are that Shaw et al. [20] do not allow a constant term (because full health is anchored at 1) and that they identified a broader group of statistically significant interaction terms, albeit specified in a different way. One criticism of both approaches is that they are relatively blunt in their approach to interactions. For example, if we consider the interactions concerning dimensions being at level 3, the effect of there being a number of dimensions at level 3 is independent of the specific dimensions at that level.

International Comparisons

The final question this article considers is the extent to which the use of these different algorithms affects the preference scores associated with the 243 states in EQ-5D space and, thus, whether the choice of model is likely to alter resource allocation decisions. Our results, comparing the wider range of countries using all states defined by EQ-5D space, are shown in Figures 2 and 3 (note that the UK algorithm is smooth because it has been selected as the base case).

We have compared the algorithms to the benchmark in groups of three. When algorithms from Denmark, Germany, and The Netherlands are compared with the UK study, they generate similar preference scores across the range of health states. Generally, these lie above the UK figures but follow the same trend. This suggests that the various dimensions of the EQ-5D have the same approximate relative importance in these countries, but the absolute disutility attached to worsening in the health state in general is estimated to be lower. Regarding the apparent tendency for the UK algorithm to provide health state valuations that are lower than those for other algorithms, it is worth noting that a modified Research Group on the Measurement and Valuation of Health protocol was used in a repeat experiment in a UK population [27], which produced scores generally higher than those derived from the Dolan et al. algorithm [27].

Divergence from this trend can be seen in the countries shown in Figure 3. The Spanish model does not appear to be systematically different from the UK model but displays more variance from the UK model than the Northern European results, suggesting different emphasis between dimensions. The Japanese results are less than those of all other models for mild health states (as a result of a large constant term in the N3 model) but, for worse states, lie above all other models. Under the Japanese model, there are very few states considered worse than death. Additionally, the Japanese results show considerable variance relative to the UK figures. In comparing the Japanese results with the UK, this seems to be the result of a relatively high importance being associated with mobility and a relatively low importance being associated with pain and discomfort, and anxiety and depression. The US study follows a similar pattern to the Japanese results but displays less variability relative to the UK. This unwillingness to trade off quantity of life for quality of life in Japan and the US means that the spread of HRQoL scores is lower in these countries. As noted by Luo et al. and Noyes et al., this will lead to interventions being less cost-effective in CUA because the quality of life gain is likely to be smaller [28,29].

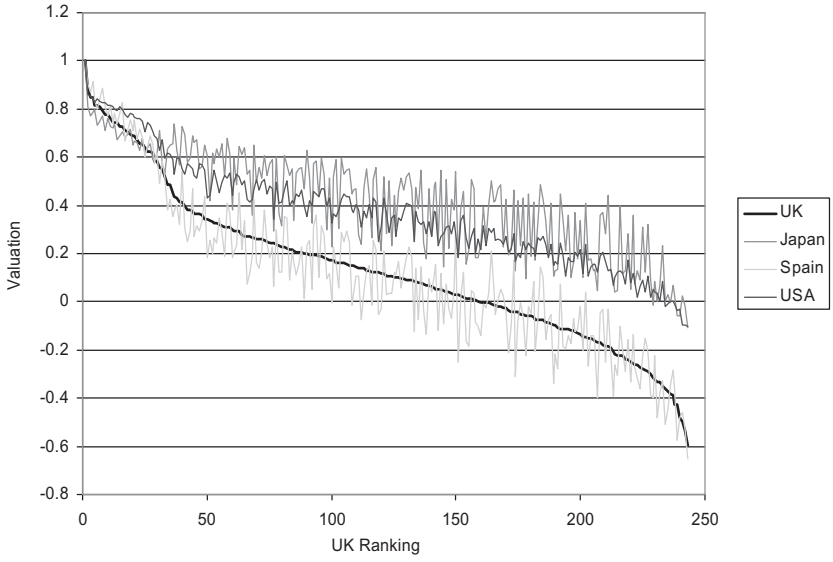

Figure 3 Comparison of UK and non-Northern European countries.*
*This figure uses the UK preference ordering as a baseline for comparison.

The uncertain element in interpreting these results is to identify whether the differences in models are a result of genuine differences in national attitudes toward ill health or whether they are the product of different study designs (including any difference caused by translation issues). In support of the former is the fact that Figure 2 suggests convergence between countries in a geographic locality (Northern Europe). Nevertheless, we believe that to firmly identify a trend in models between countries, we would require a greater number of studies than currently exist, preferably using data collected by using the same mechanism and at the same time point. The analysis of subgroups within a population is also of potential interest because it may identify what drives health state valuation patterns, both within a population and potentially between populations. Potential explanatory factors might include wealth, income, religion, or health expenditure.

Conclusions

This article identifies a number of key methodological questions in the construction of population-level EQ-5D/TTO value sets. The number of states that need to be directly valued is considered, and the best solution may depend on whether it is worthwhile to look for interaction terms. We identified study design issues with the sets of states most commonly selected to be directly valued. The decision regarding number of states leads into a number of questions regarding the choice of algorithm. Then, we identified competing approaches for the valuation of states considered to be worse than death and identified that the approach used by Shaw et al. [20] makes valuations heavily dependent on a parameter of model design (specifically the minimum period of the state considered in the TTO) that should have no effect on the valuation.

Whether country-specific algorithms are necessary is a difficult question that we have only partly addressed. There are clear divergences between countries in their valuations, in terms of both their willingness to trade quantity of life for quality and their relative importance of the five dimensions of the EQ-5D. Our findings indicate that a proportion of the divergences in algorithms are likely to be attributable to genuine cultural differences rather to than methodological differences between studies, which suggests that country-specific algorithms are of importance. This is particularly true in countries that engage in substantial economic evaluation such as Canada and Australia, which are currently reliant on using algorithms derived from countries asserted to be comparably similar in terms of attitude to health.

Source of financial support: Financial support for this study was provided entirely by NHMRC Project Grant (403303). The funding agreement ensured the authors' independence in designing the study, interpreting the data, writing, and publishing the report.

References

1 Gold M. Cost-Effectiveness in Health and Medicine. New York: OUP, 1996.
2 Claxton K, Sculpher M, Drummond M. A rational framework for decision making by the National Institute for Clinical Excellence (NICE). Lancet 2002;360:711–5.
3 Taylor R. Using health outcomes data to inform decision-making: government agency perspective. Pharmacoeconomics 2001;19 (Suppl. 2):S33–8.
4 Department of Health and Ageing. Guidelines for preparing submissions to the pharmaceutical benefits advisory committee (version 4.2). Canberra, 2007. Available from: http://www.Health.Gov.Au/internet/main/publishing.Nsf/content/pbacguidelines-index [Accessed June 30, 2008].
5 Brazier J, Deverill M, Green C. A review of the use of health status measures in economic evaluation. Health Technol Assess 1999;3:i–iv, 1–164.
6 Spilker B. Quality of Life and Pharmacoeconomics in Clinical Trials. Philadelphia, PA: Lippencott-Raven, 1996.
7 Hawthorne G, Richardson J, Day NA. A comparison of the assessment of quality of life (AQoL) with four other generic utility instruments. Ann Med 2001;33:358–70.
8 Bansback N, Davis S, Brazier J. Using contrast sensitivity to estimate the cost-effectiveness of verteporfin in patients with predominantly classic age-related macular degeneration. Eye 2007; 21:1455–63.
9 Dolan P, Gudex C, Kind P, Williams A. The time trade-off method: results from a general population study. Health Econ 1996;5:141–54.
10 Dolan P. Modelling valuations for health states: the effect of duration. Health Policy 1996;38:189–203.
11 Tsuchiya A, Ikeda S, Ikegami N, et al. Estimating an EQ-5D population value set: the case of Japan. Health Econ 2002;11: 341–53.

12 Busschbach JJ, Weijnen T, Nieuwenhuizen M, et al. A comparison of EQ-5D time trade-off values obtaIned in Germany, the United Kingdom and Spain. In: Brooks R, Rabin R, De Charro F, eds. The Measurement and Valuation of Health Status Using EQ-5D: A European Perspective. Berlin: Springer, 2003.
13 Badia X, Roset M, Herdman M, Kind P. A comparison of United Kingdom and Spanish general population time trade-off values for EQ-5D health states. Med Decis Making 2001;21:7–16.
14 Bjork S, Norinder A. The weighting exercise for the Swedish version of the EuroQol. Health Econ 1999;8:117–26.
15 Devlin NJ, Hansen P, Kind P, Williams A. Logical inconsistencies in survey respondents' health state valuations—a methodological challenge for estimating social tariffs. Health Econ 2003;12:529–44.
16 Dolan P. Modelling valuations for EuroQol health states. Med Care 1997;35:1095–108.
17 Greiner W, Claes C, Busschbach JJ, Graf von der Schulenburg JM. Validating the EQ-5D with time trade off for the German population. Eur J Health Econ 6:124–30.
18 Jelsma J, Hansen K, De Weerdt W, Graf von der Schulenburg JM. How do Zimbabweans value health states? Popul Health Metr 2003;1:11.
19 Lamers LM, McDonnell J, Stalmeier PF, et al. The Dutch tariff: results and arguments for an effective design for national EQ-5D valuation studies. Health Econ 2006;15:1121–32.
20 Shaw JW, Johnson JA, Coons SJ. US valuation of the EQ-5D health states: development and testing of the D1 valuation model. Med Care 2005;43:203–20.
21 Wittrup-Jensen KU, Lauridsen JT, Gudex C, et al. Estimating Danish EQ-5D tariffs using the time trade-off (TTO) and visual analogue scale (VAS) methods. In: Norinder A, Pedersen KL, Roos P, eds. Proceedings of the 18th Plenary Meeting of the EuroQol Group. Copenhagen: EuroQol Group, 2001.
22 Parkin D, Devlin N. Is there a case for using visual analogue scale valuations in cost-utility analysis? Health Econ 2006;15:653–64.
23 Bleichrodt H, Pinto JL, Abellan-Perpinan JM. A consistency test of the time trade-off. J Health Econ 2003;22:1037–52.
24 Stiggelbout AM, Kiebert GM, Kievit J, et al. Utility assessment in cancer patients: adjustment of time tradeoff scores for the utility of life years and comparison with standard gamble scores. Med Decis Making 1994;14:82–90.
25 Unic I, Stalmeier PF, Verhoef LC, van Daal WA. Assessment of the time-tradeoff values for prophylactic mastectomy of women with a suspected genetic predisposition to breast cancer. Med Decis Making 1998;18:268–77.
26 Dolan P, Gudex C, Kind P, Williams A. A social tariff for EuroQol: results from a UK general population study. Centre for health economics York discussion paper no. 138 York: Centre for Health Economics, 1995.
27 Macran S, Kind P. Valuing the EQ-5D health states using a modified MVH protocol: preliminary results. Plenary Meeting of the EuroQol Group. Stiges, Spain: Institute de Salut Publica de Catalunya, 1999.
28 Nan L, Johnson JA, Shaw JW, Coons SJ. A comparison of EQ-5D index scores derived from the US and UK population-based scoring functions. Med Decis Making 2007;27:321–6.
29 Noyes K, Dick AW, Holloway RG. The implications of using US-specific EQ-5D preference weights for cost-effectiveness evaluation. Med Decis Making 2007;27:327–34.